

POLSKA AKADEMIA NAUK  
Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego

Andrzej Buda

**DYNAMIKA I STRUKTURA  
KORELACJI NA RYNKACH  
FINANSOWYCH I FONOGRAFICZNYCH**

Praca doktorska wykonana  
pod kierunkiem dr hab. Jarosława Kwapienia

Kraków, 2022



## Streszczenie

Współczynnik korelacji między akcjami definiuje podobieństwo między badanymi walorami i ma znaczenie w konstrukcji optymalnego portfela akcji – czyli takiego, który osiąga maksymalny zysk przy jak najmniejszym ryzyku. Dlatego pierwszą część mojej pracy poświęcam korelacjom na rynkach finansowych i oprócz porównania rynków wschodzących (WIG 20) i dojrzałych (DJIA, DAX, FTSE 100) przy użyciu metod fizyki (wykorzystywane wcześniej w badaniach szkieł spinowych przestrzenie ultrametryczne i drzewa minimalnej rozpiętości) wprowadzam nowe pojęcie – czas życia korelacji. Jest on pomocny w określeniu optymalnego okna czasowego historii cen, potrzebnego do obliczenia współczynników korelacji. Bowiem zbyt krótkie okna czasowe sprawiają, że współczynniki korelacji między walorami pod wpływem fluktuacji cen stają się losowe, z kolei zbyt duże okna czasowe nie odzwierciedlają aktualnych trendów gospodarki, która podlega przecież zmianom. Moje wyniki porównuję z już znanymi rezultatami. Zwłaszcza, że na wszystkich badanych rynkach drzewa minimalnej rozpiętości na ogół odzwierciedlają sektory gospodarcze stworzone przez badane spółki. W dalszej części badam rynki walutowe, które opierają wartość pieniądza w zależności od waluty bazowej, co oczywiście wpływa na strukturę i rozkład korelacji między walutami. Porównuję też trwałość drzew minimalnej rozpiętości w oparciu o różne waluty bazowe.

Drugą część pracy poświęcam rynkom fonograficznym, które są szczególnym przypadkiem rynków towarowych, a wybrałem je do analizy głównie z powodu dostępności dokładnych danych o sprzedaży płyt. Przeprowadzam analizę analogiczną jak na rynkach finansowych definiując notowania artysty na listach sprzedaży podobnie jak notowania spółek na giełdzie. W efekcie otrzymuję interesujące korelacje między artystami, które na drzewie minimalnej rozpiętości MST odzwierciedlają tradycyjne podobieństwo stylistyczne dla wszystkich gatunków, z wyjątkiem muzyki pop. Bowiem o ile pozostali artyści podobnie jak akcje na giełdzie zachowują się kolektywnie tworząc sektory gatunkowe, to zamiast sektora artystów w estetyce pop otrzymuję sektor złożony z najbardziej znanych gwiazd, ale różnych gatunków. Oprócz omówienia danych doświadczalnych (sezonowość, etc.) przedstawiam dość bogatą analizę sieciową i statystyczną (gęstości prawdopodobieństwa, wykładnik Hursta, etc.) w porównaniu do rynków finansowych i innych rynków towarowych np. energii, a także trzęsień ziemi.

Ponadto aby wyjaśnić dynamikę i trajektorie sprzedaży wprowadzam opracowany wspólnie z Andrzejem Jarynowskim stochastyczny model sprzedaży płyt, który w konfrontacji z rzeczywistością i drzewami minimalnej rozpiętości odzwierciedlającymi kolektywność oddziaływań konsumenckich prowadzi do bardzo ciekawych wniosków. Bowiem okazuje się, że główną przyczyną sezonowego wzrostu sprzedaży płyt pod koniec każdego roku wcale nie jest zagęszczenie premier płytowych, ale zwiększająca rozmiar rynku dodatkowa grupa konsumentów, którzy kupują płyty przed świętami Bożego Narodzenia, co przekłada się na strukturę rynku. Pozostałe cechy dynamiki sprzedaży model stochastyczny odzwierciedla i przewiduje już całkiem skutecznie.

W trzeciej części pracy omówione zostało oddziaływanie wielu rynków na siebie w postaci najbardziej prawdopodobnych ścieżek infekcji rozumianej jako rozprzestrzenianie się popularności pojedynczych utworów (singli) w 12 krajach Europy. Na podstawie danych od końca lat 60. XX wieku dla ponad 1700 singli udało się wyodrębnić strukturę rynków w Europie i ich klastrów, które pokrywają się z bliskością kulturową bądź językową krajów i zależą od nośnika - ery analogowej, cyfrowej i Internetu. Analiza przepływu popularności singli w skali jednego kraju podzielonego na regiony – Wielkiej Brytanii (dla danych z roku 1978) także została przeprowadzona.

Proces ten został wyjaśniony także w ostatnim rozdziale z punktu widzenia modelu agentowego dla globalnego rynku singli, złożonego z rynków krajowych, stanowiących podsieci typu Barabasięgo-Albert o bezskalowej strukturze. Dynamika popularności utworu o zadanej atrakcyjności analizowana jest na zasadach znanych z epidemiologii (agresywność wirusa) i rozprzestrzeniania się innowacji. Założenia modelu zostały przetestowane zarówno w sytuacji, kiedy lokalne rynki mogą stanowić klastry jak również kiedy globalny rynek składa się z podsieci bez klastrów. Odpowiada to różnym wariantom struktury rynku w jego historii, a także bogactwu możliwej dynamiki rozprzestrzeniania się popularności w wielu wariantach – w zależności od atrakcyjności utworu, wyboru początkowego węzła sieci od którego zacznie zdobywać popularność, a także innych parametrów sieci.

## SPIS TREŚCI:

<b>0. Wstęp.....</b>	<b>6</b>
0.1 Fizyka a rynki finansowe.....	6
0.2 Rynki finansowe a pojęcie złożoności.....	9
0.3 Miary złożoności.....	14
0.4 Miary korelacji.....	15
0.5 Sieci złożone.....	19
<b>Rozdział 1. Hierarchiczna struktura rynków finansowych.....</b>	<b>26</b>
1.1 Współczynnik korelacji między walorami.....	26
1.2 Przestrzeń ultrametryczna portfela akcji.....	32
<b>Rozdział 2. Czas życia korelacji.....</b>	<b>38</b>
2.1. W poszukiwaniu optymalnego okna czasowego.....	39
2.2 Średni czas życia korelacji.....	41
<b>Rozdział 3. Korelacje między walutami.....</b>	<b>51</b>
3.1 Rynek walutowy, kursy wymiany i korelacje.....	51
3.2 Struktura drzewa minimalnej rozpiętości (MST) dla walut.....	54
<b>Rozdział 4. Korelacje między artystami. Dynamika sprzedaży płyt .....</b>	<b>60</b>
4.1 Rynek fonograficzny jako rynek towarowy.....	60
4.2 Założenia i analogie z rynkami finansowymi. Analiza sprzedaży płyt.....	64
4.3 Hierarchiczna struktura rynku fonograficznego.....	71
4.4 Statystyczne własności rynku fonograficznego.....	77
<b>Rozdział 5. Modelowanie i dynamika rynku fonograficznego. Cykl życia produktu.....</b>	<b>87</b>
5.1 Cykl życia produktu w przemyśle płytowym. Dynamika sprzedaży.....	87
5.2 Stochastyczne modelowanie krzywej sprzedaży płyt.....	96
5.2.1 Markowskie Przełączanie Stanów: przypadek stacjonarny.....	96
5.2.2 Markowskie Przełączanie Stanów: przypadek niestacjonarny.....	99
5.4 Konfrontacja wyników modelowania z rzeczywistością – konkluzje.....	102
<b>Rozdział 6. Oddziaływanie wielu rynków - rynek singli.....</b>	<b>108</b>
6.1 Ścieżki dyfuzji, cykle życia produktu między rynkami w Europie.....	108
6.2 Algorytm przyłączania.....	112
6.3 Wielka Brytania jako inicjator – przepływ popularności w ramach jednego rynku.....	118
6.4 Sieć rynkowa a sieć kulturowa w różnych erach rozwoju nośników.....	125
<b>Rozdział 7. Ścieżki zdobywania popularności na wielu rynkach – modelowanie agentowe.....</b>	<b>129</b>
7.1 Ścieżki zdobywania popularności między rynkami.....	129
7.2 Model zdobywania popularności dla rynku lokalnego.....	135
7.3 Przypadek wielu rynków.....	146
7.4 Porównanie z danymi empirycznymi.....	161
<b>Podsumowanie.....</b>	<b>167</b>
<b>Literatura.....</b>	<b>171</b>

# 0. Wstęp

## 0.1 Fizyka a rynki finansowe

Tematem mojej pracy jest analiza rynków finansowych i rynku fonograficznego pod kątem ich dynamiki i struktury korelacji. Analizy tej dokonuję w oparciu o dane empiryczne, wykorzystując metody fizyki statystycznej i ciała stałego oraz wprowadzając nowe pojęcia, jak czas życia korelacji, pokazujący ciekawe, do tej pory nieznanne, własności badanych rynków giełdowych. W przypadku rynku fonograficznego, który jest szczególnym przykładem rynku towarowego, zwieńczeniem pracy są stworzone od podstaw 2 modele - stochastyczny i agentowy, opisujące dynamikę sprzedaży płyt.

Mogłoby się wydawać, że zagadnienia te leżą poza granicami fizyki, ponieważ dotyczą finansów i funkcjonowania układów, które są tradycyjnie przedmiotem zainteresowania nauk społecznych. Tymczasem bardziej dogłębne spojrzenie zarówno na układy społeczne, jak i na przedmiot zainteresowań fizyki, pozwala na sformułowanie tezy, że nie dość, że takie układy mogą być z powodzeniem badane przez fizyków, to dodatkowo właśnie fizyka dysponuje najlepiej rozwiniętymi narzędziami, które można zastosować do takich badań. Przede wszystkim fizyka jest nauką, która poszukuje prawidłowości rządzących przyrodą, zwłaszcza tych, które są uniwersalne dla wielu zjawisk. W związku z tym fizyka wypracowała język, którym można te uniwersalne prawidłowości i zjawiska opisać. Jednym z takich fundamentalnych pojęć w fizyce jest pojęcie oddziaływania pomiędzy dwoma lub więcej ciałami lub elementami jakiegoś większego układu. Oddziaływania te mogą należeć zarówno do tzw. oddziaływań podstawowych, jak i mogą być bardziej abstrakcyjne. W tym drugim przypadku przez oddziaływanie rozumie się po prostu wpływ jednego elementu na inny. Takie rozumienie oddziaływań otwiera możliwość opisu w ten sposób układów badanych przez inne dyscypliny naukowe: układów biologicznych, społecznych, ekonomicznych, czy też związanych z przekazywaniem informacji.

Okazuje się też, że pewne zjawiska, które obserwuje się w nawet odległych od siebie układach, mają bardzo podobny charakter. Rodzi to pomysł, by układy te poddać badaniom przy użyciu tej samej metodologii. Fizyka, jak się wydaje, oferuje tu najwięcej

spośród wszystkich dyscyplin naukowych, ponieważ jest oswojona z różnorodnością układów i identyfikacją głębokich, nietrywialnych prawidłowości nimi rządzących, co często prowadzi do odkrywania praw uniwersalnych. Inne dyscypliny naukowe nie posiadają tak rozległego obszaru zainteresowania jak fizyka, ponieważ zajmują się typowo relatywnie małym, wyspecjalizowanym wycinkiem przyrody, a to przenosi się także na mniejszą rozległość, i - co za tym idzie - ogólność metodologii przez nie wytworzonej. Co więcej, dziedziny te często korzystają z praw odkrytych przez fizyków i narzędzi przez nich stworzonych. Jednym z najbardziej znaczących przykładów tego typu jest fizyka statystyczna, która jest teorią opisującą strukturę i dynamikę układów o wielkiej liczbie oddziałujących ze sobą elementów składowych. Narzędzia opracowane w ramach tej dziedziny mogą być stosowane nie tylko do opisu np. gazów czy przejść fazowych w układach fizycznych, ale również w każdym przypadku, gdy rozmiar badanego układu jest na tyle duży, że bardziej korzystne jest opisywanie go statystycznie niż analizując zachowanie poszczególnych elementów składowych.

Patrząc na ten problem z punktu widzenia filozofii nauki, można dostrzec, że historycznie fizyka włożyła najtrwalszy wkład w rozwój naszej cywilizacji praktycznie od czasów starożytnej Grecji, począwszy od Talesa z Miletu czy Arystotelesa, który właśnie fizykę uznał za najważniejszą z dziedzin, którymi się zajmował [1]. Pozytywiści XIX wieku, a przede wszystkim neopozytywiści, umieścili fizykę na pierwszym miejscu listy najbardziej naukowych dziedzin jako dziedzinę najlepiej osadzoną w doświadczeniu i logicznym wnioskowaniu, co cenili m.in. empiryści brytyjscy z Ianem Hume na czele. Fizyka znalazła się przed chemią, biologią i socjologią, która w tamtych czasach praktycznie była psychologią społeczną i balansowała na krawędzi naukowości. To właśnie neopozytywiści bowiem, bazując na osiągnięciach logiki i bacznie obserwując wielkie osiągnięcia naukowe swoich czasów – ze szczególną teorią względności i mechaniką kwantową na czele, które potwierdziły się doświadczalnie, ostatecznie ustalili kryteria naukowości. Otóż teoria, aby być naukowa, musi być falsyfikowalna i weryfikowalna, czyli zgadzać się z doświadczeniem. Fizyka oczywiście spełnia oba te warunki. Co więcej, istniejące w niej alternatywne teorie, jak np. mechanika fal i cząstek, jako równouprawnione doskonale uzupełniają się wzajemnie. Dlatego właśnie neopozytywiści jako pierwsi postulowali wyrażenie innych nauk w języku fizyki [2].

W kontekście powyższych postulatów, a także tematyki niniejszej pracy, warto zauważyć, że związki pomiędzy fizyką a teoriami ekonomicznymi i badaniami rynków finansowych

sięgają ponad dwustu lat [3]. Można tu wskazać A. Smitha, ojca współczesnej ekonomii, który inspirował się w swoich badaniach metodologią fizyki i astronomii, a w szczególności podstawowym dla tych nauk tworzeniem teorii na bazie badania prawidłowości empirycznych. Wielu XIX-wiecznych ekonomistów, takich jak L. Walras i V. Pareto, starało się przenieść idee i formalizm fizyki na grunt ekonomii, co doprowadziło do powstania wielu ważnych dla ekonomii pojęć i teorii, takich jak np. pojęcie równowagi rynkowej, inspirowane fizyką statystyczną. Z drugiej strony, w 1900 r., L. Bachelier starał się modelować ruchy cen akcji, wyrażając je przez proces stochastyczny, z dzisiejszego punktu widzenia równoważny ruchowi Browna [4], co wyprzedziło nawet dokonania Smoluchowskiego i Einsteina.

Do przenikania się dziedzin przyczynił się także rozwój techniki obliczeniowej, przez co możliwe stało się analizowanie dużych zbiorów danych. Na przykład impuls do rozwoju geometrii fraktalnej dał B. Mandelbrot, który w 1963 r. badał ceny bawełny i zauważył, że rozkład ich fluktuacji ma charakter potęgowy. Geometria fraktalna znalazła następnie zastosowanie w fizyce, do opisu struktury wielu układów naturalnych i procesów fizycznych. Olbrzymi impuls do rozwoju ekonomii, a zwłaszcza ekonometrii, dali w końcowych dekadach XX w. fizycy zatrudnieni w instytucjach finansowych, którzy dzięki znajomości fizyki statystycznej mieli odpowiednie przygotowanie, by analizować olbrzymie zbiory danych finansowych, zawierających zapisy miliardów transakcji giełdowych, często dostępnych w czasie rzeczywistym (dla porównania, w latach 60-tych B. Mandelbrot miał do dyspozycji tylko około 2 tysięcy danych).

Wreszcie zainteresowanie badaniem rynków finansowych i zagadnień ekonomicznych zaczęli przejawiać sami fizycy akademicy, doprowadzając w połowie lat 90-tych XX w. do wyłonienia się nowej dziedziny nauki, nazwanej ekonofizyką, badającej zjawiska ekonomiczne metodami fizyki, ale uwzględniając przy tym dotychczasowe osiągnięcia ekonomii. Zainteresowaniu fizyków sprzyja także niespotykana w standardowej fizyce doskonała jakość i dokładność danych empirycznych, praktycznie pozbawionych błędów systematycznych.

Rynek fonograficzny jest szczególnym przypadkiem rynku towarowego, w którym przedmiotem handlu są nagrania artystów, a ich wartością - notowania na listach sprzedaży. Dzięki nowoczesnym technologiom od niedawna gromadzone są w skali



globalnej dokładne dane pochodzące z tego rynku o jakości nie odbiegającej od jakości danych z rynków finansowych. Dzięki temu do badań rynku fonograficznego można wykorzystać analogiczne metody, wykorzystywane do badań innych rynków, a których źródłem jest m.in. fizyka statystyczna. Co interesujące, rynek fonograficzny nie był dotąd przedmiotem podobnych badań, a niniejsza praca jest w tej dziedzinie pionierską w skali światowej.

## **0.2 Rynki finansowe a pojęcie złożoności**

Zjawiska ekonomiczne, a zwłaszcza te dotyczące rynków finansowych są niezwykle trudne do opisanie i wyjaśnienia. To, co obserwuje się na rynkach finansowych, czy – szerzej – w gospodarce, jest efektem oddziaływania na te układy ogromnej liczby czynników wewnętrznych i zewnętrznych. Do tych pierwszych zaliczyć można przepływ informacji pomiędzy uczestnikami rynków, brak znajomości warunków początkowych, mikroskopowego stanu rynków i precyzyjnie wyznaczonych praw nimi rządzących czy, ogólniej, praw ekonomii. Do czynników zewnętrznych można z kolei zaliczyć zdarzenia polityczne, czynniki naturalne (katastrofy, zjawiska klimatyczne) i technologiczne oraz wpływ innych rynków. Co więcej, uczestnikami rynków są istoty inteligentne, które potrafią się uczyć i dostosowywać swoje zachowanie do zmiennych okoliczności. Olbrzymia liczba tych czynników i ich wzajemne powiązania powodują, że rynki finansowe są jednymi z najbardziej skomplikowanych układów istniejących we Wszechświecie. Wszystko to razem sprawia, że poprawny i całościowy opis tych rynków wymaga wzięcia pod uwagę informacji dostępnej na różnych poziomach organizacji: na poziomie pojedynczych inwestorów i innych uczestników obrotu gospodarczego, poziomie branż, geografii oraz zjawisk makroekonomicznych.

Potrzebę takiego całościowego opisu spotyka się w przypadku wielu innych układów występujących w przyrodzie i strukturach społecznych. Układy takie zaliczane są do klasy, która jest przedmiotem zainteresowania stosunkowo nowej dziedziny - nauki o układach złożonych. Układy tego typu charakteryzują się dużą liczbą oddziałujących ze sobą elementów składowych, a nieliniowy charakter tych oddziaływań sprawia, że całość zachowuje się inaczej niż układ, będący prostą sumą elementów składowych.

Koncepcja układów złożonych zaczęła się klarować od lat 50-tych XX w. dzięki pracom biologa Ludwiga Bertalanffy, który zajmował się teorią układów otwartych [5]. Skrytykował próby opisu żywych organizmów na gruncie termodynamiki układów zamkniętych, będąc zdania, że układy biologiczne wymagają zupełnie nowego podejścia, uwzględniającego ich otwartość i wymianę energii i informacji z otoczeniem. Z punktu widzenia dynamiki takich układów bardzo ważne jest, że Bertalanffy, wbrew stanowisku neopozytywistów, twierdził, że zmiany zachodzące w procesie ewolucji układów biologicznych zachodzą skokowo, co według dzisiejszego stanu wiedzy wydaje się być prawdą. Za rozwinięcie tej koncepcji można uznać np. późniejszą teorię *punctuated equilibria*, opisującą ewolucję jako ciąg długotrwałych stanów równowagi przerywanych gwałtownymi zmianami [6].

Od czasów prac Bertalanffy'ego pojęcie złożoności układów uległo rozwinięciu, jednak do dziś nie ma jednej uznanej powszechnie definicji, która nie budziłaby żadnych wątpliwości. Wiąże się to przede wszystkim z obecnością tego terminu w wielu odrębnych dziedzinach nauki, na przykład w psychologii, socjologii, cybernetyce, biologii, fizyce, matematyce i teorii informacji, a – co za tym idzie – wielu różnym próbom jego formalizacji (mimo, że intuicyjnie idea złożoności może być w każdej dziedzinie rozumiana podobnie).

Nauka o złożoności określa układy złożone jako układy nieredukowalne, czyli takie, w których na poziomie makroskopowym występują zjawiska, których nie da się wytłumaczyć jedynie na podstawie znajomości ich struktury mikroskopowej. Są to tzw. zjawiska emergentne. Teoria systemów definiuje te własności jako [7]:

- własności układów, których nie obserwuje w ich podukładach;
- własności, które są rezultatem dynamicznej organizacji ich elementów;
- własności globalne, które mogą być efektem mieszanym czynników wewnętrznych i zewnętrznych, związanych z przeszłością układu i warunkami środowiska;
- własności funkcjonalne na wyższym poziomie organizacji mają potencjalnie wielorakie wyjaśnienie na niższych poziomach organizacji (brak jednoznaczności pomiędzy poziomami organizacji).

W przyrodzie zjawiska emergentne są powszechnie spotykane, np. skomplikowane formy ukształtowania terenu, takie jak niektóre typy wybrzeży o strukturze fraktalnej (fiordowe, riasowe, szkierowe), łańcuchy górskie, wydmy piaszczyste, fraktalna struktura systemów rzecznych, wielkoskalowe zjawiska atmosferyczne, takie jak tornada, niektóre formacje chmur, przepływy turbulentne cieczy, życie i formacje organizmów żywych (kolonie bakterii, ławice ryb, gromady ptaków). Podobnie, zjawiska emergentne są cechą układów społecznych: pojawienie się specyficznych cech w psychologii tłumu, istnienie ruchów społecznych, struktur państwowych, wojen, pieniądza, rynków finansowych, literatury, sztuki itp. Zjawiskiem emergentnym na pograniczu biologii i zjawisk społecznych jest z kolei język naturalny i związane z nim efekty wyższego rzędu, takie jak gramatyka czy styl literacki. Z drugiej strony emergentność zjawiska, które może się wydawać najbardziej złożonym spośród nietrywialnych cech organizmów żywych, czyli świadomości, budzi kontrowersje, ponieważ jest ciągle słabo poznane [7].

Cechą układów złożonych, która prowadzi do pojawienia się zjawisk emergentnych, jest samoorganizacja. Jest to długotrwały proces ewolucji układu, który na skutek oddziaływania z otoczeniem poprzez wymianę informacji (energii), zmienia swą strukturę wewnętrzną w celu adaptacji do nowych warunków. Ta zmiana struktury, a właściwie ciąg następujących po sobie zmian, ma charakter nieodwracalny. O nieodwracalności ewolucji układów pisał już L. Dollo w latach 90-tych XIX w., który odnosił ją do układów biologicznych, ale jest to prawdą dla dowolnych układów złożonych. Z punktu widzenia fizyki, nieodwracalność ewolucji jest skutkiem drugiej zasady termodynamiki oraz własności nieliniowych układów o dynamice dysypatywnej. W tym kontekście warunki zewnętrzne i ich wpływ na układ mają charakter parametru kontrolnego, który zmieniając się, zmienia dynamikę układu opisywanego pewnymi równaniami różniczkowymi. Zmiany parametru powodują, że w pewnym momencie układ traci równowagę i zamiast jednego rozwiązania stabilnego pojawiają się dwa lub więcej rozwiązań, z których jedno, któremu odpowiada bieżący stan układu, jest niestabilne. Pojawia się stan nierównowagi i układ musi przejść do nowego stanu równowagi, który odpowiada jednemu z rozwiązań stabilnych. Występuje wtedy tzw. spontaniczne złamanie symetrii. Ciąg takich zmian stanów odpowiada więc ewolucji poprzez sekwencję stanów metastabilnych. Jeśli historia układu jest wystarczająco długa i wystąpiło wiele zdarzeń tego typu, to struktura układu może stać się dowolnie złożona. W konsekwencji samoorganizacja może występować w układach, które są zarazem elastyczne, by mogły zmieniać swą strukturę, ale również

stabilne, by mogły istnieć przez dłuższy okres czasu.

Zmienność zachowania układów samoorganizujących się jest widoczna w zjawiskach kolektywnych, które wyłaniają się z pozornego chaosu i w takim chaosie później zanikają. Do takich zjawisk w układach fizycznych można zaliczyć np. reakcję Bielewska-Zabotyńskiego w układach dyfuzyjnych, agregację ograniczoną dyfuzją (DLA) czy komórki Benarda w cieczy. Podobne efekty występują w świecie przyrody ożywionej. Na przykład ptaki mogą przez dłuższy okres czasu prowadzić indywidualny tryb życia, by w pewnym momencie zacząć grupować się w wielkie stada i odlatywać wspólnie w jednym kierunku, po czym w innym momencie stada ulegają dezintegracji i ptaki powracają do indywidualnego stylu życia. Ta łatwość przechodzenia z nieuporządkowania do uporządkowania i odwrotnie jest jedną z najbardziej charakterystycznych i uniwersalnych cech złożoności. Na rynkach finansowych i towarowych, o których mowa będzie w dalszych rozdziałach, można także zaobserwować takie kolektywne efekty. Bańki spekulacyjne, które są fazami uporządkowanego zachowania inwestorów i wyłaniają się ze nieuporządkowanego stanu równowagi rynkowej, typowo kończą się jeszcze bardziej kolektywnymi krachami, po których równowaga i nieuporządkowanie zostają przywrócone. To samo dotyczy sezonowych okresów wzmożonej sprzedaży płyt na rynku fonograficznym, wywołanych przez stadne zachowania konsumentów. Przyczyny takich zachowań zostaną wyjaśnione w niniejszej pracy.

Jednym z ciekawszych zjawisk spotykanych w układach złożonych jest tzw. samoorganizująca się krytyczność (*self-organized criticality*). Jest to forma samoorganizacji, w której ewolucja układu dąży do osiągnięcia lub przywrócenia stanu krytycznego. Stanem krytycznym w tym rozumieniu jest stan, w którym układ wykazuje gotowość do dowolnie silnej odpowiedzi na zewnętrzne zaburzenie. Jest to stan pokrewny stanowi układu w okolicach punktu krytycznego przejścia fazowego drugiego rodzaju, gdzie obserwuje się pojawianie się struktur o dowolnym rozmiarze, od najmniejszych, do porównywalnych z rozmiarem całego układu (np. w zjawisku perkolacji). Mechanizm samoorganizującej się krytyczności opisuje model stożka piasku, tworzony przez dodawanie pojedynczych ziaren [8]. W modelu tym stożek, początkowo płaski, stopniowo zwiększa kąt swojej tworzącej aż osiągnie on taką wartość, że stożek ulega osypaniu i spłaszczeniu, po czym proces zaczyna się od nowa. Stan krytyczny odpowiada tu takiemu kątowi, w którym pojedyncze ziarno piasku może spowodować zarówno małą lokalną

lawinę, jak i globalną katastrofę. Istotą procesu jest jednak to, że kąt krytyczny jest swoistym atraktorem, do którego stan stożka nieustannie dąży. Model samoorganizującej się krytyczności został wykorzystany do wyjaśnienia powszechności szumu  $1/f$ , występującego w przyrodzie np. w zjawisku trzęsień ziemi, wylewach rzek, muzyce, rytmie ludzkiego serca, sieciach neuronowych, a nawet w funkcjonowaniu mózgu[9].

Szum  $1/f$  jest przykładem jednego z typów zależności potęgowej, często obserwowanych w układach naturalnych. Mimo że zależności potęgowe mogą mieć bardzo różne przyczyny, w tym także dość trywialne[10], to charakteryzują także zjawiska krytyczne – np. wiele wielkości fizycznych skaluje się potęgowo w pobliżu punktu krytycznego. Dlatego też ich obecność w danych empirycznych może być przesłanką, że odpowiedni układ znajduje się w stanie krytycznym. Z tego m.in. powodu panuje pogląd, że prawa potęgowe są jedną z cech układów złożonych. Z drugiej strony zależności potęgowe są obecne w strukturach fraktalnych, stąd wiele układów uważanych za złożone przejawia strukturę tego typu.

Fraktale, które są obiektami samopodobnymi, posiadają strukturę hierarchiczną. W przypadku układów złożonych taka hierarchia struktur może wynikać bądź z ich fraktalnej struktury na danym poziomie organizacji, bądź z wielopoziomowej struktury organizacyjnej. Oba typy hierarchiczności są świetnie widoczne na rynkach finansowych. Globalny rynek dzieli na rynki w poszczególnych częściach świata i w pojedynczych krajach, rynki krajowe dzielą się na sektory gospodarcze, te dzielą się na podsektory, te na jeszcze mniejsze gałęzie branżowe, a te z kolei na niewielkie grupy powiązanych firm. (Analogiczną strukturę ma globalny rynek fonograficzny, który można podzielić na rynki lokalne, ale też na gatunki i podgatunki muzyczne.) Z innego punktu widzenia uczestnicy rynków finansowych, którzy są elementami mikroskopowej struktury rynków, mogą być organizacjami (firmami, funduszami inwestycyjnymi itp.), składającymi się z ludzi. Ludzie, jako organizmy żywe, są zbudowani z komórek, które to komórki mają wewnętrzną strukturę organelli, zbudowanych z molekuł. To tworzy wspomnianą wielopoziomową strukturę, w której jedne poziomy są pierwotnymi elementami wyższych poziomów.

### 0.3 Miary złożoności

Jak już to zostało wspomniane wyżej, układy złożone wymagają równoległego opisu na kilku poziomach organizacji struktury. Choć teoretycznie możliwe, wyprowadzenie opisu takich układów z wiedzy o budowie i oddziaływaniach pomiędzy pojedynczymi ich elementami nie jest wykonalne w praktyce. Dzieje się tak z powodu nieznaności warunków początkowych, co w przypadku zjawisk nieliniowych wyklucza dokładne prześledzenie ewolucji układu w czasie (tzw. czułość na warunki początkowe), a wielka liczba stopni swobody układu wyklucza użycie współcześnie dostępnej mocy obliczeniowej. Te problemy sprawiają, że opis układów złożonych jest albo jedynie jakościowy, albo do badań trzeba wykorzystać te ich aspekty, które poddają się łatwo analizie.

Dużą rolę w badaniu układów złożonych odgrywa analiza danych empirycznych. Dane te w postaci sygnałów (serii czasowych) są dostępne po wyborze odpowiednich zmiennych, które poddają się pomiarom. Własności układów są w związku z tym określane w sposób pośredni poprzez własności tych sygnałów. Złożoność układów przy tym może, choć nie musi, oznaczać złożoności sygnałów.

Pojawia się tu w ogóle problem, jak zdefiniować stopień złożoności ciągu liczb. W przeciwieństwie do definicji złożoności układów, złożoność sekwencji liczbowych można definiować w sposób ilościowy i ścisły. Niestety, można to zrobić na wiele sposobów, bez jednej uniwersalnej definicji[11-12]. Pierwszą z zaproponowanych miar złożoności była złożoność algorytmiczna, która opierała się na długości minimalnego algorytmu, który daną sekwencję może odtworzyć. Jej wadą było przypisywanie nieskończonej złożoności sekwencjom przypadkowym, w związku z czym konieczne było wprowadzenie innych wielkości, które sobie z tym potrafiły poradzić: np. złożoności efektywnej czy głębokości logicznej. Niestety, wielkości te zazwyczaj są trudne do wyliczenia i przez to ich stosowalność bardzo ograniczona. O wiele prościej wylicza się miary oparte na pojęciu entropii informacyjnej:

$$H_I = - \sum_{i=1}^N p_i \ln p_i \quad , \quad (0.1)$$

gdzie  $p_i$  oznacza, że prawdopodobieństwo, że sygnał przyjmie wartość  $i$  z  $N$  dozwolonych wartości (możliwe jest też zdefiniowanie entropii w sposób ciągły). I choć entropia jest użyteczną miarą, to – podobnie jak złożoność algorytmiczna - może osiągać najwyższą wartość dla sygnałów, które są wynikiem serii losowej. Ważną zaletą entropii i metod określania złożoności ją wykorzystujących (np. złożoności efektywnej) jest to, że może być stosowana nie tylko do sygnałów, ale też do układów. W tym wypadku zamiast prawdopodobieństwa wartości sygnału używa się prawdopodobieństwa, że układ jest w danym stanie.

Z faktu, że obiekty fraktalne mogą być intuicyjnie odbierane jako obiekty o dużej złożoności strukturalnej, wielkości charakteryzujące fraktale mogą też służyć jako ilościowe miary złożoności, wykorzystywane zarówno do obiektów, jak i sygnałów. Takimi wielkościami są wymiary fraktalne i wykładniki Hursta. W przypadku multifraktali lub sygnałów multifrakalnych, które są splotami wielu różnych jednorodnych fraktali, a przez to bardziej złożone od jednorodnych frktali, miarą ich złożoności może być też tzw. spektrum multifrakalne[13]. Miary te stosuje się powszechnie w analizie danych pochodzących z rynków finansowych[12]. W przypadku rynku fonograficznego wydaje się, że mają one jednak mniejsze zastosowanie.

Złożoność układu można też opisywać poprzez identyfikację efektów kolektywnych rozumianych jako redukcja efektywnej liczby stopni swobody układu. Na poziomie analizy sygnałów empirycznych można to zrobić, badając korelacje pomiędzy sygnałami związanymi z różnymi stopniami swobody. Im więcej sygnałów jest od siebie zależnych i im silniej są one zależne, tym kolektywność jest większa. W tym kontekście takimi miarami mogą być omówione w dalszej części pracy współczynniki korelacji Pearsona i Spearmana oraz informacja wzajemna, wraz z ich rozkładami.

## 0.4 Miary korelacji

Niech dany będzie układ, składający się z  $N$  stopni swobody. Dla stopnia  $\alpha$  możemy zmierzyć pewną zmienną związaną z tym stopniem swobody i utworzyć serię czasową wyników pomiarów  $x_\alpha(i)$ , gdzie  $i$  numeruje kolejne chwile czasu: 1,...,. Zatem układ ten

można opisać przy pomocy macierzy danych  $\mathbf{X}$  o elementach:

$$X_{\alpha i} = \frac{1}{\sigma_{\alpha}} (x_{\alpha}(i) - \bar{x}_{\alpha}) \quad (0.2)$$

gdzie  $\sigma_{\alpha}$  oznacza wariancję, a od sygnału  $x_{\alpha}(i)$  odejmujemy jego wartość średnią, aby wartości były unormowane.

Jednym z podstawowych pojęć opisujących ilościowo współzależność badanych wielkości  $X$  i  $Y$  jest współczynnik korelacji, zdefiniowany jako:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (0.3)$$

gdzie  $\text{cov}(X, Y)$  jest kowariancją zmiennych losowych  $X$  i  $Y$ , a  $\sigma_X$  oraz  $\sigma_Y$  są odchyleniami standardowymi  $X$  i  $Y$ . Z kolei  $\mu_X$  i  $\mu_Y$  oznaczają średnie wartości zmiennych  $X$  i  $Y$ , a  $E$  oznacza wartość oczekiwaną. Jeśli zastosujemy zdefiniowany wzorem (0.3) współczynnik korelacji do wybranej próbki danych w postaci oryginalnego ciągu, to wówczas otrzymamy współczynnik korelacji Pearsona dla tej próbki, który dla odmiany oznaczmy przez  $r$ :

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (0.4)$$

Możemy go prościej zapisać jako:

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right), \quad (0.5)$$

gdzie:

$$s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{oraz} \quad \frac{X_i - \bar{X}}{s_X}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

oznaczają, odpowiednio, odchylenie standardowe, zmienną unormowaną i średnią.

W danych empirycznych mamy jednak często do czynienia z tzw. obserwacjami odstającymi, czyli takimi, które znacząco odbiegają wartością od pozostałych, a zatem mają silniejszy od nich wpływ na współczynnik korelacji Pearsona. Wówczas współczynnik Pearsona nie spełnia dobrze swej roli., gdyż jest zdominowany przez pojedyncze zdarzenia ekstremalne.



Alternatywną miarą korelacji, która jest mniej wrażliwa na obserwacje odstające, jest współczynnik korelacji Spearmana, zdefiniowany jako:

$$r_s = \text{corr}(R_x, R_y), \quad (0.6)$$

gdzie  $\text{corr}(R_x, R_y)$  jest współczynnikiem korelacji Pearsona, ale w odniesieniu do rang wartości zmiennych losowych  $X$  i  $Y$  (w tabeli wartości uszeregowanych w kolejności od największej do najmniejszej). Korelacja rang Spearmana może być opisana jako współczynnik kierunkowy prostej najlepiej dopasowanej (metodą najmniejszych kwadratów) do zbioru par rang. O ile korelacja Pearsona odnosi się do liniowej zależności między zmiennymi, to korelacja rang Spearmana wychwytuje także nieliniową zależność, o której mowa we wzorze (0.7). W dalszej części pracy stosuję współczynnik korelacji Pearsona, który jest używany powszechnie w analizie danych finansowych, gdyż interesują mnie korelacje sygnałów w czasie, a współczynnik Spearmana nie zawiera informacji o czasie wystąpienia danego wyniku pomiaru.

Oba powyższe współczynniki korelacji są liniową miarą podobieństwa, co widać z ich definicji, gdzie wartości sygnałów wpływają liniowo na wartość odpowiedniego współczynnika korelacji. Tymczasem w ogólności podobieństwo sygnałów może się opierać o nieliniowe zależności, które nie mają żadnego wpływu na wartości współczynników korelacji. Nieliniowe zależności można wprawdzie uwzględnić, tworząc nieliniowe współczynniki korelacji, gdzie do odpowiednich wzorów (0.4) i (0.6) wartości sygnałów wchodzi w różnych potęgach, jednak w takiej sytuacji należałoby analizować wartości całej rodziny współczynników, różniących się potęgami. W celu jednoczesnego uwzględnienia zależności różnych typów można wprowadzić do analizy informację wzajemną, opartą na entropii informacyjnej. Definiuje się ją jako różnicę pomiędzy entropiami pojedynczych sygnałów, a ich entropią wspólną:

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \quad (0.7)$$

gdzie człon  $H(X, Y)$  jest entropią wspólną i odpowiada za współzależności między oboma sygnałami. Sygnały te są zależne, jeśli  $0 < H(X, Y) < H(X)$ . Informacja wzajemna jest czuła na zależności wszystkich typów – liniowe i nieliniowe.

Jeśli analizie poddawane są więcej niż dwa sygnały, to mamy do czynienia z analizą wielokanałową. W przypadku dużej liczby sygnałów opis przy pomocy wartości miar korelacji dla ich par staje się uciążliwy. Inny, dodatkowy problem występuje w przypadku uogólnienia wzoru (0.7) na więcej sygnałów, gdyż wówczas wymiar problemu rośnie i pojawia się niedobór danych, by wyniki były wiarygodne statystycznie.

Dla zmiennych losowych zarówno korelacja rang Spearmana, jak i współczynnik korelacji Pearsona mogą tworzyć macierz korelacji, która spełnia 4 następujące kryteria:

- jest kwadratowa,
- jest symetryczna,
- wszystkie elementy leżące na diagonalu są równe 1,
- wszystkie elementy macierzowe należą do przedziału od -1 do 1.

Macierz tę można zapisać jako iloczyn dwóch macierzy danych (wzór (0.2)), z których jedna jest transponowana:

$$\mathbf{M} = \mathbf{T}^1 \mathbf{X} \mathbf{X}^T, \quad (0.8)$$

w wyniku czego otrzymujemy macierz w następującej postaci:

$$\mathbf{M} = \begin{bmatrix} r_1(t_1) & r_1(t_2) & \dots & \dots & \dots & r_1(t_i) \\ r_2(t_1) & r_2(t_2) & \dots & \dots & \dots & r_2(t_i) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_N(t_1) & r_N(t_2) & \dots & \dots & \dots & r_N(t_i) \end{bmatrix} \quad (0.9)$$

której elementami są współczynniki korelacji Pearsona między sygnałami  $\alpha$  i  $\alpha'$ . Wyznaczenie macierzy korelacji jest pomocne w identyfikacji efektów kolektywnych i szumu, dzięki szeregowi swoich własności i dobrze określonymu układowi odniesienia w postaci odpowiedniego zespołu macierzy przypadkowych, tj. zespołu macierzy Wisharta [14]. Dla tego zespołu jest dostępnych szereg ścisłych wyników teoretycznych, w tym

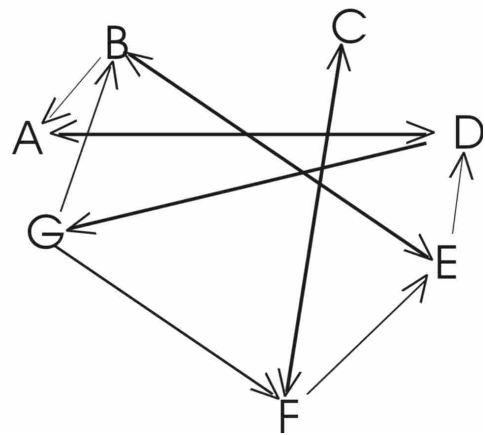
rozkład gęstości wartości własnych[15]. Własności empirycznej macierzy korelacji mogą w ten sposób służyć do identyfikacji efektów nieprzypadkowych, związanych z rzeczywistymi korelacjami w układzie. Przykładem użycia macierzy korelacji i teorii macierzy przypadkowych są analizy rynków finansowych, w których poszczególne sygnały odpowiadają fluktuacjom cen akcji[16-17]. W tej pracy macierze korelacji są podstawą, na której opiera się analiza sieciowa powiązań między walorami na rynkach.

## 0.5 Sieci złożone

W układach, które nie są zbiorem przypadkowo dobranych, niezwiązanych ze sobą elementów, do opisu struktury i oddziaływań między tymi elementami wykorzystuje się teorię grafów (sieci), która jest częścią matematyki dyskretnej. Przedmiotem zainteresowania tej teorii jest graf  $G$ , będący zbiorem węzłów  $v_i$  i połączeń między nimi  $e_{jk}$ , gdzie  $i, j, k = 1, \dots, n$ . Koncepcja grafów ma już długą historię, która sięga XVIII w., gdy L. Euler rozpatrywał tzw. problem mostów królewieckich. Została ona znacznie rozwinięta w latach 50-tych XX w., gdy węgierscy matematycy P. Erdős i A. Rényi rozważali własności grafów o przypadkowych połączeniach. Dopiero jednak pod koniec ubiegłego stulecia nastąpiła prawdziwa eksplozja zainteresowania wykorzystaniem grafów do analizy danych empirycznych, zwłaszcza pochodzących z badań układów złożonych.

Podstawową cechą grafów jest kierunkowość połączeń; jeśli połączenia są symetryczne, to graf jest nieskierowany, natomiast w przeciwnym razie mamy do czynienia z grafem skierowanym, w którym połączenia mają swój początek i koniec. Oba typy grafów, choć pozornie wydają się obiektami geometrycznymi, mogą w istocie być łatwo wyrażone w postaci algebraicznej przy pomocy tzw. macierzy sąsiedztwa. Dla grafów nieskierowanych macierz ta jest symetryczna, natomiast dla skierowanych – asymetryczna. Inną ważną cechą grafów jest rodzaj połączeń, które mogą być ważne lub nieważne (binarne). Przykładowy graf skierowany o binarnych połączeniach został zaprezentowany poniżej w formie macierzy i w formie graficznej.

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



Rys. 0.1 Sieć skierowana

Element macierzowy (1,4) (pierwszy rząd, czwarta kolumna) odpowiada połączeniu pomiędzy węzłami A i D. Gdyby macierz z rysunku 0.1 była symetryczna, odpowiadałaby sieci nieskierowanej.

Podstawowym parametrem każdego węzła jest liczba połączeń (stopień, krotność)  $k$  z innymi węzłami, która w przypadku sieci skierowanych dzieli się na liczbę połączeń przychodzących i wychodzących. Sieć skierowana jest wygodna do opisu jednostronnych oddziaływań między elementami np. sieci wpływów, natomiast w mojej pracy ograniczam się do użycia grafów nieskierowanych ze względu na symetrię problemów, którymi się zajmuję. W zależności od własności statystycznych rozkładu krotności węzłów i mechanizmu tworzenia się połączeń wyróżnić można kilka ważnych typów sieci (termin ten będzie używany wymiennie z grafem).

Historycznie pierwszym rozważanym w fizyce typem sieci były sieci regularne, w których każdy węzeł był charakteryzowany przez taką samą wartość  $k$ . Sieci tego typu były wykorzystywane zwłaszcza w fizyce ciała stałego do opisu struktury krystalicznej. Kolejnym typem grafów były tzw. sieci przypadkowe, wprowadzone przez Erdösa i Rényiego (tzw. model ER), w których połączenia między węzłami tworzyły się w sposób losowy przez wybór pary węzłów, tak by sieć miała określoną wartość średnią  $\langle k \rangle$ , a rozkład krotności był dany przez rozkład dwumianowy:

$$P(\text{deg}(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}, \quad (0.10)$$

gdzie  $n$  – całkowita liczba węzłów w sieci.

Dla  $np = \text{const}$  oraz dużych  $n$  przechodzi on w rozkład Poissona:

$$P(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!} \quad (0.11)$$

Modyfikacją modelu ER jest sieć, która tworzy się nie przez wypełnianie połączeniami istniejących węzłów, ale poprzez dodawanie nowych do już istniejących węzłów, losując za każdym razem węzeł, z którym łączony jest nowy. Rozkład połączeń w takiej sieci jest także dany rozkładem (0.11). Przez długi czas uważano, że sieci typu ER dobrze opisują własności sieci rzeczywistych. Tymczasem z upływem lat okazało się, że sieci rzeczywiste posiadają własności mocno niekiedy odbiegające od tego modelu. Sieci typu ER są jednorodne i żaden węzeł nie jest statystycznie wyróżniony (mówi się w tym wypadku o *sieciach demokratycznych*), natomiast w wielu sieciach rzeczywistych występuje silna niejednorodność i pewne węzły są faworyzowane. Prowadzi to do potęgowych rozkładów krotności. A.-L. Barabási i R. Albert jako pierwsi pokazali, że sieć WWW, do tej pory uważana za sieć typu ER, ma strukturę bezskalową (potęgową):

$$P(k) \sim k^{-\gamma} \quad (0.12)$$

gdzie wykładnik  $\gamma$  przyjmuje wartości w przedziale  $2 < \gamma < 3$  [18]. Potęgowy rozkład  $k$  implikuje istnienie niewielkiej liczby węzłów o dużej krotności (tzw. hubów) oraz dużej liczby węzłów o niewielkich krotnościach. Implikuje także istnienie korelacji strukturalnych w sieci, o ile  $\gamma < 3$  [19]. Okazuje się, że tego typu topologię wykazuje wiele innych sieci rzeczywistych, takich jak: fizyczne sieci komputerowe (hardware)[20], sieci transportowe (połączenia lotnicze, kolejowe itp.)[21], sieci metaboliczne[22], sieci opisujące rynki finansowe[23], a zwłaszcza sieci powiązań społecznych: sieci współpracy artystów[18], kolaboracji naukowych[24], cytowań publikacji naukowych[25], sieci rozprzestrzeniania się epidemii[26], czy sieci znajomości w mediach społecznościowych[27].

Rozpowszechnienie sieci o potęgowych rozkładach opisujących bardzo odmienne od siebie układy spowodowało potrzebę wprowadzenia nowego modelu rozwoju sieci, który w efekcie dawałby tego typu topologię. Najbardziej popularnym modelem jest model wprowadzony przez Barabásiego i Albert (tzw. model BA), w którym mechanizmem wzrostu sieci jest preferencyjne przyłączanie[18]. Polega ono na przyłączaniu nowych

węzłów do sieci za pomocą przypadkowych połączeń z istniejącymi węzłami, przy czym prawdopodobieństwo przyłączenia do konkretnego węzła dane jest przez krotność tego węzła. Taki mechanizm wydaje się być intuicyjny, ponieważ np. w przypadku cytowań publikacji naukowych, szansę otrzymać kolejne cytowania mają przede wszystkim te artykuły, które już zdobyły duży rozgłos i dużą liczbę cytowań.

Sieci bezskalowe posiadają inne ciekawe cechy, które sprawiają, że naturalne układy tak często preferują tę właśnie topologię. Przede wszystkim mała liczba węzłów kluczowych dla struktury sieci powoduje, że przypadkowy wyeliminowanie takiego węzła jest mało prawdopodobne (najwięcej jest węzłów peryferyjnych i mało istotnych). Ponieważ przyroda nie ma informacji o tym, które węzły są hubami, ten typ topologii dobrze zdaje egzamin. (Niestety, problem staje się poważny, gdy atak na sieć przeprowadzi istota inteligentna, która ma informację o położeniu hubów, wtedy taką sieć bezskalową łatwo zniszczyć; przykładem są blackoutu elektryczne i ataki hackerskie w internecie). Drugą istotną własnością takich sieci jest ich optymalność pod kątem zasobów potrzebnych do ich skonstruowania – posiadają krótkie drogi łączące węzły między sobą, nawet jeśli te węzły są od siebie odległe w przestrzeni. Miarą, która może to opisać, jest tzw. średnia długość najkrótszej ścieżki, która dla grafów binarnych ma postać:

$$L = \frac{1}{n(n-1)} \sum_{i,j,i \neq j} L_{ij} \quad (0.13)$$

gdzie  $L_{ij}$  jest długością najkrótszej z możliwych dróg pomiędzy daną parą węzłów  $i,j$  (przez drogę rozumie się liczbę pokonanych pośrednich połączeń). Taka definicja ścieżki jest dopuszczalna tylko dla grafów spójnych, tzn. takich, w których wszystkie węzły są dostępne za pośrednictwem istniejących połączeń. W przypadku sieci przypadkowych ER wraz ze wzrostem  $n$  średnia długość najkrótszej ścieżki rośnie jak:

$$L(n) \sim \frac{\log n}{\log \langle k \rangle} \quad (0.14)$$

Sieci bezskalowe nie są opisywane w taki sam sposób i dla nich  $L$  rośnie znacznie wolniej:

$$L(n) \sim \log \log n \quad (0.15)$$

Oprócz długości ścieżek sieci, opisujących zwartość ich struktury, do ważnych cech należy własność gronowania (klasteryzacji). Polega ona na tym, że sąsiadujące ze sobą węzły tworzą w pełni połączone trójki. Miarą opisującą gronowanie jest tzw. współczynnik gronowania, który jest równy prawdopodobieństwu, że pomiędzy dwoma sąsiadami danego węzła także istnieje połączenie. Lokalnie ten współczynnik dla binarnych sieci skierowanych jest zdefiniowany jako:

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}. \quad (0.16)$$

natomiast dla sieci nieskierowanych:

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}. \quad (0.17)$$

Dla całej sieci globalny współczynnik gronowania jest średnią z  $C_i$  po wszystkich węzłach:

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^n C_i \quad (0.18)$$

i jego wartość określa stosunek istniejących połączeń do wszystkich możliwych połączeń w tej sieci. Sieci rzeczywiste mają typowo dość duży współczynnik globalny, podczas gdy np. sieci ER wykazują małą wartość gronowania.

Wiele sieci rzeczywistych posiada posiada topologię połączeń typu małego świata, gdzie wartość średniej długości najkrótszej ścieżki rośnie stosunkowo wolno ze wzrostem liczby węzłów  $n$ :

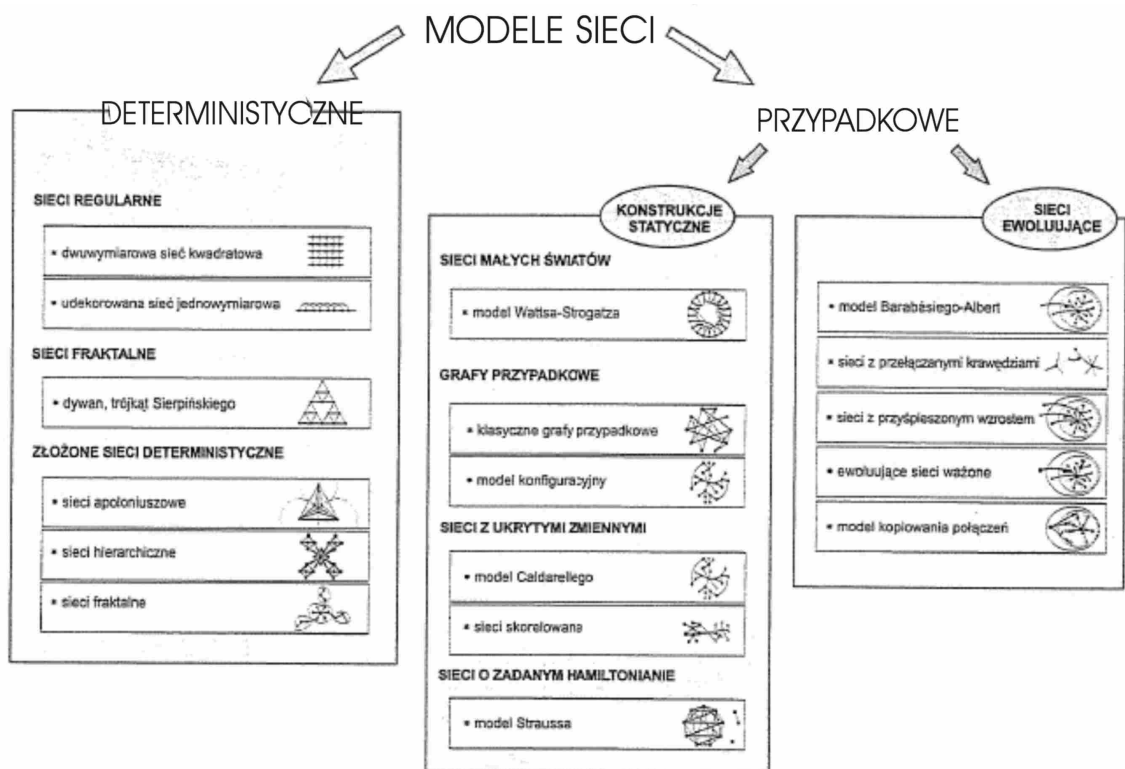
$$L(n) \sim \log n \quad (0.19)$$

a jednocześnie współczynnik gronowania ma dużą wartość.  $L$  rośnie szybciej niż w przypadku sieci bezskalowych i co najwyżej tak szybko jak w sieciach ER, a od własności sieci ER sieci małego świata odróżnia wartość współczynnika gronowania. Nazwa sieci małego świata ma swoje źródło w eksperymencie S. Milgrama, w którym określona została średnia odległość pomiędzy przypadkowo wybranymi ludźmi w sieci znajomości[28]. Okazała się ona zaskakująco mała:  $L \approx 6$ . Dla porównania, współcześnie ta długość jest jeszcze mniejsza i wynosi 4,74 (wg serwisu społecznościowego Facebook) [29].

Tzw. problem małego świata był przez długi czas uważany za ciekawostkę, aż do momentu, gdy D.J. Watts i S. Strogatz opublikowali model, w którym własności sieci małego świata można otrzymać w niezwykle prosty sposób, przełączając losowo drobny ułamek istniejących połączeń w sieci regularnej[30]. Model ten, zwany modelem WS, daje sieć o małej wartości  $L$  i dużej wartości  $\langle C \rangle$ . Obecnie wiadomo, że sieci małego świata są powszechne w przyrodzie i strukturach społecznych. Podzbiorem sieci małego świata są np. sieci bezskalowe, nazywane czasem sieciami ultramałego świata, gdyż dla nich średnia długość najkrótszej ścieżki jest opisywana wzorem (0.15).

Jest wiele innych miar charakteryzujących sieci, np. asortatywność, która odnosi się do korelacji między węzłami różnych typów. Dla różnych układów wielkość ta waha się od przypadku dysasortatywnego, gdy węzły o dużej krotności łączą się najchętniej z węzłami o małej krotności (np. Internet, sieci zależności pokarmowych w ekosystemach, sieci reprezentujące rynki finansowe[12]), do asortatywnego, gdy węzły tych samych typów grupują się razem (sieci społecznościowe, sieci współautorstwa publikacji naukowych) [31]. W przypadku rynku fonograficznego spodziewać się należy *mieszania asortatywnego*, ponieważ największe gwiazdy, o dużej liczbie współpracowników, często współpracują z artystami o podobnej randze. Inną miarą jest rozmiar sieci lub, w przypadku sieci składających się z oddzielnych fragmentów, największego jej spójnego podzbioru (*giant cluster*). Jest to istotna miara w przypadku sieci, w których mamy do czynienia z pojawianiem się lub znikaniem połączeń, gdyż można dzięki niej określić, czy w sieci ma miejsce odpowiednik fizycznego zjawiska perkolacji. Samo zjawisko perkolacji ma olbrzymie znaczenie dla problemu funkcjonowania sieci, ponieważ przejście z sieci spójnej do rozłącznych jej fragmentów w przypadku pewnych układów powoduje destrukcję sieci[32].





**Rys. 0.2.** Klasyfikacja różnych modeli sieci, na podstawie [19]

Klasyfikację modeli sieci przedstawia Rys. 0.2. W klasyfikacji tej najważniejszym kryterium jest podział na sieci przypadkowe i deterministyczne, a następnie na statyczne i ewoluujące. W tej pracy przedmiotem badań będą głównie sieci ewoluujące, powstałe w sposób indeterministyczny.

Współcześnie obserwuje się niebywały rozwój teorii sieci, co pozwoliło zarówno na odkrycie wielu interesujących zjawisk w układach rzeczywistych, wyrażanych przez reprezentacje sieciowe, jak i opracowanie nowych metod badania tych układów. Jednym z najbardziej spektakularnych osiągnięć analizy sieciowej jest wydobywanie informacji, które nie są bezpośrednio dostępne obserwacjom. Chodzi tu o identyfikację pewnych ukrytych poziomów struktury układów, odtwarzanie nieznanymi, ale istniejącymi fragmentami struktury oraz przewidywanie przyszłych zdarzeń. Przykładem są tu sieci kontaktów w organizacjach przestępczych, przewidywanie miejsca i czasu przestępstw, przewidywanie wyroków sądów na podstawie znajomości składów sędziowskich i przewidywanie decyzji konsumenckich na podstawie historii zakupów.

# Rozdział 1. Hierarchiczna struktura rynków finansowych

## 1.1 Współczynnik korelacji między walorami.

Na rynkach kapitałowych podstawą do określenia przyszłej wartości danego waloru jest wykorzystywana analiza fundamentalna w połączeniu z analizą techniczną. Pierwsza z nich służy do określenia bieżącej wartości danego waloru, np. produktu czy akcji spółki, na podstawie przesłanek wynikających z oceny sytuacji na rynku, kondycji przedsiębiorstwa, jakości zarządzania itd. i porównywania tej wartości z aktualną wyceną tego waloru przez rynek giełdowy, natomiast druga z nich ma charakter czysto statystyczny i pozwala na przewidywanie ruchu cen  $P(t)$  na podstawie ich historii bez wnikania w ich relację z rzeczywistą (pozagiełdową) wartością waloru. Analiza techniczna jest wykorzystywana przez inwestorów znacznie częściej niż analiza fundamentalna, ponieważ jej stosowanie nie wymaga aż takich kompetencji i dostępu do wielkiej ilości danych, często niejawnych, jak ma to miejsce w przypadku analizy fundamentalnej. Ponadto w przypadku danych o cenach i transakcjach giełdowych mamy do czynienia lepszą ich jakością i mniejszym zaszumieniem niż w przypadku tzw. danych fundamentalnych. Ze względu na statystyczny charakter analizy technicznej, inwestor dysponuje też w tym przypadku nieporównywalnie większym zestawem narzędzi analitycznych o dobrym umocowaniu formalnym.

Podstawowym pojęciem na rynkach finansowych jest cena  $P(t)$ , czyli cena waloru (akcji, towaru, waluty itd.) w chwili  $t$ , a także stopa zwrotu  $Y(t)$ , określająca zmianę ceny wobec wartości w chwili poprzedniej:

$$Y(t) = P(t) - P(t-1). \quad (1.1)$$

Oczywiście w ten sposób można badać rynki finansowe zarówno na przestrzeni minut i sekund, jak i na przestrzeni wielu miesięcy i lat. Jednak w perspektywie długoterminowej wpływ na wartość cen akcji – a tym samym także stóp zwrotu  $Y(t)$  – może mieć inflacja. Oprócz tego postrzeganie ludzkie jest zdominowane przez różnice względne (np. procentowe wzrosty cen), a nie przez różnice bezwzględne – np. w kontekście wrażeń zmysłowych opisuje to prawo Webera-Fechnera[33]. Dlatego, aby zminimalizować wpływ powyższych efektów na zmiany cen, wygodniej jest używać logarytmicznej stopy zwrotu,

definiowanej następująco:

$$Y(t) = \ln P(t) - \ln P(t-1) . \quad (1.2)$$

Oprócz ewolucji ceny konkretnego waloru, duże znaczenie inwestycyjne ma podobieństwo tej ewolucji do ewolucji cen innych walorów będących przedmiotem obrotu na rynku. Przykładowo, jeśli obserwowany jest wzrost lub spadek ceny jednej spółki z danego sektora gospodarki, jest duża szansa, że ruchy te pociągną za sobą wzrost lub spadek cen innych spółek z tego sektora. Fakt ten może prowadzić do osiągnięcia większych zysków inwestycyjnych, ale jest związany bezpośrednio ze zwiększonym ryzykiem w przypadku, gdy inwestor ulokuje kapitał w jednej branży lub w jednym typie towaru.

Miarą podobieństwa najczęściej wykorzystywaną w analizie technicznej jest współczynnik korelacji Pearsona, który w przypadku zdefiniowanych wzorami (1.1) i (1.2) stóp zwrotu ma postać:

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2) (\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} , \quad (1.3)$$

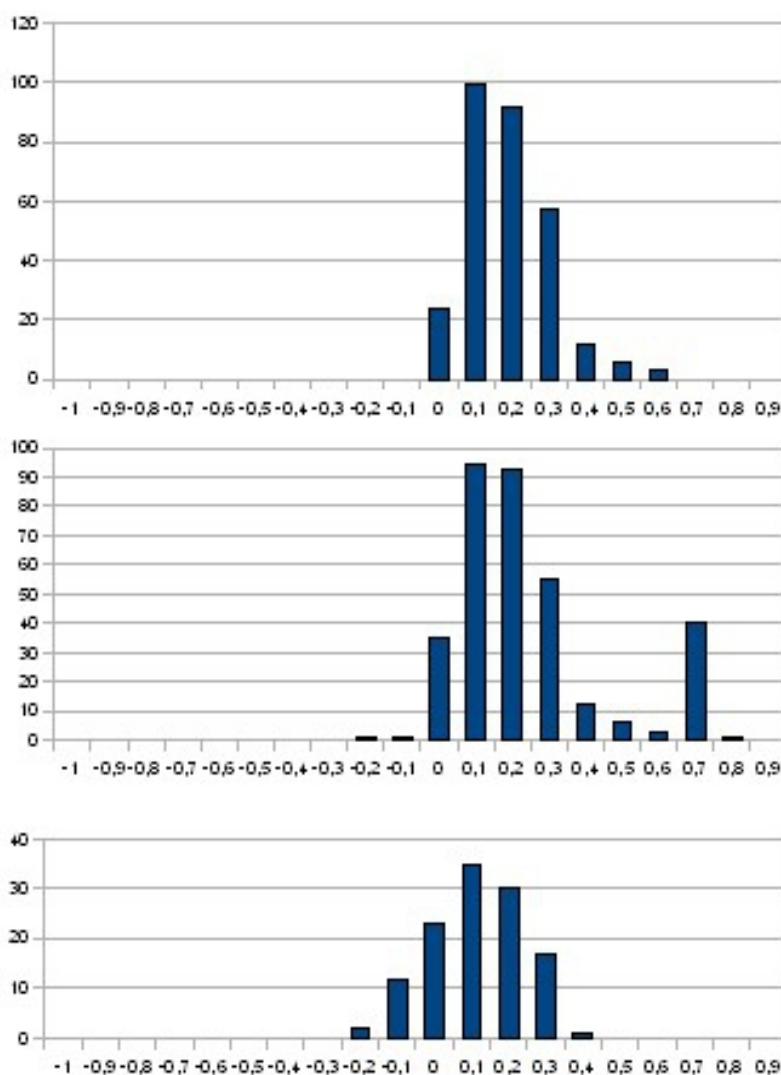
gdzie  $Y_i$  oraz  $Y_j$  oznaczają stopy zwrotu akcji  $i$  oraz  $j$ , a średnie wartości i odchylenia standardowe obliczane są na podstawie znajomości historycznych cen w danym okresie czasu. Współczynnik (1.3) pozwala na określenie siły korelacji pomiędzy walorami, a także odróżnienie korelacji dodatnich od antykorelacji. Przykładowo: wzrost ceny paliwa powoduje wzrost cen akcji spółek paliwowych oraz spadek cen akcji spółek produkujących samochody, natomiast spadek cen paliw determinuje zjawisko przeciwne – czyli spadek cen akcji spółek paliwowych, a wzrost samochodowych. Zatem współczynnik korelacji między spółkami branży paliwowej i samochodowej będzie ujemny, natomiast współczynnik korelacji między akcjami należącymi do tej samej branży będzie dodatni. W praktyce współczynnik korelacji wykorzystywany jest rutynowo w wyznaczaniu ryzyka inwestycyjnego konkretnych portfeli i określaniu składu optymalnego portfela, który maksymalizuje zysk przy minimalnym ryzyku lub odwrotnie (tzw. teoria optymalnego portfela sformułowana przez Markowitza [34]).

W mojej pracy analizuję korelacje między akcjami tworzącymi portfele głównych indeksów giełdowych na rynkach dojrzałych (*established markets*), takich jak amerykański Dow Jones Industrial Average (DJIA) i niemiecki Deutsche Aktienindex (DAX), oraz na rynkach wschodzących (*emerging markets*) – tu przykładem jest polski WIG20. Do analizy korelacji wybrałem ten sam okres od 1.01.1997 do 1.11.2004, tak aby porównać, jak zachowywały się w nim różne rynki. Jako wartość  $P(t)$  brałem pod uwagę cenę w chwili zamknięcia sesji giełdowej w danym dniu. Wybrałem 30 akcji tworzących główny indeks giełdowy DJIA na giełdzie w Nowym Jorku, 26 z 30 akcji tworzących DAX i 20 akcji tworzących WIG20. Główne indeksy giełdowe w praktyce składają się z portfeli akcji o największych obrotach, a ich skład zmienia się co kilka miesięcy. Zwłaszcza, że pojawiają się też nowe walory. Dlatego w obliczeniach uwzględniłem te spółki, które były notowane na giełdach przez cały badany okres, tak aby ich liczba w każdym z trzech portfeli akcji była porównywalna. Brałem także pod uwagę 67 ze 100 akcji tworzących indeks Financial Times Stock Exchange (FTSE 100) na giełdzie londyńskiej, jednak ten dojrzały rynek w niektórych przypadkach wykluczyłem z dalszych badań ze względu na zbyt dużą liczbę akcji w porównaniu z trzema pozostałymi rynkami.

W Tabeli 1.1 przedstawiłem pary spółek należących do indeksu DJIA o największych wartościach współczynnika korelacji pomiędzy ich akcjami w badanym okresie czasu, natomiast w Tabeli 1.2 – pary o najmniejszych wartościach tego współczynnika. Z danych zgromadzonych w obu tabelach widać, że dla akcji tworzących DJIA wartości współczynnika korelacji zawierają się w granicach 0,0-0,72. Co charakterystyczne, nie obserwuje się żadnych par ujemnie skorelowanych. Wynika to prawdopodobnie z tego, że na dojrzałych rynkach można handlować specjalnie przygotowanymi portfelami akcji wchodzących w skład konkretnego indeksu giełdowego, co nieuchronnie prowadzi do skorelowania dodatniego ruchów cen tych akcji. Ewentualne ujemne korelacje mogłyby się pojawiać w krótkich okresach czasu (na skutek zdarzeń dotyczących konkretnych spółek), ale w tym przypadku, gdy okno czasowe obejmuje długi okres zdominowany przez korelacje dodatnie, korelacje ujemne nie są już widoczne.

**Tabela 1.1** Silnie skorelowane pary z portfela DJIA:

Pary akcji	Współczynnik korelacji
C-JPM	0.72
AXP-C	0.68
JPM-AXP	0.65
GE-AXP	0.61
C-GE	0.59
JPM-GE	0.56
IP-DD	0.5
JNJ-MRK	0.5
IP-AA	0.5



**Rys. 1.1** Rozkłady współczynników korelacji uzyskane dla par akcji tworzących portfele DJIA (górny wykres), DAX (środkowy wykres) i DJIA (dolny wykres).

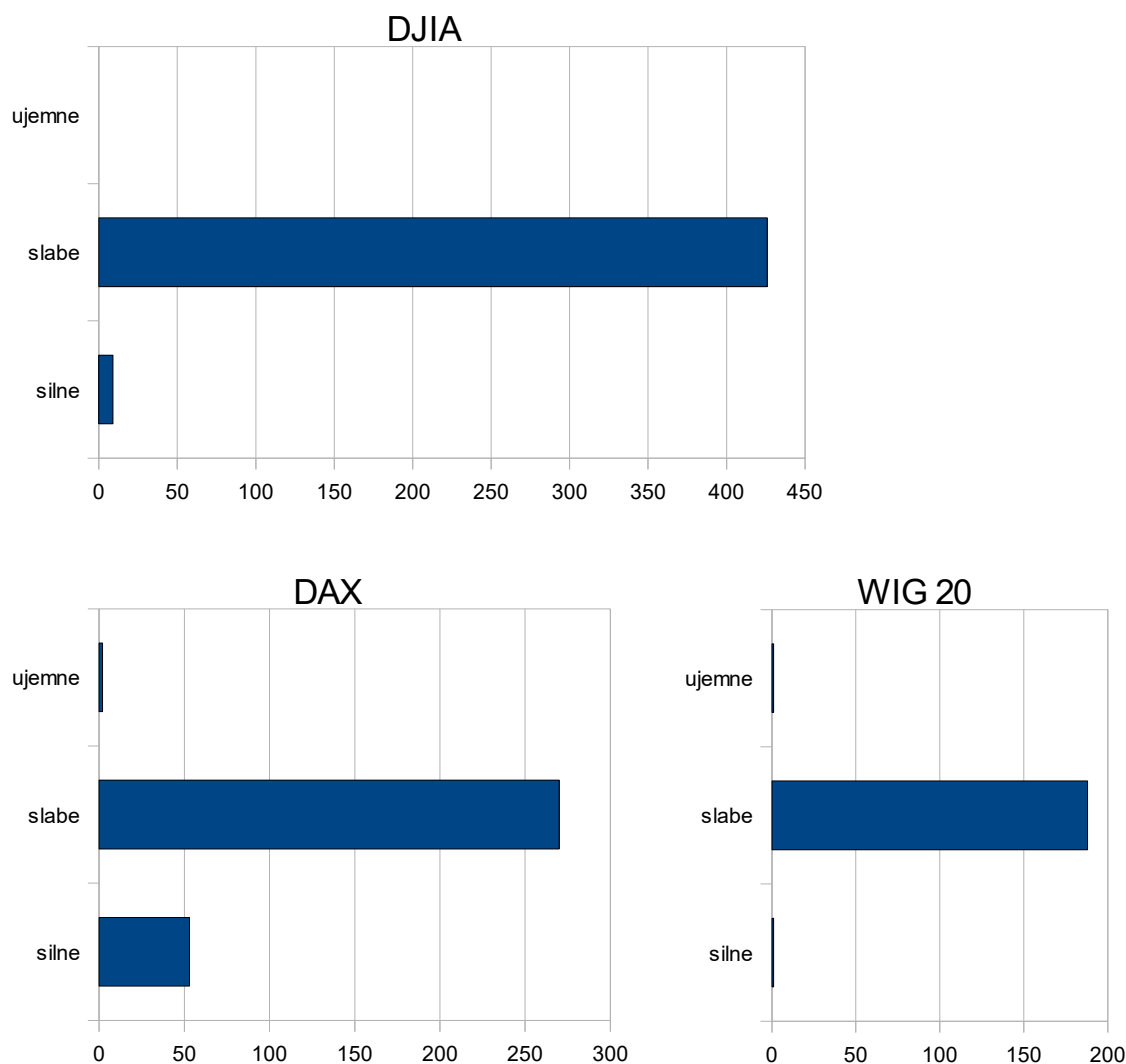
Rozkłady współczynników korelacji dla wszystkich trzech analizowanych rynków przedstawia Rysunek 1.1. Jak widać na tych wykresach, wartości tych współczynników tworzą ciągłe rozkłady. W praktyce taka ciągłość może być utrudnieniem dla podejmowania decyzji inwestycyjnych, zwłaszcza gdy portfele konstruuje się ręcznie, a nie w sposób algorytmiczny. Dlatego wygodniej jest wyrazić siłę korelacji w postaci pewnych poziomów, z których każdy będzie reprezentował pewien przedział wartości współczynników Pearsona. Moją propozycją jest wyróżnienie trzech takich poziomów: poziomu silnych korelacji, słabych korelacji i ujemnych korelacji:

1. Korelacje silne  $\rho \in [1/2, 1]$  ;
2. Korelacje słabe  $\rho \in [0, 1/2)$  ;
3. Korelacje ujemne  $\rho \in [-1, 0)$  .

Wyróżnienie poziomu ujemnych korelacji jest podyktowane tym, że z punktu widzenia teorii portfelowych uwzględnienie ujemnie skorelowanych par spółek w portfelu obniża jego ryzyko [34].

Rysunek 1.2 przedstawia liczbę par spółek o różnej sile korelacji dla trzech analizowanych rynków. Najsilniej skorelowanym rynkiem okazuje się niemiecki DAX, pośrodku lokuje się DJIA, a najsłabsze korelacje obecne są na rynku WIG20. W szczególności w indeksie DAX istnieje tylko jedna ujemnie skorelowana para spółek, tj. ALV (Allianz AG) i DBK (Deutsche Bank), natomiast najsilniej skorelowana jest para BAY (Bayer AG) i BASF (0,83). Podobnie, jedna ujemnie skorelowana para jest obecna w WIG20: EXB (Exbud) i MEXP (Metalexport), podczas gdy najsilniej skorelowana spółka to JLF (Jelfa) i KBH (Kable Holding) – 0,50.

W przypadku większej liczby rozpatrywanych akcji zamiast analizy współczynników korelacji dla poszczególnych par spółek, których liczba rośnie jak  $n^2$ , wygodnie przeprowadzić rachunek macierzowy, wykorzystując do tego celu zdefiniowaną w poprzednim rozdziale macierz korelacji (0.9). Jej własności, takie jak np. widmo wartości własnych pozwalają na określenie globalnych, średnich własności rynku, ale także na identyfikację grup walorów powiązanych ze sobą silniej niż z innymi walorami obecnymi na rynku.



**Rys. 1.2.** Liczba korelacji ujemnych, słabych i silnych na badanych rynkach (DJIA, DAX, WIG 20).

W ekonofizyce w odniesieniu do rynku akcji rachunek macierzowy po raz pierwszy zastosowany został w pracach [16, 35] oraz [17], w których pokazano, że w ten sposób można wyróżnić średnie zachowanie rynku jako całości, można zidentyfikować sektory rynku oraz mniejsze grupy silnie ze sobą powiązanych firm bez konieczności rozpatrywania poszczególnych współczynników korelacji. Od tamtego czasu metodologia macierzy korelacji stała się rutynowo używanym narzędziem do analizy danych finansowych złożonych z wielu równoległe rejestrowanych szeregów czasowych [12; 36-37]. Oprócz swojego globalnego charakteru, macierze korelacji mają dobrze określony układ odniesienia w postaci macierzy przypadkowych (tzw. zespół macierzy Wisharta), których własności są znane. W ten sposób można stwierdzić, czy własności macierzy empirycznych odpowiadają rzeczywistym korelacjom (czyli czy zawierają efekty kolektywne), czy też mają charakter szumu.

## 1.2 Przestrzeń ultrametryczna portfela akcji

W mojej analizie zamiast bezpośredniego rozpatrywania własności macierzy korelacji i określania struktury rynku na ich podstawie, traktuję macierz jako punkt wyjścia do analizy sieciowej, a w szczególności do tworzenia drzew minimalnej rozpiętości (ang. *minimal spanning tree*, MST), które także mogą wiarygodnie opisywać strukturę rynku. Drzewa MST przedstawiają zależności pomiędzy elementami układu jako strukturę w przestrzeni metrycznej, opartej na metryce euklidesowej  $d(i,j)$ :

$$\begin{aligned} \text{(i)} \quad & d(i,i) = 0 && \text{wtedy i tylko wtedy gdy} && i = j \\ \text{(ii)} \quad & d(i,j) = d(j,i) && && \\ \text{(iii)} \quad & d(i,j) \leq d(i,k) + d(k,i) . && && \end{aligned} \quad (1.4)$$

Drzewa MST mogą obrazować strukturę korelacji pomiędzy elementami danego układu, jeśli zastąpi się współczynnik korelacji, który nie spełnia np. pierwszego z warunków (1.4), przez odległość  $d(i,j)$ , określoną wzorem:

$$d(i,j) = \sqrt{2(1 - \rho_{ij})} \quad , \quad (1.5)$$

wprowadzonym w pracy [38]. Odległość ta spełnia wszystkie warunki (1.5); spełnienie warunku drugiego wynika z tego, że dla danej pary sygnałów współczynnik korelacji jest symetryczny.

Rozpatrując zbiór  $n$  szeregów czasowych, macierz korelacji jest określona przez  $n(n-1)/2$  współczynników korelacji (na diagonalu, gdzie mamy korelacje między tymi samymi akcjami, występują z założenia jedynki) i stosując wzór (1.5) można otrzymać taką samą liczbę odległości  $d(i,j)$  i utworzyć tzw. macierz odległości  $\mathbf{D}$ . Drzewo MST tworzone jest przez redukcję macierzy  $n \times n$  do grafu o  $n-1$  krawędziach i charakteryzuje się własnością minimalnej sumarycznej długości krawędzi (minimalną rozpiętością).

Drzewa MST są przydatnym narzędziem do analizy złożonych systemów hierarchicznych w przestrzeni ultrametrycznej. (W fizyce od dawna wiadomo, że koncepcja ultrametryczności jest ściśle związana z koncepcją hierarchiczności, co zaobserwowano w



szkłałach spinowych [39], które są pierwowzorem sfrustrowanych układów nieuporządkowanych [40]). Odległość w przestrzeni ultrametrycznej definiujemy zgodnie z dwoma pierwszymi warunkami (1.4) oraz dodatkowo warunkiem zastępującym warunek trzeci (1.4):

$$d(i,j) < \max [ d(j,k), d(k,i) ] \quad (1.5)$$

Uwzględnienie takiej metryki pozwala na wprowadzenie diagramów hierarchicznych [38], które wykorzystane zostaną w dalszych rozdziałach pracy.

Procedurę tworzenia drzewa MST podaje, znany od lat 60-tych XX w., algorytm Kruskala [41-42]. MST powstaje najpierw od połączenia pary akcji najsilniej skorelowanych. W przypadku portfela DJIA są to City Group (C) i JP Morgan (JPM) (patrz Tabela 1.1), których współczynnik korelacji 0.72 zgodnie ze wzorem (1.4) daje połączenie między akcjami odległością równą 0.75. Kolejna najsilniej skorelowana para AXP-C ma odległość 0.8, a zatem dopisujemy ją do poprzedniej gałęzi od strony C. W efekcie z pierwszych dwóch najbliższych połączeń otrzymujemy:

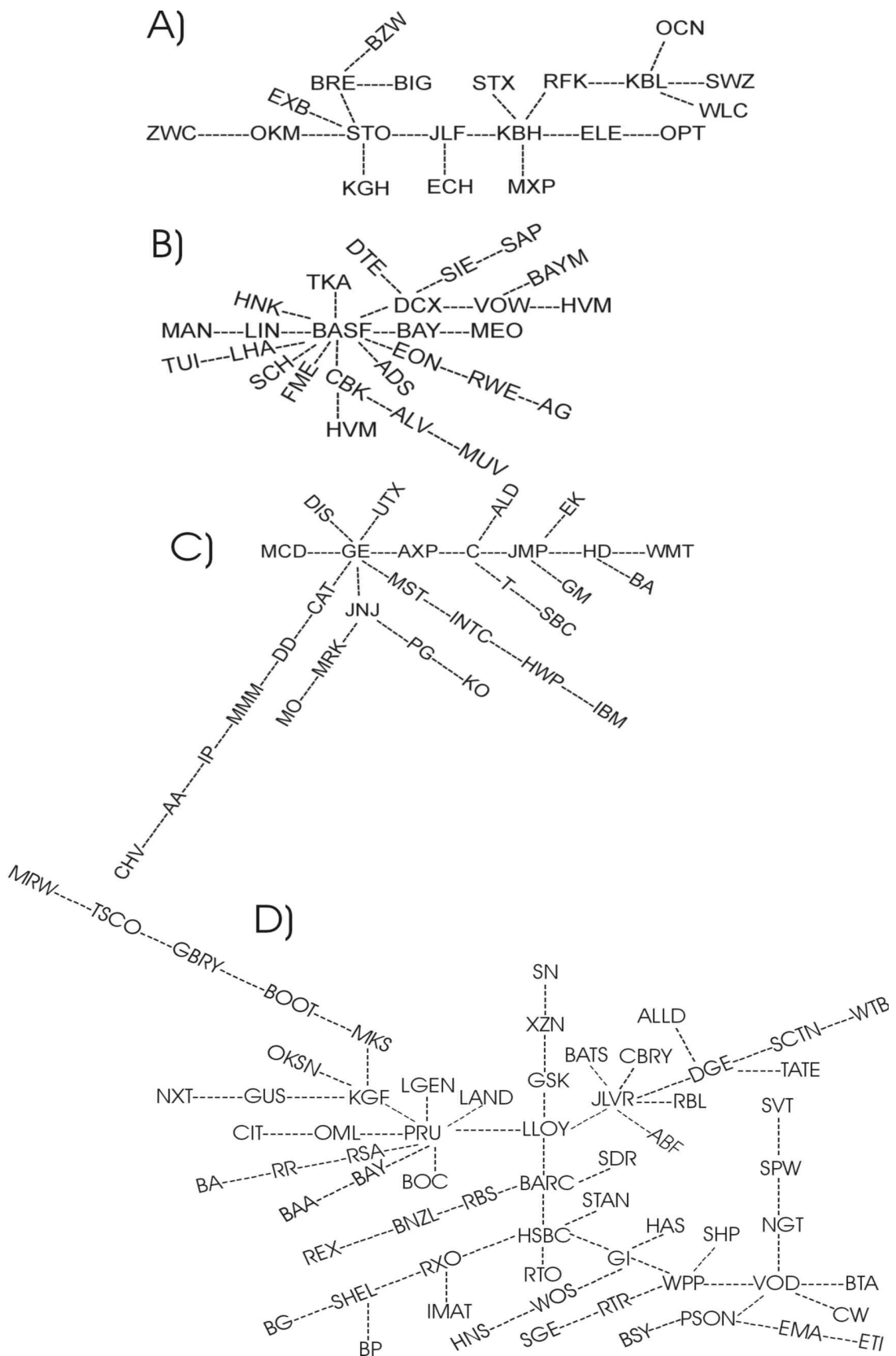
AXP-C-JPM .

Następna para z Tabeli 1.1. to: JPM-AXP, o odległości 0.83. Tyle, że obie te spółki są już ze sobą połączone za pośrednictwem C, a zatem nie musimy rysować nowego połączenia. Przy okazji widzimy, że metryka  $d$  spełnia nierówność trójkąta, bowiem:

$$0.83 < 0.75 + 0.80 .$$

Kolejna para C-GE przynosi nam dopisanie GE do drzewa od strony C i tak samo postępujemy z kolejnymi parami – czyli dorysowujemy jedynie spółki, których nie było wcześniej. W ten sposób otrzymałem drzewa minimalnej rozpiętości dla wszystkich czterech rynków, które przedstawiłem na Rysunku 1.3. Atutem takiego przedstawienia portfeli jest odzwierciedlenie sektorów i podsektorów gospodarczych na drzewie MST, bowiem gałęzie drzewa świetnie pokrywają się z gałęziami ekonomii zarówno na wschodzących, jak i dojrzałych rynkach.

Na każdym z przedstawionych drzew widoczna jest hierarchia węzłów (spółek), wśród których można wskazać węzły dominujące, posiadające największą liczbę połączeń z sąsiadami: dla rynku amerykańskiego jest to GE, dla angielskiego - PRU i ULVR, dla niemieckiego – BASF, a dla polskiego – KBH, a także węzły o znaczeniu lokalnym o mniejszej liczbie sąsiadów niż węzły główne (np. na rynku amerykańskim są to C, JMP, na angielskim - HSBC, BARC, WPP, VOD czy KGF) i węzły peryferyjne, połączone z jednym lub dwoma sąsiadami. Struktura tego typu jest charakterystyczna dla rynków finansowych, w tym zwłaszcza dla rynków akcji, co zostało pokazane w kilku wcześniejszych pracach na ten temat. Mantegna zbadał korelacje wśród spółek amerykańskich indeksów DJIA i S&P500 i pokazał, że w latach 1989-1995 dominującą spółką był także GE [38]. Podobny wynik uzyskano w pracy [43] dla danych z lat 1981-1999 dla spółek z indeksu S&P500, w pracy [44] dla 100 spółek o największej kapitalizacji na rynku amerykańskim (jako grupa niezwiązanych z konkretnym indeksem) w latach 1995-1998 i w pracy [37] dla 1000 największych spółek z tego samego rynku z lat 1998-1999. Na tle wymienionych tu prac, moje wyniki pokazują, że struktura rynku jest stabilna na przestrzeni lat, gdyż dane analizowane przeze mnie obejmowały okres do 2004 roku.



**Rys. 1.3** Drzewa minimalnej rozpiętości MST uzyskane dla akcji tworzących portfele: WIG20 (A), DAX (B), DJIA (C) i FTSE 100 (D). Długość krawędzi pomiędzy symbolami spółek jest związana z wartością  $d(i,j)$  określoną wzorem (1.5): im dłuższa krawędź, tym większa wartość  $d(i,j)$ .

Drzewa MST zilustrowane na rysunku 1.3 pokazują też, że ich struktura odzwierciedla podział gospodarki na sektory branżowe, gdyż spółki z tego samego sektora często ze sobą sąsiadują. Ten wynik także potwierdza, rezultaty uzyskane dla danych z wcześniejszych okresów czasu [38,43,37]. Przykładowo, w obrębie indeksu DJIA spółki informatyczne i komputerowe (MSFT, INTC, HWP, IBM) są zlokalizowane na wspólnej gałęzi, przyłączonej do spółki GE, która jest konglomeratem skupiającym wiele branż. Tak samo spółki tworzące artykuły spożywcze i codziennego użytku (JNJ, PG, KO, MRK, MO) znajdują się na jednej gałęzi. Z drugiej strony spółki surowcowe, takie jak CHV i AA, są stosunkowo oddalone (w sensie metrycznym) od centrum sieci, co interpretuje się w ten sposób, że charakter ich produkcji jest najbardziej pierwotny dla innych gałęzi gospodarki, a ich cykl koniunkturalny istotnie wyprzedza cykle w innych branżach. Podobne zależności obserwuje się na giełdzie londyńskiej (BG, SHEL, BP).

Centralność spółek w strukturze drzew MST może wynikać z różnych czynników, z których największą rolę odgrywa wielkość firmy i dywersyfikacja obszarów jej zaangażowania (przykładem są tu konglomeraty amerykańskie, takie jak widoczny w centrum rysunku 1.3C koncern General Electric, a także z wielkości sektora, do którego dana spółka należy i jego aktualna kondycja ekonomiczna (spółki Citigroup, C na rynku amerykańskim i Prudential, PRU na rynku angielskim). Z kolei peryferyjny charakter spółek może wiązać się albo z unikalnością branży w obrębie rozpatrywanego zbioru spółek (Thyssen-Krupp, TKA na rynku niemieckim czy Disney, DIS na rynku DJIA), albo z indywidualnym charakterem dynamiki akcji, związanym z jej kondycją ekonomiczną lub innymi zdarzeniami jej dotyczącymi (Optimus, OPT).

**Oznaczenia symboli spółek:**

**Skład DJIA:** AA Alcoa Inc., AIG American International Group, AXP American Express Company, BA Boeing Co., C Citygroup, CAT Caterpillar, CVX Chevron, DD E I Du Pont De Nemours & Company, DIS Walt Disney Co., EK Eastman Kodak, GE General Electric, HD Home Depot Inc., HON Honeywell International Inc., HPQ Hewlett-Packard Co., IBM, INTC Intel Corp, IP International Paper Co., JNJ Johnson & Johnson, JPM JPMorgan Chase, KO Coca-Cola Co., MCD McDonald's, MMM Millennial Media Inc., MO Altria Group Inc., MRK Merck & Co. Inc., MSFT Microsoft Corp, PG Procter & Gamble Co, T AT&T Inc., TRV Travelers Companies Inc., UTX United Technologies Corp., WMT Wal-Mart Stores Inc.

**Skład DAX:** Adidas ADS, Allianz ALV, BASF BASF, Bayer BAYN, Beiersdorf BEI, Manufacturing BMW, Commerzbank CBK, Continental CON, Daimler DAI, Deutsche Bank DBK, Deutsche Börse DB1, Deutsche Lufthansa LHA, Deutsche Post DPW, Deutsche Telekom DTE, E.ON EOAN, Fresenius FRE, Fresenius Medical Care FME, HeidelbergCement HEI, Henkel HEN3, Infineon Technologies IFX, K+S SDF, Lanxess LXS, Linde LIN, Merck MRK, Munich Re MUV2, RWE, SAP, Siemens Industrial SIE, ThyssenKrupp TKA, Volkswagen Group VOW3,

**Skład WIG20:** BIG BIG Bank, BRE Bank Rozwoju Eksportu, BZWBK Bank Zachodni, ECH Echo, ELE Elektrim, EXB Exbud, JLF Jelfa, KBL NKT Cables SA., KBH Kable Holding, KGH KGHM, MXP Mostostal Expo, OCN Ocean, OKM Okocim, OPT Optimus, RFK Rafako, STO Stomil, STX Stalexport Autostrady SA, SWC, WLC Wólczanka, ZWC Zywiec

**Skład FTSE100:** Aberdeen Asset Management ADN, Admiral Group ADM, Aggreko AGK, AMEC AMEC, Anglo American plc AAL, Antofagasta ANTO, ARM Holdings ARM, Associated British Foods ABF, AstraZeneca AZN, Aviva AV, Babcock International BCKIF, BAE Systems BTAS, Barclays BARC, BG Group BG, BHP Billiton BLT, BP BP, British American Tobacco BTI, British Land Co BLND, BSKYB BSY, BT Group BT.A, Bunzl BNZL, Burberry Group BRBY, Capita CPI, Carnival plc CUK, Centrica CPYYY, Coca-Cola HBC AG CCH, Compass Group CPG, CRH plc CRHCY, Croda International CRDA, Diageo DGE, Eurasian Natural Resources EVR, Evraz EVR, Experian EXPN, Fresnillo plc FRES, G4S GFS, GKN GKN, GlaxoSmithKline GSK, Glencore International GLEN, Hammerson HMSO, Hargreaves Lansdown HL, HSBC HSBA, IMI plc IMI, Imperial Tobacco Group ITY, InterContinental Hotels Group IHG, International Consolidated Airlines Group SA IAG, Intertek Group ITRK, ITV plc ITV, J Sainsbury plc SBRY, Johnson Matthey JMPY, Kingfisher plc KNGF, Land Securities Group LAND, Legal & General LGGNY, Lloyds Banking Group LLOY, Marks & Spencer Group MKS, Meggitt MGGT, Melrose plc MRO, Morrison Supermarkets MRW, National Grid plc NGGTF, Next plc NXT, Old Mutual OML, Pearson plc PSORF, Petrofac PFC, Prudential plc PRU, Randgold Resources RRS, Reckitt Benckiser RB, Reed Elsevier REL, Resolution Limited FLG, Rexam REX, Rio Tinto Group RIO, Rolls-Royce Group RR, Royal Bank of Scotland Group RBS, Royal Dutch Shell RDSA, RSA Insurance Group RSA, SABMiller SAB, Sage Group SGE, Schroders SDR, Serco Group SRP, Severn Trent SVT, Shire plc SHPGY, Smith & Nephew SNN, Smiths Group SMIN, SSE plc SSEZY, Standard Chartered STAN, Standard Life SL, Tate & Lyle TATE, Tesco TSCO, TUI Travel TT, Tullow Oil TLW, Unilever ULVR, United Utilities UU, Vedanta Resources VED, Vodafone Group VOD, Weir Group WEIR, Whitbread WTRZ, Wolseley plc WOSYY, Wood Group WG, WPP plc WPPGY, Xstrata XTA

## Rozdział 2. Czas życia korelacji

W definicji współczynnika korelacji (0.3) istotna jest długość okna czasowego, w którym dokonywana jest obserwacja wartości sygnałów. Im dłuższe okno, tym lepsza statystyka, ale jednocześnie większy wpływ odległych historycznie zdarzeń. W związku z tym pojawia się pytanie, czy długość okna czasowego ma wpływ na jakość i stabilność wyników.

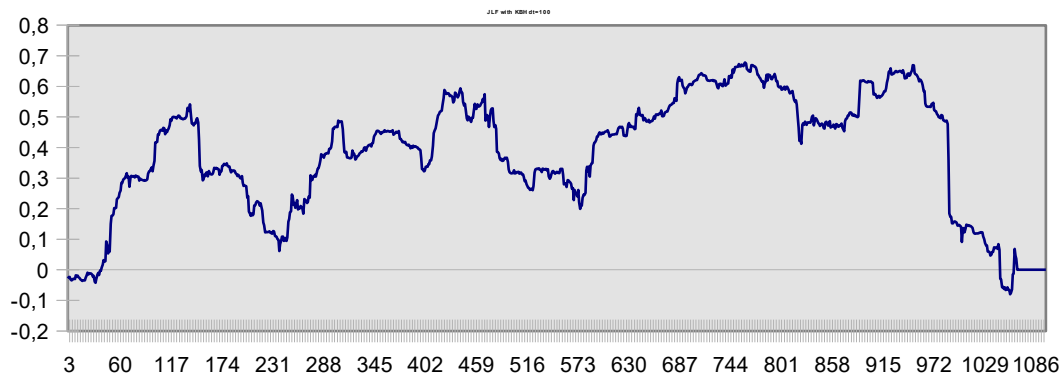
Teorie portfelowe [34, 45] zakładają stacjonarność danych w czasie, co w przypadku danych finansowych nie jest niestety spełnione. Ta niestacjonarność wpływa na stabilność współczynnika korelacji, którego wartość także jest zmienna. Przykład ewolucji współczynnika Pearsona dla spółek JLF (Jelfa) i KBH (Kable) należących do WIG20, obliczonego na podstawie okna czasowego 100 poprzedzających dni transakcyjnych w latach 1997-2004 zaprezentowany jest na Rysunku 2.1. Współczynnik ten ewoluował przyjmując wartości od -0.1 do 0.7. Jak pokazują duże wahania wartości na rysunku, badania nad dynamiką i sensem stosowania w tym kontekście współczynnika korelacji mają dla inwestorów i ekonomistów duże znaczenie.

Przede wszystkim warto zauważyć, że są dwa konkurencyjne spojrzenia na pojęcie współczynnika korelacji w odniesieniu do danych finansowych. Jedno z nich zakłada, że współczynnik korelacji dla danej pary walorów jest niezmienny i ma dokładnie określoną wartość, natomiast jego praktyczna zmienność wynika tylko z efektu estymowania na podstawie skończonej próbki danych historycznych [46]. Drugie spojrzenie, podzielane przez autora, zakłada zmienność warunków zewnętrznych (sytuacja gospodarcza, kondycja firmy itd.), wpływających na wycenę walorów, a zatem także na rzeczywistą wartość współczynnika korelacji między nimi.

W dalszej części tego rozdziału zbadane zostanie zachowanie współczynnika korelacji jako funkcji czasu i rozmiaru okna czasowego  $\Delta t$ :

$$\rho = \rho(t, \Delta t) \quad (2.1)$$

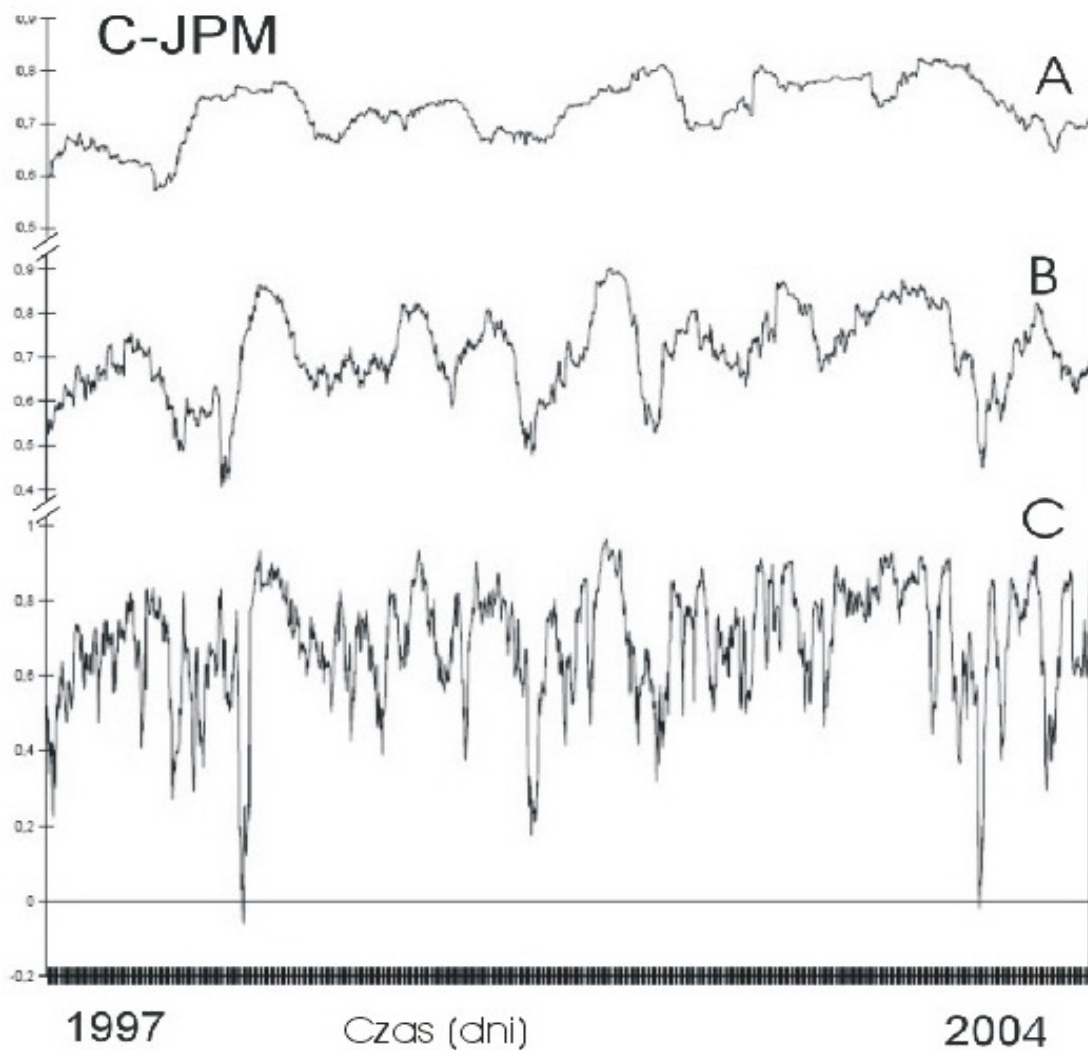
oraz określona zostanie metodologia wyznaczania optymalnej długości tego okna.



**Rys. 2.1** Ewolucja współczynnika korelacji (oś pionowa) w latach 1997-2004 między spółkami JLF i KBH należącymi do WIG20 obliczona dla okna czasowego 100 dni transakcyjnych, przesuwanego wzdłuż osi czasu (oś pozioma).

## 2.1. W poszukiwaniu optymalnego okna czasowego

Rysunek 2.1 pokazuje, że dla danych dwóch sygnałów precyzyjna wartość współczynnika korelacji ewoluje i, co za tym idzie, ma charakter lokalny. Dlatego do obliczonej przykładowo wartości 0,72 współczynnika korelacji między Citygroup (C) i JP Morgan (JPM), tak jak do wszystkich innych wartości, należy podchodzić z dystansem. Na Rysunku 2.2 pokazałem ewolucję współczynnika korelacji dla okien czasowych o różnej długości  $\Delta t$ . Z rysunku wynika, że dla zbyt krótkiego okna wykres ewolucji współczynnika korelacji zbliża się do szumu, bo jest pod wielkim wpływem fluktuacji cen i pokazuje losowość, zamiast odzwierciedlać powiązania gospodarcze. Gdy  $\Delta t$  rośnie, wykres się wygładza i stabilizuje w okolicach swojej średniej wartości równej 0,72. Problem w tym, że ewolucja gospodarki jest procesem naturalnym i ekonomiści podchodzą do sprawy w sposób praktyczny i nie używają zbyt dużego okna czasowego w swoich obliczeniach [47]. Zwłaszcza, że współczynnik korelacji zmienia się w czasie jak średnia ruchoma, a zatem, obciążony zbyt długą historią cen, nie uchwyci aktualnych zmian gospodarczych z najnowszych dni. Istnieje zatem potrzeba, aby mieć narzędzie, które pomoże ustalić optymalne okno czasowe  $\Delta t$ , dla którego obliczony współczynnik korelacji będzie odporny na fluktuacje cen, a jednocześnie wrażliwy na zmiany gospodarcze.



Rys. 2.2 Ewolucja czasowa współczynnika korelacji  $\rho(t, \Delta t)$  (oś pionowa) między akcjami spółek Citigroup (C) i JP Morgan (JPM) dla różnych okien czasowych [15]:  $\Delta t = 200$  dni (A),  $\Delta t = 65$  dni (B),  $\Delta t = 20$  dni (C).

Dla ruchomego okna czasowego  $\Delta t = 200$  dni współczynnik korelacji  $\rho(t)$  między C i JPM zmienia się, ale ciągle pozostaje na poziomie silnym (w sensie klasyfikacji wprowadzonej w Rozdziale 1). Kiedy okno czasowe  $\Delta t$  zmniejsza się z 200 do 20 dni handlowych, wykres coraz bardziej przypomina szum. Fakt ten dowodzi, że rynek nie jest statyczny, ale dynamiczny. Proces zmian cen nie jest stacjonarny, więc korelacje pod wpływem dynamiki rynku zależą od czasu, bo spółki i oddziaływania między nimi zmieniają się. Wówczas  $\rho(t)$  przyjmuje wartości ze wszystkich poziomów korelacji: słabych, silnych i ujemnych. Zachowanie współczynnika korelacji przedstawione powyżej jest typowe dla wszystkich rozważanych przeze mnie walorów i wszystkich rynków finansowych. Również w literaturze ten problem był poruszany, a niestabilność korelacji dotyczy także takiej globalnej miary, jak największa wartość własna macierzy korelacji, silnie zmienna w czasie [12,17].

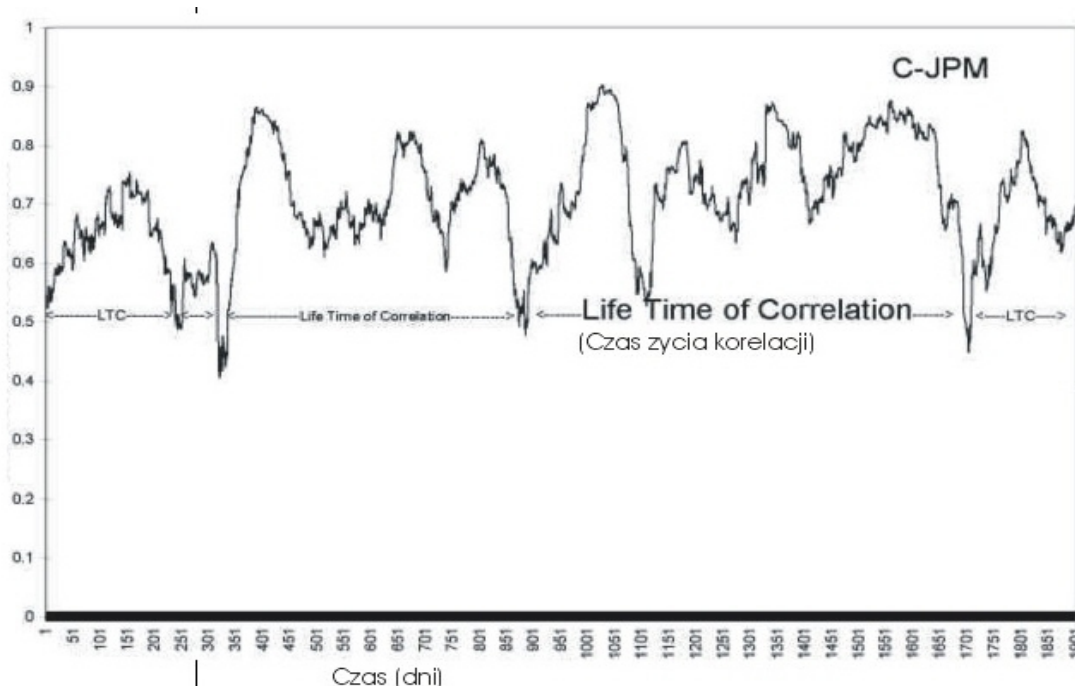


## 2.2 Średni czas życia korelacji

Ponieważ istnieje potrzeba uchwycenia trwałości korelacji, użytecznym narzędziem, które w tym celu wprowadzam, jest  $\tau_{ij}^{(k)}(\Delta t)$  - *czas życia korelacji* (ang. *life-time of correlation*), określony dla pary szeregów czasowych  $i, j$ . Definiuję go jako okres, podczas którego współczynnik korelacji  $\rho(t, \Delta t)$  jest permanentnie na silnym poziomie [48]. Na przykład dla  $\Delta t$  równego 65 dni współczynnik korelacji  $\rho(t)$  między C a JPM osiąga silny poziom 6 razy (Rys. 2.3). Mamy zatem dla tej pary spółek 6 różnych czasów życia korelacji, które można następnie uśrednić, tworząc *średni czas życia korelacji*:

$$\langle \tau_{ij}(\Delta t) \rangle = \frac{1}{K} \sum_{k=1}^K \tau_{ij}^{(k)}(\Delta t) \quad (2.2)$$

gdzie  $k$  numeruje następujące po sobie kolejne czasy życia korelacji ( $k=1, \dots, K$ ) dla ustalonej pary  $i, j$ .



Rys.

2.3 Ewolucja współczynnika korelacji  $\rho(t, \Delta t)$  między spółkami C a JPM uzyskana dla  $\Delta t = 65$  dni. Czas życia korelacji jest zdefiniowany jako czas, w którym współczynnik korelacji jest permanentnie na silnym poziomie [48].

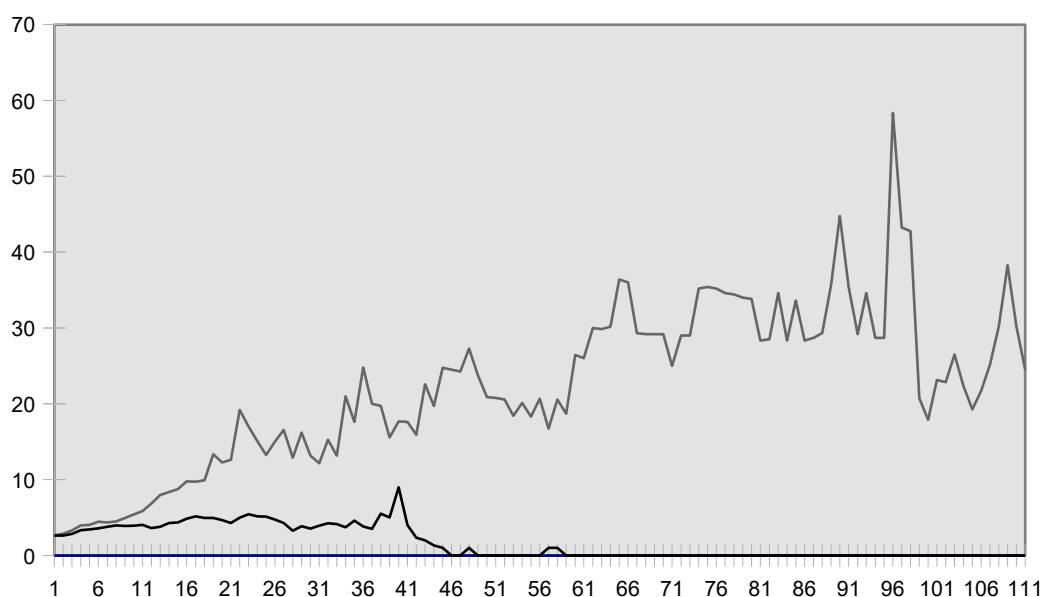
Podobnie jak  $\rho_{ij}(t, \Delta t)$ , średni czas życia korelacji (2.2) także zależy od okna czasowego  $\Delta t$ . Na przykład dla  $\Delta t = 65$  dni średni czas życia korelacji dla pary C-JPM wynosi 233,125 dni handlowych, czyli prawie kalendarzowy rok ( $\approx 250$  dni handlowych).

Generalnie średni czas życia korelacji jest miarą trwałości korelacji w badanym okresie.

Pod tym kątem zbadałem każdą z  $N=435$  par akcji w portfelu DJIA. Dla bardzo małego  $\Delta t$  współczynnik korelacji  $\rho(t, \Delta t)$  opiera się na chwilowych zmianach wartości danych pod wpływem fluktuacji cen. Dla wszystkich par średnie czasy życia korelacji zachowują się tak samo. Ze wzrostem  $\Delta t$  rosną zarówno wartości  $\tau_{ij}^{(k)}(\Delta t)$ , jak i  $\langle \tau_{ij}(\Delta t) \rangle$ , ponieważ wykresy  $\rho(t, \Delta t)$  znane z Rys. 2.2 wygładzają się. Jest to efekt nakładania się zbiorów danych, bowiem dla większych  $\Delta t$  współczynnik korelacji  $\rho(t, \Delta t)$  i  $\rho(t+1, \Delta t)$  w chwili następnej opiera się na prawie identycznych, nakładających się danych. W efekcie dla dużych okien czasowych współczynnik korelacji w kolejnych chwilach praktycznie przestaje się zmieniać.

Dla silnie skorelowanych par jak C-GE z portfela DJIA  $\tau_{ij}^{(k)}(\Delta t)$  rośnie wraz z  $\Delta t$ . Z drugiej jednak strony pary słabo skorelowane jak GM-MO (których współczynnik korelacji wynosi zaledwie 0,12) dla dostatecznie dużego  $\Delta t$  osiągają średni czas życia korelacji  $\langle \tau_{GM-MO}(\Delta t) \rangle$  równy 0. Ma to swoje źródło w tym, że dla odpowiednio dużego okna czasowego wykres ewolucji współczynnika korelacji się wygładza i stabilizuje poniżej progu 0,5, który definiuje silne korelacje. A zatem wszystkie pary, dla których współczynnik korelacji wyliczony dla całego sygnału ma wartość poniżej 0,5, muszą ostatecznie posiadać zerowy średni czas życia korelacji.

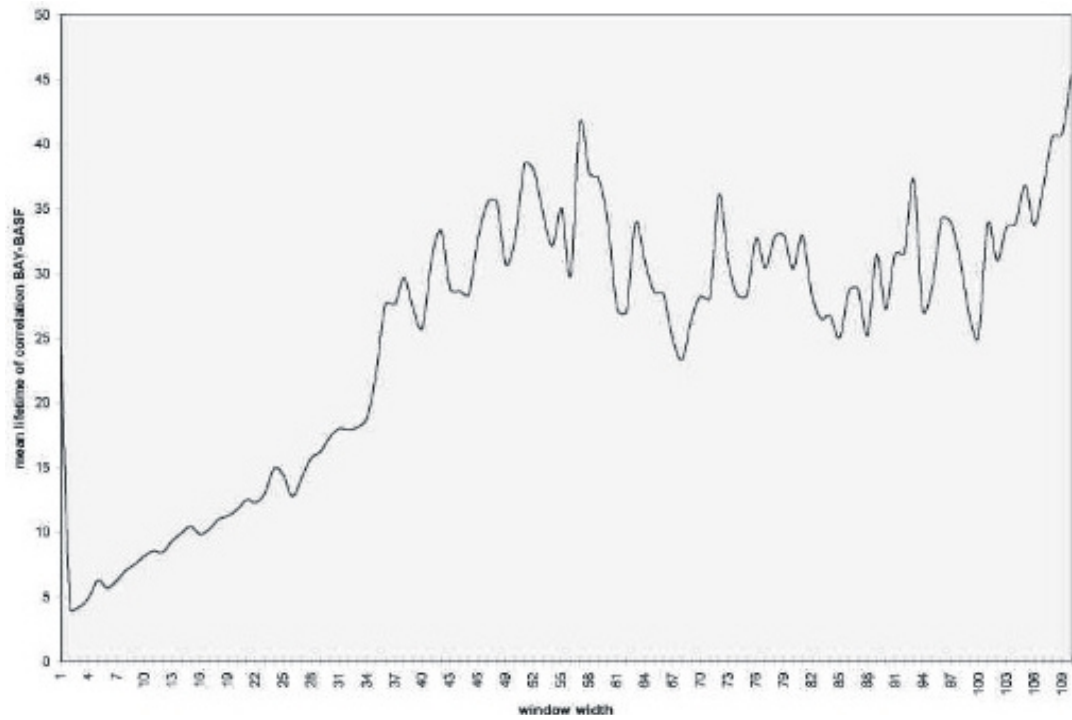
Na rysunku 2.4 pokazuję średni czas życia korelacji dla dwóch różnych par spółek: silnie skorelowanej C-GE, łączącej spółkę czysto finansową z konglomeratem wielobranżowym (w tym także zajmującym się bankowością) i słabo skorelowanej GM-MO, łączącej branżę motoryzacyjną i tytoniową. Widzimy, że dla  $\Delta t > 60$  dni średni czas życia korelacji  $\langle \tau_{GM-MO}(\Delta t) \rangle$  maleje do zera, w przeciwieństwie do średniego czasu życia korelacji  $\langle \tau_{C-GE}(\Delta t) \rangle$ , który nadal rośnie. Z kolei dla zbyt małego okna czasowego  $\Delta t < 10$  dni oba średnie czasy życia korelacji są małe i mają podobną wartość, więc ich poziom jest w praktyce nierozróżnialny.



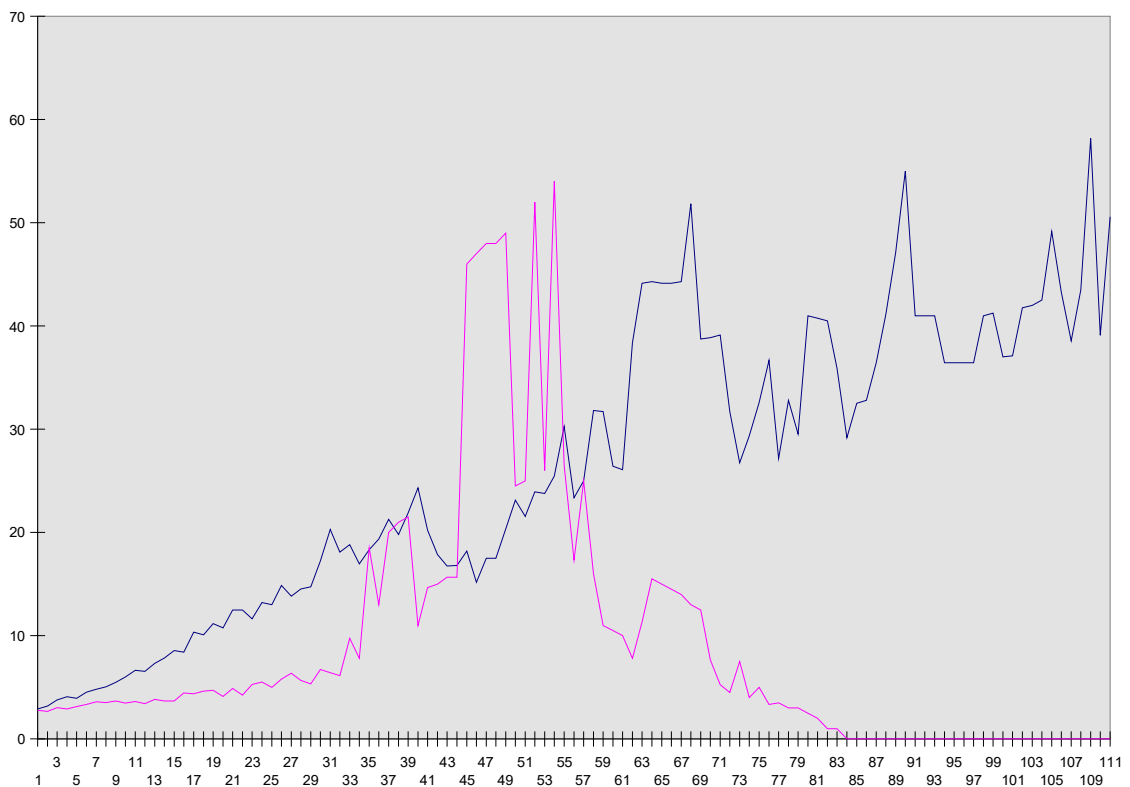
**Rys. 2.4.** Średni czas życia korelacji (oś pionowa) dla silnie skorelowanej pary C-GE (górną linią) oraz słabo skorelowanej pary GM-MO (dolną linią) z portfela DJIA w zależności od długości okna czasowego  $\Delta t$ , wyrażonej w dniach handlowych (oś pozioma) [48].

Dla dużych  $\Delta t$  średni czas życia korelacji ma coraz większe wahania. Źródłem tego zjawiska jest malejąca liczba okresów, w których korelacja jest silna, co powoduje, że pogarsza się statystyka przy uśrednianiu czasów życia korelacji (2.2). Przykładowo:  $\langle \tau_{C-GE}(\Delta t) \rangle$  wynosi 59 dni dla  $\Delta t = 106$  dni i gwałtownie maleje do 38 dni dla  $\Delta t = 108$  dni. Oznacza to, że nawet niewielka zmiana  $\Delta t$  może powodować dramatyczne zmiany  $\tau_{ij}^{(k)}(\Delta t)$ . To zjawisko zaobserwowałem dla par na wszystkich badanych rynkach. Służy temu choćby przykład silnie skorelowanej pary BASF-BAY należącej do DAX (Rys. 2.5). Fakt ten sugeruje, że okno czasowe  $\Delta t$  potrzebne do obliczania współczynnika korelacji między akcjami (1.3) nie może być zbyt duże. Analogiczny wynik przedstawiono w pracy [43], w której badano czas połowicznego rozpadu drzewa minimalnej rozpiętości MST dla portfela akcji DJIA. Czas połowicznego rozpadu drzewa  $t_{1/2}(\Delta t)$  zdefiniowano jako czas, po którym rozpadnie się połowa połączeń między akcjami w tym drzewie. Oczywiście on także zależy od okna czasowego  $\Delta t$ .

Warto zwrócić uwagę na Rys. 2.6, który przedstawia nietypową sytuację, gdy para, która jest słabo skorelowana dla dużych wartości  $\Delta t$ , a silnie – dla średnich. Jej średni czas życia korelacji dla średnich okien może przewyższać czas życia korelacji dla pary,



**Rys. 2.5.** Średni czas życia korelacji (oś pionowa) dla pary BASF-BAY (rynek niemiecki) w zależności od długości okna czasowego  $\Delta t$  (oś pozioma) [48].



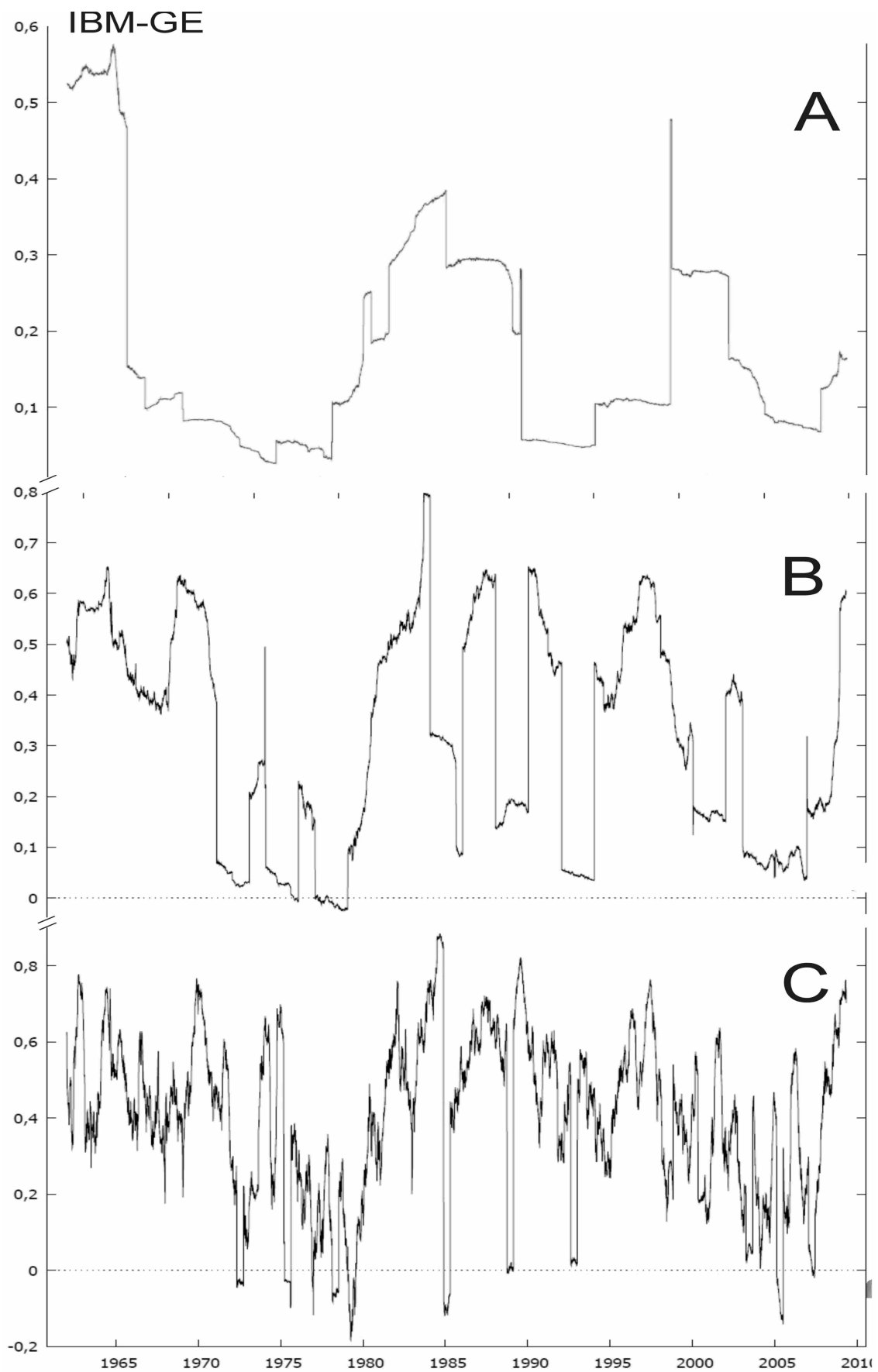
**Rys. 2.6.** Średni czas życia korelacji (oś pionowa) dla pary CHV-SBC (ciemniejsza linia) oraz słabo korelowanej pary CHV-T (jaśniejsza linia) z portfela DJIA w zależności od długości okna czasowego  $\Delta t$  (oś pozioma).

która w długich oknach czasowych jest silniej skorelowana. Na wszystkich badanych rynkach odnotowałem tylko jeden taki przypadek – Chevron (CHV) i AT&T (T). Typowa sytuacja wygląda natomiast tak, jak na Rys. 2.4.

Interesujące jest zbadanie średniego czasu życia korelacji w dłuższej perspektywie czasowej niż kilka lat. Problem w tym, że indeksy DJIA, DAX i WIG 20 z definicji zawierają spółki o największych obrotach, a zatem zmieniają swój skład co parę miesięcy. Zwłaszcza, że rozwój gospodarczy prowadzi do upadku starych firm i wejścia nowych. W tej sytuacji ze wszystkich badanych rynków wybrałem najstarszą dostępną parę – General Electric oraz IBM, które do portfela akcji DJIA należą jako jedyne od ponad kilkudziesięciu lat. Wziąłem pod uwagę dane od 25 czerwca 1959 do 25 czerwca 2011, a więc okres 52 lat.

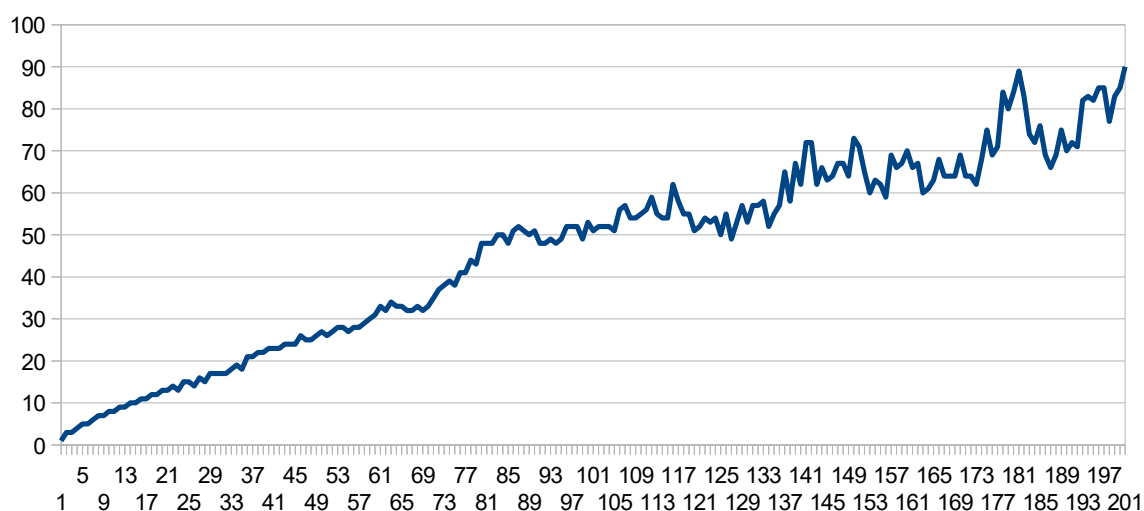
We wcześniejszych badaniach na podstawie historii codziennych cen zamknięcia sesji w okresie od 1.01.1997 do 1.11.2004 – a więc na kilkuletnim oknie czasowym – otrzymałem współczynnik korelacji między GE i IBM równy 0,11. Jednak kiedy sięgnąłem do bardziej odległej historii, okazało się, że spółki te bywały też silnie skorelowane wcześniej – do roku 1964. Później ich korelacja była słaba, ale na ogół utrzymywała się na stałym poziomie i jeśli podlegała zmianom, to były to zmiany skokowe (Rys. 2.7). Najważniejsza z nich – przejście z silnej korelacji do słabej - miała miejsce po 7 kwietnia 1964 r., kiedy IBM pod kierownictwem Tomasza Watsona Jr – syna założyciela firmy - ogłosił wprowadzenie pierwszej w świecie rodziny komputerów IBM System 360. Komputer ten zapoczątkował rodzinę kompatybilnych maszyn i spopularyzował pracę na odległość (terminale podłączone przez linie telefoniczne). Do roku 1968 sprzedano 14 tys. maszyn, a do roku 1978 System 360 znalazł zastosowanie zarówno komercyjne, jak i naukowe, a IBM, który dotąd zajmował się energetyką czy wspieraniem lotów w kosmos, od 1964 roku związał się już ściśle z technologią informatyczną.

Z kolei dla General Electric, założonej w 1892 roku przez Tomasza Edisona, technologia to zaledwie jedna z kilku dziedzin. Pozostałe, którymi GE się zajmuje, to energetyka, rynki kapitałowe, konsumenckie czy przemysłowe, dzięki czemu spółka ta zawsze znajduje się w centrum drzewa minimalnej rozpiętości MST (Rys. 1.3C) dla indeksu DJIA, w którym oryginalnie jest od chwili jego powstania, czyli od 1892 roku, gdy obejmował on jedynie 12 spółek.



**Rys. 2.8.** Ewolucja czasowa współczynnika korelacji  $\rho(t, \Delta t)$  między IBM a GE dla różnych okien czasowych:  $\Delta t = 2000$  dni (A),  $\Delta t = 500$  dni (B) i  $\Delta t = 105$  dni (C).]

Porównanie okien czasowych, dla których wyznaczyłem ewolucję współczynnika korelacji między GE a IBM (Rys. 2.7 A,B,C) pokazuje, że dla ogromnych okien czasowych  $\Delta t = 2000$  dni, czyli rzędu 6 lat, obserwujemy mniej skoków, a zatem na wartość współczynnika korelacji wpływają jedynie zdarzenia o długofalowych skutkach. Co prawda podany jako przykład IBM System 360 od wielu lat nie jest już używany (obecnie istnieją jedynie niesprawne urządzenia), to sam kierunek rozwoju spółki IBM poszedł w branżowo trochę inną stronę niż GE. Jednak dla mniejszych okien czasowych, na przykład  $\Delta t = 500$  dni, skoków obserwujemy więcej, a zatem zaszły kolejne zdarzenia wpływające na korelację GE-IBM, ale o mniejszej wadze ekonomicznej. Dopiero dla okna czasowego rzędu  $\Delta t = 105$  dni, a więc 5 miesięcy, obserwujemy wpływ szumu we fluktuacjach cen. Widzimy jednak, że jest on znacznie mniejszy niż skoki samego współczynnika korelacji (sytuację dla podobnych okien czasowych jak na Rys. 2.7C przedstawiłem już wcześniej na Rys. 2.2 dla pary C-JPM, ale dla węższego wycinka lat 1997-2004).



**Rys. 2.9.** Średni czas życia korelacji (oś pionowa) dla GE-IBM, najstarszej pary z portfela DJIA, wyznaczony na podstawie danych z lat 1959-2011 (oś pozioma oznacza długość okna czasowego  $\Delta t$ ).

Widzimy zatem, że ewolucja współczynnika korelacji dla dużych okien czasowych przebiega skokowo, w zależności od wagi ekonomicznej wydarzeń. Ma to oczywiście wpływ na średni czas życia korelacji tej pary spółek, który wraz z oknem czasowym  $\Delta t$  nie zanika (Rys. 2.8). Został on bowiem obliczony i uśredniony na podstawie historii cen obejmującej okres, gdy korelacje między GE a IBM były zarówno słabe, jak i silne. Możliwe jest także wprowadzenie średniego czasu życia korelacji zdefiniowanego dla grupy szeregów czasowych poprzez uśrednienie średnich czasów życia korelacji dla poszczególnych par sygnałów. Dla rynku akcji uśrednianie takie można przeprowadzić np. po wszystkich parach spółek w obrębie danego indeksu. Wielkość tę nazywam

indeksowym czasem życia korelacji i definiuję wzorem:

$$\langle \tau(\Delta t) \rangle_{IND} = \frac{1}{N} \sum_{ij} \langle \tau_{ij}(\Delta t) \rangle \quad (2.3)$$

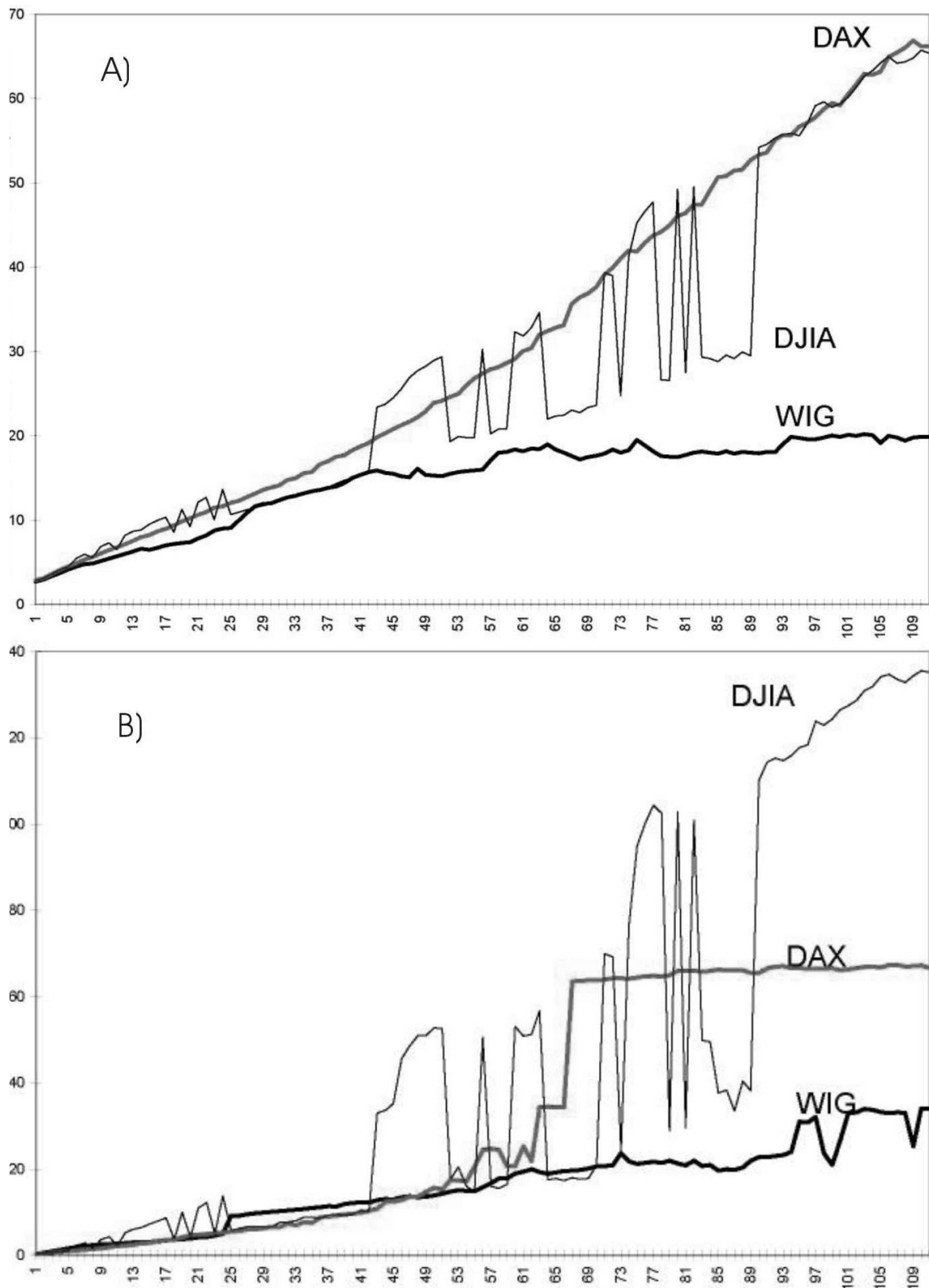
gdzie  $N$  jest liczbą par akcji należących do portfela WIG20, DAX czy DJIA, natomiast  $ij$  oznaczają spółki ( $i \neq j$ ). Taką definicję wprowadzam po to, by porównać indeksowe czasy życia korelacji w różnych portfelach akcji oraz wykryć zróżnicowanie korelacji na danym rynku za pomocą odchylenia od indeksowego czasu życia (Rys. 2.10). Portfele DJIA i DAX zawierają silnie skorelowane pary akcji, co powoduje, że indeksowy czas życia korelacji rośnie wraz z  $\Delta t$ . Mimo, że indeksowy czas życia korelacji dla obu portfeli ma podobną wartość, wahania jego wartości w obrębie portfela DJIA są znacznie większe. Jak wcześniej wspomniałem, jedynie kilka par z portfela DJIA jest silnie skorelowana, dlatego ich średni czas życia korelacji rośnie, podczas gdy dla pozostałych par maleje do zera. Wówczas jedynie silnie skorelowane pary dają wkład do indeksowego czasu życia korelacji DJIA, co osłabia statystykę. Dlatego odchylenia od indeksowego czasu życia korelacji w portfelu DJIA powstają pod wpływem tych kilku silnie skorelowanych par akcji – w przeciwieństwie do niemieckiego indeksu DAX, gdzie większość par akcji jest silnie skorelowana.

Na obu dojrzałych rynkach indeksowy czas życia korelacji dla akcji tworzących indeksy DJIA i DAX rośnie osiągając 54 dni dla  $\Delta t = 90$  dni. Z drugiej strony na rozwijającym się rynku polskim indeksowy czas życia korelacji dla WIG20 nie przekracza 20 dni handlowych. Polski rynek jest względnie płytki i dlatego niewielki nawet napływ kapitału może silnie destabilizować korelacje między akcjami. Widzimy zatem, że średni czas życia korelacji między akcjami zachowuje się inaczej na rynkach dojrzałych, a inaczej na rynkach rozwijających się (Rys. 2.10). Porównanie indeksowego czasu życia korelacji z czasami średnimi dla poszczególnych par spółek stanowić może pomoc w wyodrębnieniu tych spółek, których korelacje są znacznie trwalsze niż typowo na rynku. To mogłoby ułatwić inwestorom wybór składników portfela akcji.

Z praktycznego punktu widzenia warto przeanalizować i wykorzystać optymalne okno czasowe do wykrycia, dla jakiego  $\Delta t$  zachodzi zróżnicowanie korelacji na każdym z trzech rynków. Dla krótkich okien czasowych odchylenie standardowe od indeksowego czasu życia korelacji rośnie z  $\Delta t$  bardzo powoli. Co więcej, dla  $\Delta t = 2, 3, 4, 5$  dni odchylenie standardowe na wszystkich rynkach jest mniejsze niż 1 dzień handlowy. Taka sytuacja



powoduje, że niemożliwe jest rozróżnienie wartości korelacji między akcjami na podstawie ledwie kilkudniowej historii cen zamknięcia sesji giełdowej.



**Rys. 2.10.** A) Indeksowy czas życia korelacji dla akcji tworzących indeksy DJIA, DAX i WIG 20. B) Odchylenie standardowe od średnich czasów życia korelacji dla par akcji tworzących indeksy DJIA, DAX i WIG 20. Oś pozioma na obu wykresach oznacza okno czasowe  $\Delta t$  [48].

Podobnie jak dla portfeli DJIA i DAX, dla WIG20 odchylenie standardowe od indeksowego czasu życia korelacji rośnie wraz z  $\Delta t$ , ale wzrost ten jest słabszy, gdy  $\Delta t > 60$  dni, bo w w portfelu akcji WIG20 znajduje się tylko jedna silnie skorelowana para (JLF-KBH), a wszystkie pozostałe są słabo skorelowane. Zatem brak jest wyraźnego różnicowania korelacji i odchylenie standardowe od indeksowego czasu życia korelacji dla akcji z WIG20 słabo rośnie z długością okna czasowego. Ciekawa jest analiza odchylenia standardowego średnich czasów życia korelacji dla akcji wchodzących w skład indeksu DAX. Okazuje się, że dla  $\Delta t > 65$  dni odchylenie standardowe przestaje rosnąć i jest praktycznie stałe. Oznacza to, że okno czasowe  $\Delta t = 66$  dni jest wystarczające i optymalne do wykrycia różnicowania korelacji między parami akcji na rynku niemieckim.

Powyższe wyniki uzyskane dla dużego zespołu danych z rynków dojrzałych i wschodzącego wskazują, że stosowane w praktyce przez ekonomistów okno czasowe o długości  $\Delta t$  rzędu kilku miesięcy jest uzasadnione [47].

Zjawisko zaniku korelacji między fluktuacjami cen na rynkach akcji dla krótkich okien czasowych zostało odkryte przez Tomasza Eppsa [49] i nosi nazwę efektu jego imienia. Efekt ten wynika z niezerowego czasu potrzebnego na dotarcie informacji do wszystkich lub przynajmniej większości aktywnych inwestorów. W notowaniach cen objawia się to asynchronicznością ich zmian na bardzo krótkich skalach czasowych, gdzie dominuje szum, oraz bardziej synchronicznymi ruchami cen na dłuższych skalach czasowych, gdy informacja zostanie już rozproszona na całym rynku. W związku z tym w celu rzetelnego przybliżenia wartości współczynnika korelacji, szerokość poruszającego się okna czasowego  $\Delta t$  powinna być dużo większa od skali asynchroniczności. (Możliwe jest także oszacowanie korelacji bez stosowania dużych okien  $\Delta t$  metodą dekompozycji korelacji danych skumulowanych [46]; metoda ta nie jest jednak wiarygodna dla dłuższych skal czasowych, które interesują mnie w tej pracy.)

## Rozdział 3. Korelacje między walutami

### 3.1 Rynek walutowy, kursy wymiany i korelacje.

W poprzednich rozdziałach przeanalizowałem dynamikę i strukturę korelacji na rynkach giełdowych na podstawie historii cen. Warto w tym miejscu podkreślić, iż wartość spółki na giełdzie mierzona jest w walucie danego kraju – stąd spółki tworzące portfel DJIA mają ceny w dolarach (USD), spółki tworzące DAX w euro (EUR), a tworzące WIG20 mają ceny podane w polskich złotych (PLN). Dlatego właśnie analizowałem każdy z rynków osobno, bo wartości cen akcji na każdym z rynków są podane w różnych walutach i mimo stosowania logarytmicznej stopy zwrotu (1.2) nie można lekceważyć takich wskaźników jak inflacja, która ma wpływ na ceny akcji i walut, które są bazowe dla cen spółek notowanych na giełdzie. Dlatego zdecydowałem się zbadać także korelacje między walutami.

W przeciwieństwie do notowań akcji na rynkach giełdowych, czy notowań na rynkach towarowych, pojęcie „absolutnej wartości” waluty jest o wiele bardziej skomplikowane, ponieważ nie istnieje obiektywny układ odniesienia, w którym wycena danej waluty byłaby możliwa w równie naturalny sposób, jak ma to miejsce na innych rynkach. Inną różnicą w stosunku do rynków akcji czy rynków towarowych jest całkowita globalizacja rynku walutowego, który stanowi jedną całość bez względu na geografę, w związku z czym dni handlowe na tym rynku są takie same dla wszystkich walut świata.

Rynek walutowy (Forex) leży w obszarze zainteresowania ekonofizyki od samego początku, choć liczba prac mu poświęconych jest wyraźnie mniejsza od liczby prac dedykowanych rynkom akcji. Do najciekawszych wyników zawartych w literaturze zaliczyć należy odkrycie skalowania potęgowego fluktuacji logarytmicznych kursów walutowych [50, 51] i ich modułów [52], multifraktalnej struktury tych fluktuacji [53-54] oraz hierarchicznej ich struktury [55-56], która zmienia się w czasie na skutek wydarzeń na rynkach [55, 12].

W moich badaniach wziąłem pod uwagę  $N=38$  walut z różnych krajów świata, tj. takich, które są całkowicie wymienne i których wartość jest określana swobodnie przez rynek.

W ten sposób obok wykluczenia z analizy walut zupełnie egzotycznych, takich jak np. waluty Wenezueli czy Tanzanii, pominąłem także niektóre istotne waluty, których wycena nie jest całkowicie rynkowa (m.in. w wyniku nałożonych parytetów) – np. dolar singapurski czy duńska korona. Dane, które analizowałem, stanowią dzienne notowania kursów wymiany wszystkich par walut z analizowanego koszyka, czyli łącznie 1406 par, i obejmują okres 1999-2012.

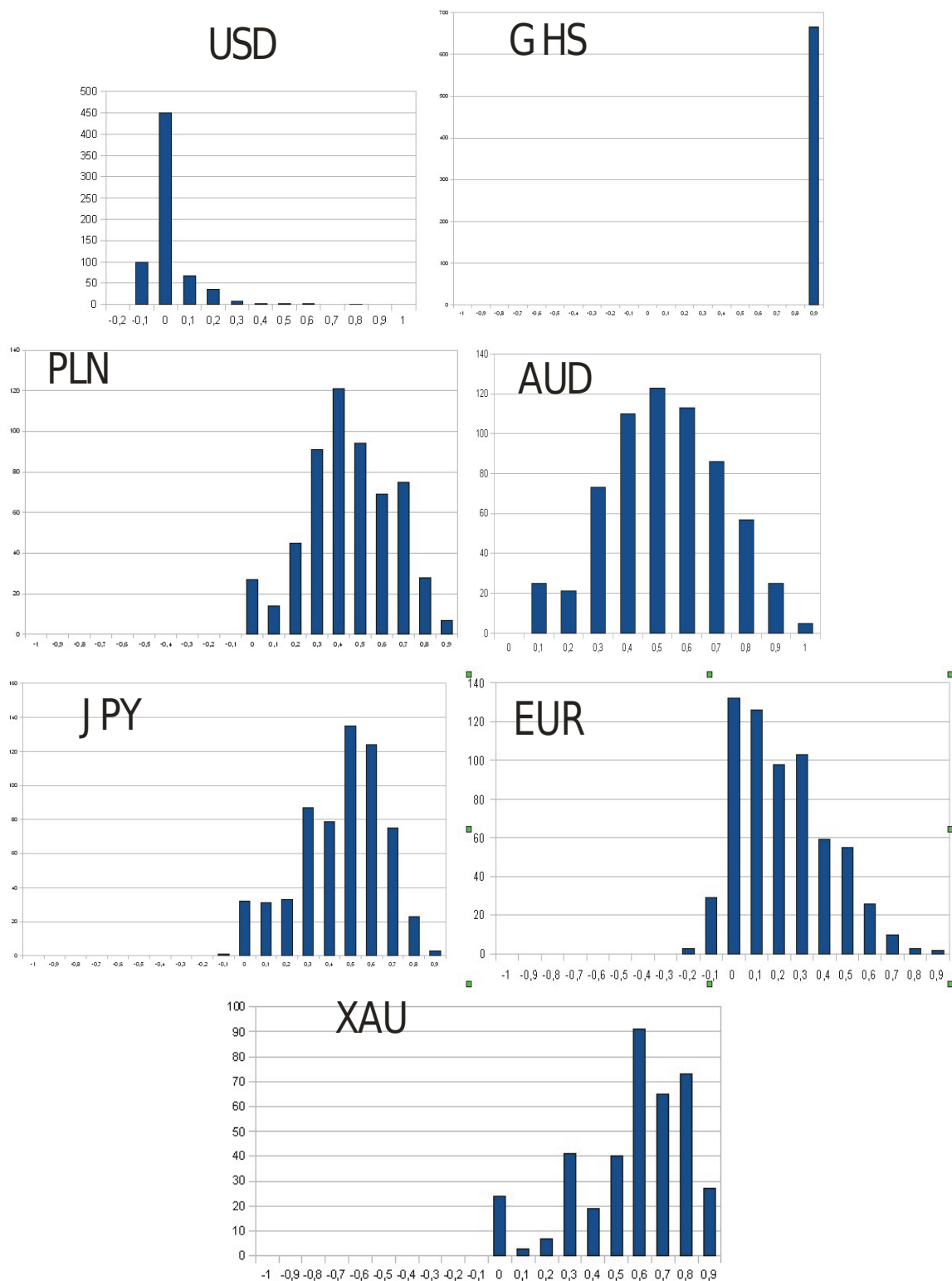
Kurs wymiany dwóch walut zapisuje się w formie ułamkowej, np. EUR/USD, co wyraża cenę 1 euro w dolarach amerykańskich. W ten sposób licznik ułamka  $X/Y$  definiuje *walutę bazową*  $X$ . Jeśli weźmie się pod uwagę tylko te kursy wymiany, które mają tę samą walutę bazową, dla wszystkich pozostałych  $N-1$  walut otrzymuje się wygodny układ odniesienia, dzięki któremu można porównywać ich zachowanie. Oczywiście, jeśli wprowadzi się inną walutę bazową, powstanie inny układ odniesienia, wobec którego waluty będą mogły mieć inne zachowanie.

Rozpatrywanie kursów wymiany walut definiuje układ mający  $N(N-1)$  stopni swobody. Analiza takiego układu byłaby uciążliwa i pochłaniałaby dużo mocy obliczeniowej. Na szczęście na stopnie swobody tego układu nałożone są więzy w postaci dwóch związków:

$$X/Y = (Y/X)^{-1} \quad (3.1)$$

$$X/Y * Y/Z = X/Z \quad , \quad (3.2)$$

które pozwalają na efektywną eliminację większości stopni swobody. Jeśli ponadto dokona się wyboru waluty bazowej, to liczba rozpatrywanych stopni swobody zredukuje się do  $N-1$ . Relacja (3.2) nosi nazwę zależności trójkąta i sprawia, że niemożliwy jest tzw. arbitraż trójkątny, w którym można byłoby osiągnąć zysk poprzez cykliczne wymiany walut objętych tym związkiem. (W praktyce taki arbitraż jest obecnie możliwy jedynie na niezwykle krótkich skalach czasu rzędu ułamków sekund, na których zachodzą transakcje automatyczne, niwelujące wszystkie różnice kursowe.) Rynek walutowy z tego powodu cechuje wysoka efektywność w rozumieniu hipotezy efektywnego rynku [57].



**Rys. 3.1.** Rozkład współczynników korelacji dla 7 wybranych walut bazowych (USD, GHS, PLN, AUD, JPY, EUR, XAU), tworzących kursy wymiany ze wszystkimi pozostałymi  $N-1$  walutami. Okres notowań dziennych obejmuje lata 1999-2012.

Logarytmiczne stopy zwrotu dla kursów walutowych są zdefiniowane, w analogii do wzoru (1.2), przez:

$$R_{XY}(t) = \ln [X/Y(t)] - \ln [X/Y(t-1)] \quad . \quad (3.3)$$

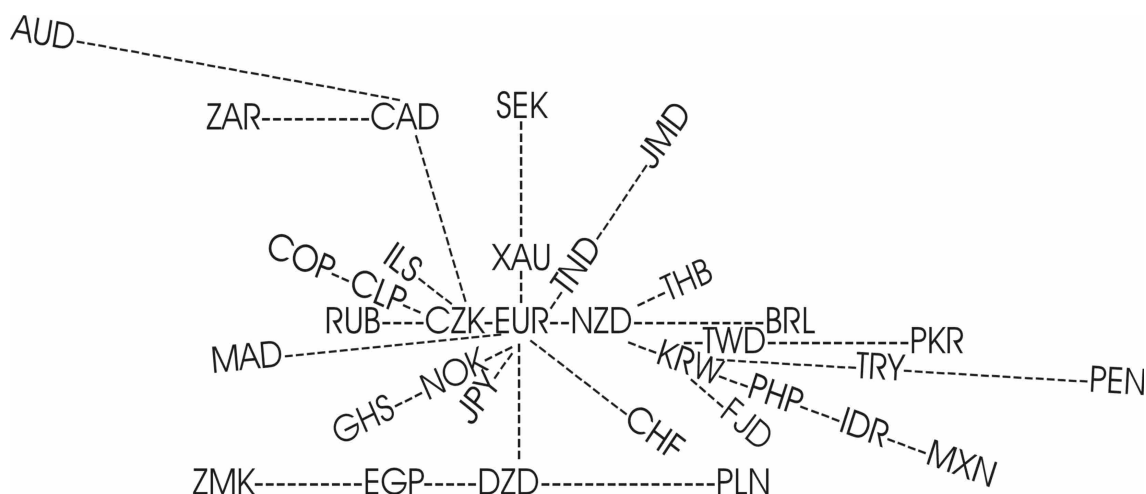
W ten sposób zależności (3.1)-(3.2) prowadzą do następujących relacji pomiędzy stopami zwrotu:

$$R_{XY}(t) + R_{YX}(t) = 0 \quad (3.4)$$

$$R_{XY}(t) + R_{YZ}(t) + R_{ZX}(t) = 0 \quad . \quad (3.5)$$

Szeregi czasowe stóp zwrotu (3.3) są podstawą do wyliczenia współczynników korelacji Pearsona pomiędzy parami walut dla ustalonej waluty bazowej. Rozkłady wartości współczynników korelacji  $\rho_Z(X,Y)$  dla przykładowych walut bazowych Z przedstawione zostały na Rys. 3.1. Rozkłady te charakteryzują się silną zależnością od wyboru Z (Rys. 3.1). Na przykład jeśli walutą bazową był dolar amerykański (USD), to rozkład korelacji między walutami jest asymetryczny i dominują w nim słabe korelacje, skoncentrowane w okolicach 0, w przeciwieństwie do rozkładu dla waluty AUD, który jest skoncentrowany w okolicach 0.5 i przypomina kształtem rozkład normalny (choć w istocie nim nie jest). Przyczyny tych różnic staną się jasne w dalszej części rozdziału.

### 3.2 Struktura drzewa minimalnej rozpiętości (MST) dla walut.



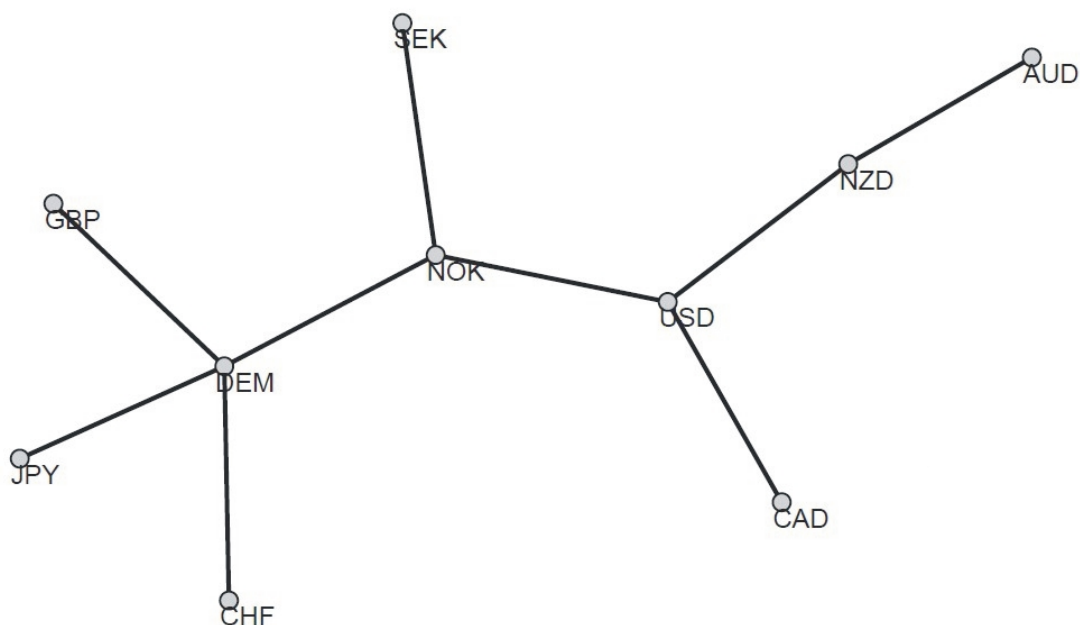
**Rys. 3.2** Drzewo minimalnej rozpiętości MST dla 37 kursów wymiany USD do pozostałych walut. Długość krawędzi jest związana z odległością metryczną (wzór (1.5)).

Względny charakter wyceny waluty powoduje, że wybór waluty bazowej ma istotny wpływ na strukturę rynku. Jak pokazałem w poprzednich rozdziałach, stworzone za pomocą algorytmu Kruskala drzewo MST łączy sklasteryzowane grupy walorów i na ogół nie wykrywa antykorelacji (ujemnie skorelowanych par). Obecność ustalonej waluty bazowej ułatwia interpretację struktury drzewa MST, ponieważ eliminuje dowolność doboru kursów wymiany w obrębie danej pary walut (np. GBP/CHF vs. CHF/GBP). Dowolność taka mogłaby skutkować wprowadzeniem antykorelacji zamiast korelacji dodatnich (z uwagi na wzór 3.4) i efektywnym pominięciem jakiejś silnie skorelowanej pary w strukturze drzewa tylko dlatego, że rozpatrywany był nieodpowiedni kierunek wymiany.

Drzewo minimalnej rozpiętości dla dolara amerykańskiego jako waluty bazowej przedstawiłem na Rys. 3.2. W centrum drzewa znajduje się euro (EUR), które cechuje największa krotność ( $k=9$ ), natomiast większość pozostałych walut jest od niego znacznie oddalona i ma charakter węzłów peryferyjnych. Wyjątek stanowią waluty NZD, CZK i KRW, których krotności wskazują na ich znaczenie jako lokalnych centrów sieciowych. Walutą najbardziej odległą od pozostałych jest dolar australijski (AUD), który z większością walut skorelowany jest słabo ujemnie, a jego najbliższym sąsiadem jest dolar kanadyjski (CAD), z którym skorelowany jest na poziomie 0,003. Tak duża niezależność AUD jest na tyle intrygująca, że użyłem tej waluty jako bazowej, licząc kolejne drzewo MST (Rys. 3.3). Innymi walutami silnie peryferyjnymi wobec USD są polski złoty (PLN) i nowy sol peruwiański (PEN). Na rysunku 3.2 można, mimo słabych korelacji, wyróżnić grona węzłów odpowiadających walutom z krajów zbliżonych geograficznie: (ZMK, EGP, DZD), (KRW, PHP, IDR, TWD, PKR), (EUR, CZK, RUB, NOK).

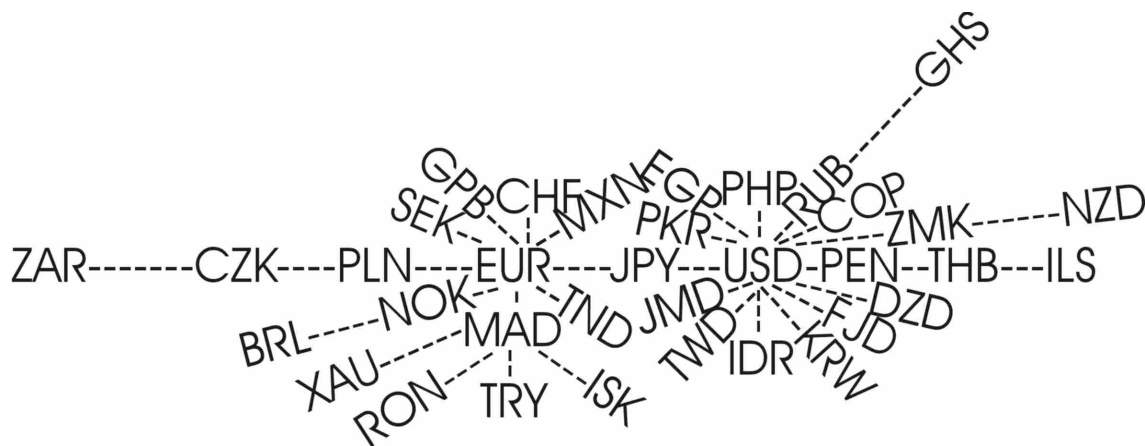
O wiele lepiej geograficzny rozkład walut odzwierciedla Rys. 3.3. Tutaj widoczne są dwa silne centra w postaci USD ( $k=14$ ) i EUR ( $k=9$ ) oraz lokalne centrum w dinarze marokańskim (MAD,  $k=5$ ). Pozostałe waluty są przyłączone do głównych centrów ub tworzą łańcuchy ze swoimi najbliższymi sąsiadami. Całkowicie odmienna struktura tego drzewa od drzewa na Rys. 3.2 i wyłonienie się dużego klastra walut skupionych wokół USD świadczy o ogromnym wpływie dolara amerykańskiego na gospodarkę światową. Z tego powodu waluty, dla których punktem odniesienia w ich ewolucji jest USD i które tworzą wraz z nim klastr na Rys. 3.3, przy braku uwzględnienia tej waluty w analizie poprzez potraktowanie jej jako waluty bazowej, zostają przyłączone do innych walut w

sposób albo całkiem przypadkowy, jak ma to miejsce w przypadku PEN, albo tworzą mniejsze klastry z walutami geograficznie zbliżonymi, co jest podyktowane np. lokalnymi związkami gospodarczymi (słabszymi niż ich związek ze Stanami Zjednoczonymi).



Rys. 3.3 Drzewo minimalnej rozpiętości MST dla 17 walut względem złota jako waluty bazowej [55]

Euro, które było centrum na Rys. 3.2, nadal pozostaje centrum na Rys. 3.3, co wynika z siły tej waluty na rynku międzynarodowym. Po usunięciu USD to euro staje się centralną walutą świata, co jest zgodne z intuicyjnym postrzeganiem tej waluty jako nr 2 na rynku. Zanim euro osiągnęło tę pozycję, przed 1999 r., istniało wiele walut europejskich, które potem weszły w jego skład. Ślad EUR w postaci marki niemieckiej DEM można zaobserwować w danych sprzed 1999 roku na Rys. 3.4, który pochodzi z pracy [55]. Marka niemiecka miała wówczas centralną rolę w gospodarce europejskiej, którą później przejęło EUR [58].



Rys. 3.4 Drzewo minimalnej rozpiętości MST dla 37 walut względem waluty bazowej AUD



Podobnie jak na rynku giełdowym, drzewa minimalnej rozpiętości dla walut zmieniają się w czasie wraz z ewolucją gospodarczą. W poprzednim rozdziale pokazałem, że miarą trwałości korelacji może być czas życia korelacji lub czas połowicznego rozpadu drzewa MST, a rachunki czasu życia korelacji wykonałem dla wszystkich kilkuset par na trzech badanych rynkach. Ze względu na względność wyceny walut, w przypadku każdej z 38 walut bazowych odpowiednia macierz korelacji zawiera  $(N-1)(N-2)/2$  różnych elementów będących współczynnikami korelacji. W celu określenia stabilności struktury korelacyjnej rynku walutowego, analizuję stabilność drzew MST za pomocą zdefiniowanych w pracy [43] współczynników przetrwania. Wyrażają one liczbę połączeń, które przetrwały kolejny krok w stosunku do liczby wszystkich połączeń w drzewie MST. Współczynnik można zdefiniować restrykcyjnie, jako tzw. *średni wielokrokowy współczynnik przetrwania*:

$$\Sigma^B(\delta) = \frac{1}{n_w} \sum_{i=1}^{n_w} \frac{\#\{E^B(i) \cap E^B(i+1) \cap \dots \cap E^B(i+\delta)\}}{N-1} \quad (3.6)$$

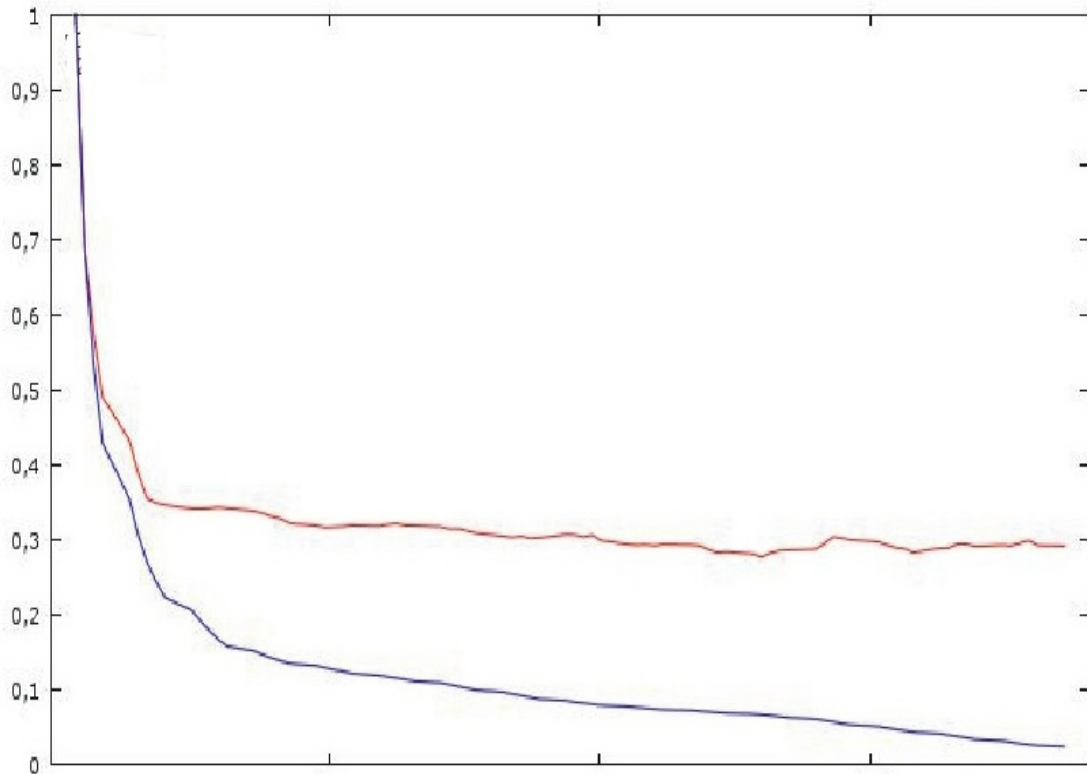
czyli poprzez usunięcie dla danego położenia okna czasowego  $i$  wszystkich połączeń, które znikają w kolejnych  $\delta$  krokach i porównanie liczby tych zachowanych połączeń do ich całkowitej liczby  $N-1$  w początkowym oknie. Współczynnik można zdefiniować również nierestykcyjnie, jako tzw. *jednokrokowy współczynnik przetrwania*:

$$\sigma^B(\delta) = \frac{1}{n_w} \sum_{i=1}^{n_w} \frac{\#\{E^B(i) \cap E^B(i+\delta)\}}{N-1} \quad (3.7)$$

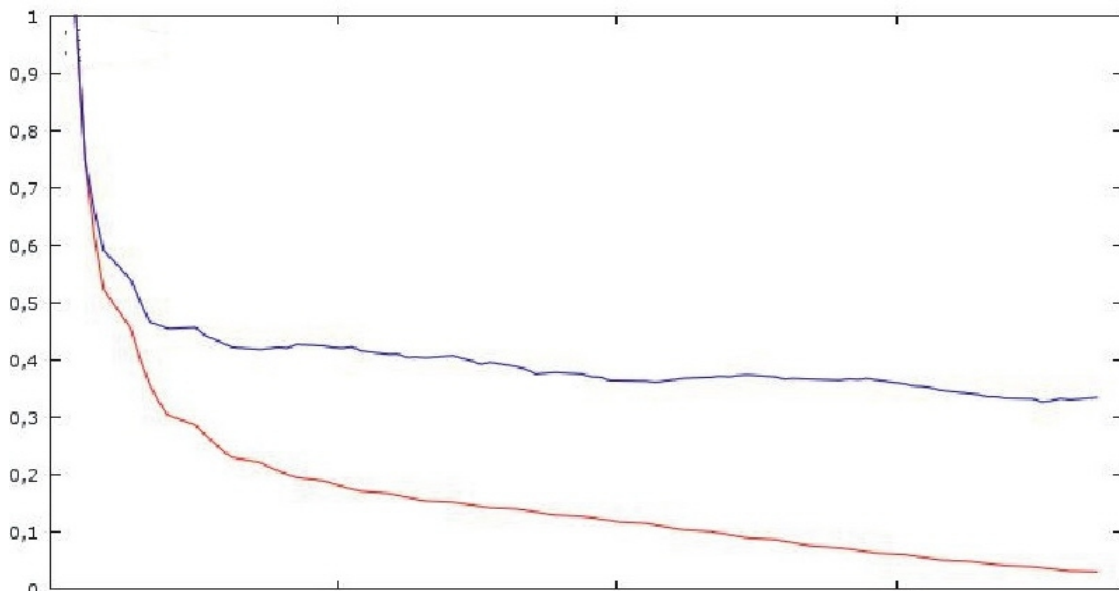
czyli poprzez porównanie liczby zachowanych połączeń w oknie oddalonym o  $\delta$  kroków od początkowego położenia okna  $i$  do pierwotnej liczby połączeń  $N-1$ . W obu wzorach  $E^B(i)$  oznacza zbiór połączeń drzewa MST w oknie  $i$ , natomiast  $n_w$  oznacza liczbę wszystkich położenia okna początkowego, po których dokonujemy uśrednienia. W ten sposób otrzymamy „korytarz” określający mniej więcej, ile połączeń w drzewie przetrwa kolejne kroki.

Obliczenia wykonano dla drzewa określonego w oknie czasowym 100 dni. Wyniki przedstawione zostały na Rys. 3.5 i 3.6 dla dwóch wybranych walut bazowych: USD i AUD. Zanik połączeń w obu przypadkach ma podobną strukturę, natomiast drzewo dla AUD jest statystycznie stabilniejsze. Wynik ten interpretuję jako efekt niezależności dynamiki prezentowanej przez dolara australijskiego (Rys. 3.1D). Niezależność ta jest spowodowana oparciem gospodarki australijskiej na wydobyciu i handlu surowcami

naturalnymi, co zbliża ją do gospodarek takich państw, jak RPA czy Kanada, a czyni nieco odleglejszą od gospodarki europejskiej czy gospodarki USA.



**Rys. 3.5** Średnie współczynniki przetrwania: jednokrokowy  $\sigma^B(\delta)$  (krzywa górna) oraz wielokrokowy  $\Sigma^B(\delta)$  (krzywa dolna) w funkcji czasu  $\delta$  wyrażonego przez kolejne położenia okna czasowego. Wartości współczynników oznaczają ułamek liczby krawędzi drzewa MST, które przetrwały w kolejnych krokach. Walutą bazową jest USD.



**Rys. 3.6** Te same wielkości co na Rys. 3.5, ale obliczone dla waluty bazowej AUD.

Wyniki dla wybranych walut bazowych USD i AUD uzyskane w oparciu o drzewa MST dla danych z lat 1999-2012 przedstawiłem na Rys. 3.5 i 3.6. Okazało się, że dolar

australijski jako waluta bazowa w porównaniu z dolarem amerykańskim daje większą stabilność drzewa MST, prawdopodobnie dlatego, że sam jako waluta najbardziej niezależna od pozostałych podlega mniejszym wahanom. Operując konkretnymi walutami i porównując odpowiednie wyniki z rezultatami wcześniejszej analizy przedstawionej w pracy [55], otrzymałem mniej stabilne drzewa, ale rozpatrywałem późniejszy okres czasu i – w przeciwieństwie do [55] – rozpatrywałem konkretne waluty bazowe, a nie średnią po wszystkich walutach.

**Wykaz badanych walut:**

Dolar australijski (AUD), Dolar kanadyjski (CAD), Frank szwajcarski (CHF), Peso chilijskie (CLP), Peso kolumbijskie (COP), Korona czeska (CZK), Dinar algierski (DZD), Funt egipski (EGP), Euro (EUR), Dolar Fidżi (FJD), Cedi ghańskie (GHS), Funt brytyjski (GBP), Rupia indonezyjska (IDR), Szekeł izraelski (ILS), Korona islandzka (ISK), Dolar jamajski (JMD), Jen japoński (JPY), Won południowokoreański (KRW), Dirham marokański (MAD), Peso meksykańskie (MXP), Korona norweska (NOK), Dolar nowozelandzki (NZD), Nowy sol peruwiański (PEN), Peso filipińskie (PHP), Rupia pakistańska (PKR), Złoty polski (PLN), Nowa leja rumuńska (RON), Rubel rosyjski (RUB), Korona szwedzka (SEK), Baht tajski (THB), Dinar tunezyjski (TND), Nowa lira turecka (TRY), Nowy dolar tajwański (TWD), Dolar amerykański (USD), Złoto (XAU), Rand południowoafrykański (ZAR), Kwacha zambijska (ZMK)

## **Rozdział 4. Korelacje między artystami. Dynamika sprzedaży płyt.**

### **4.1 Rynek fonograficzny jako rynek towarowy.**

Zanim powstały rynki finansowe – o których pisałem w poprzednich rozdziałach – przepływ kapitału odbywał się na zasadzie wymiany towarowej. Odkąd Fenicjanie wynaleźli pieniądze, rynki towarowe stały się osobną kategorią, podobnie jak giełda czy rynek wymiany walut. Wraz z rozwojem i wzrostem potrzeb człowieka, rynki towarowe przestały już dotyczyć wyłącznie potrzebnych w przemyśle surowców, gdyż treść intelektualna i duchowa z biegiem czasu także stała się towarem. W ten sposób Biblia stała się największym bestsellerem książkowym od czasu wynalezienia druku przez Gutenberga. Kolejne wynalazki pozwalały utrwalać także muzykę i film, jednak dopiero wynalazek fonografu (1877) Tomasza Edisona sprawił, że muzyka wydawana na płytach stała się początkiem przemysłu muzycznego. Dodam, że ten sam Tomasz Edison założył opisaną w drugim rozdziale firmę General Electric.

Początkowo patefon – następca fonografu - był droгим wynalazkiem, a posiadanie płyt było bardzo elitarne i dotyczyło ludzi bogatych. Dlatego pierwszym artystą, który sprzedał ponad milion egzemplarzy swojej płyty był legendarny włoski tenor Enrico Caruso – na początku XX wieku, tak jak w poprzednim stuleciu, wśród elit najpopularniejsza była opera. Podobnie jak w przypadku rynków finansowych, na kondycję rynku fonograficznego miały wpływ wielkie wydarzenia historyczne – I i II Wojna Światowa, czy kryzys roku 1928. W efekcie dominujący w pierwszej połowie XX wieku jazz nie przełożył swojej popularności na milionowe nakłady płyt. Tym bardziej, że używany do produkcji płyt winyl miał strategiczne znaczenie dla przemysłu obronnego i w efekcie podczas II Wojny Światowej Brytyjczycy zamiast kupować płyty sami ochotniczo oddawali już posiadane egzemplarze, celem zastosowania tak odzyskanego winylu do obrony kraju. Między innymi z tego powodu pierwszym artystą, który jak Caruso zaczął sprzedawać swoje płyty w milionach egzemplarzy, był pochodzący z Liverpoolu zespół The Beatles, a więc przedstawiciel kolejnej epoki – rock and rolla.

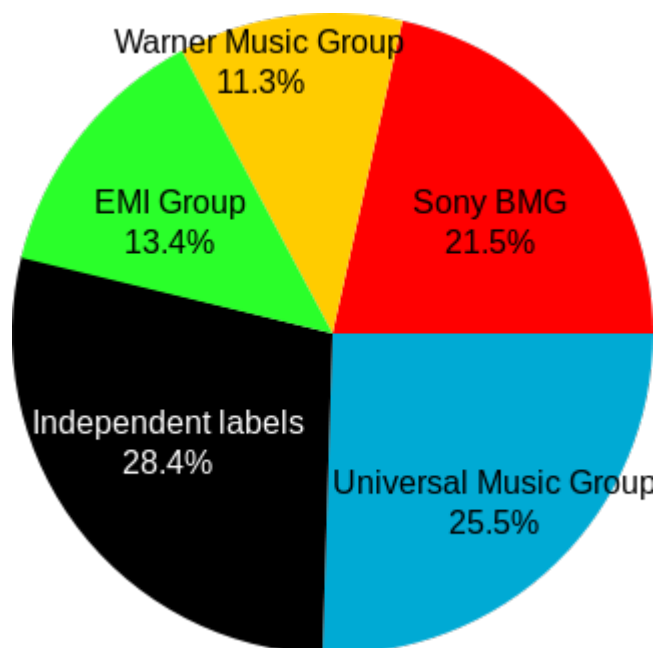
Nowym gatunkom na rynku towarzyszył rzecz jasna rozwój technologiczny. W efekcie w 1948 roku wytwórnia Columbia wprowadziła na rynek płytę długogrającą (*longplay*), która na jednej stronie mieściła około 20 minut nagrań. Był to ukłon zwłaszcza wobec twórców jazzowych, jak Duke Ellington, który mógł wyjść poza trwające 3 minuty kompozycje, odtwarzane z prędkością 78 obrotów na minutę, ale także twórców operowych, wydanie opery na winylu zajmowało bowiem wcześniej często kilkanaście nośników.

W ślad za longplayem o rozmiarze 12" odtwarzanym z prędkością 33 1/3 obrotu na minutę pojawiły się kolejne formaty, jak np. odtwarzany z prędkością 45 obr./min. 7" singiel. Na początku XXI wieku, wraz z powszechnym dostępem do Internetu, singiel przybrał elektroniczny format mp3, o wiele bardziej praktyczny niż płyta kompaktowa, która mogła pomieścić 74 minuty. Wynalazek odtwarzanej laserowo płyty kompaktowej okazał się skuteczny jedynie w przypadku albumów, CD skutecznie bowiem wyparł z rynku długogrającą płytę winylową, która dziś zamiast szerokich mas ma znaczenie jako produkt tylko dla grona audiofilów i koneserów, a więc podobnej elity, jak 100 lat wcześniej.

W moich badaniach nad rynkiem fonograficznym będę brał pod uwagę cotygodniową sprzedaż płyt długogrających – albumów. Zwłaszcza, że od czasów słynnej płyty The Beatles *Sgt Pepper's Lonely Hearts Club Band* (1967), większość udziału w rynku ma sprzedaż albumów. Ponadto definicja sprzedaży singla jest niejednoznaczna, np. na wiodącym rynku Stanów Zjednoczonych na notowania singla na liście branżowego tygodnika Billboard prócz sprzedaży ma wpływ także liczba pobrań plików mp3 oraz częstotliwość odtwarzania w stacjach radiowych i telewizyjnych. Zupełnie inaczej jest w Wielkiej Brytanii, gdzie pozycje na listach ustalane są zazwyczaj na podstawie sprzedaży singli. Tymczasem notowania albumów we wszystkich krajach świata ustalane są jednoznacznie – na podstawie ich cotygodniowej detalicznej sprzedaży.

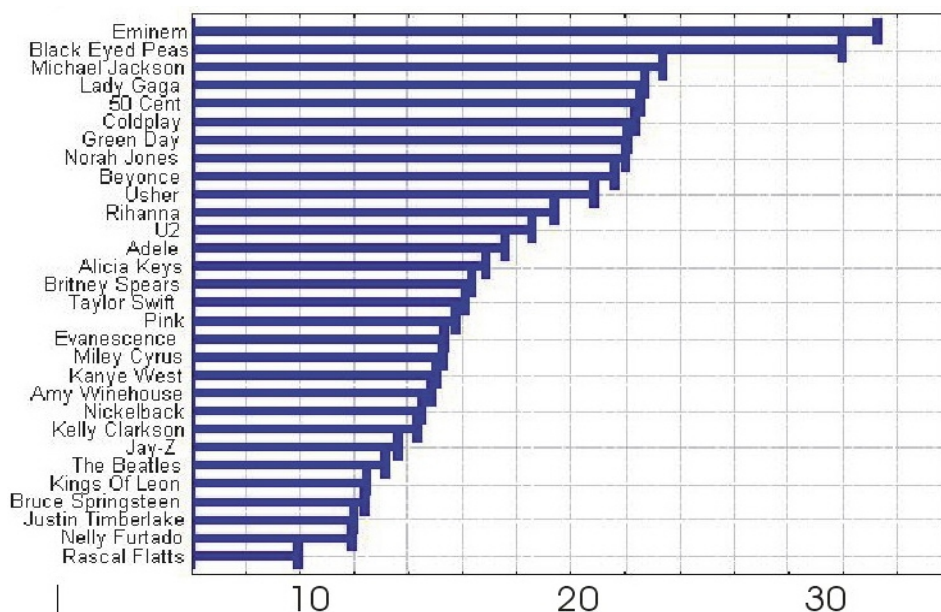
Pierwsza w świecie lista sprzedaży płyt została opublikowana przez tygodnik Billboard w 1949 roku (w Wielkiej Brytanii w 1953). Takie cotygodniowe listy były obarczone błędem pomiaru, powstawały bowiem w wyniku badań opinii o sprzedaży w losowo wybranych sklepach. W efekcie dochodziło do manipulacji, z których najsłynniejsze to oszustwa Elkie Brooks, ale także działania Briana Epsteina, który jako manager The Beatles kupił

10 tys. egz. singla swoich podopiecznych *Love Me Do* tylko po to, aby umieścić ich na liście przebojów.

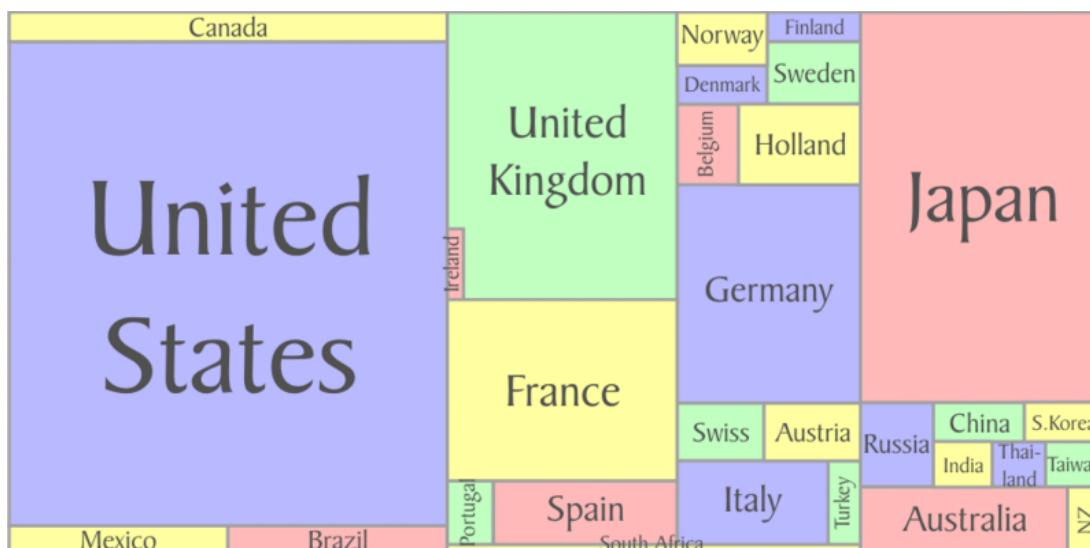


Rys. 4.1. Podział rynku fonograficznego ze względu na wytwórnice (2005) [59].

Dopiero od 1991 dzięki możliwości odczytywania kodów kreskowych sprzedanych płyt w Stanach Zjednoczonych działa system *Nielsen Soundscan*, który podaje sprzedaż z dokładnością do 1 egzemplarza. Wkrótce system objął także Kanadę i Wielką Brytanię, łącznie 14 tysięcy sklepów. W efekcie w 2003 roku IFPI (International Federation Of Phonographic Industry), zrzeszająca kraje, w których sprzedawane są oryginalne płyty, zintegrowała system na tyle, aby podawać cotygodniową sprzedaż danej płyty z dokładnością do 100 egzemplarzy w skali globalnej. Największą sprzedaż wszechczasów – prawie 100 milionów egzemplarzy - osiągnął wydany w 1982 roku album *Thriller* Michaela Jacksona [59], jednak jeśli chodzi o sprzedaż cotygodniową, to najpopularniejsze tytuły osiągają kilkadziesiąt tysięcy, a w chwili premiery nawet milion nabywców. W efekcie dokładność, z jaką znamy sprzedaż albumów, jest podobna do tej, z jaką podaje się kursy walut czy ceny akcji i wynosi dla tych rynków około 0,01%. Pomiar jest zatem dokładny. Dzieje się tak także dlatego, że większość – ponad 70% światowego rynku – jest pod kontrolą czterech wielkich koncernów: EMI, Universal, Warner i Sony BMG (Rys. 4.1). Mają one w swojej ofercie 30 najpopularniejszych na świecie artystów (Rys. 4.2), których uwzględniłam w moich badaniach.



Rys. 4.2. Całkowita sprzedaż płyt 30 najpopularniejszych artystów w latach 2003-2011 (w milionach egzemplarzy).



Rys. 4.3. Geografia rynku fonograficznego (2003) z uwzględnieniem położenia i wielkości lokalnego rynku sprzedaży [60].

Technologia - prócz wzrostu dokładności danych o sprzedaży – miała również wpływ na sposób słuchania muzyki. O ile słuchanie z taśm odeszło w zapomnienie już kilkanaście lat temu, to tradycyjne winyle stanowią około 3% rynku, a ich sprzedaż w ostatniej dekadzie systematycznie rośnie. Jednak większość, bo około 70-80% rynku stanowią płyty kompaktowe, a pozostałą część sprzedaż cyfrowa w postaci mp3 i dzwonek na telefony komórkowe. Oczywiście wraz z upowszechnieniem Internetu w drugiej połowie lat 90-tych sprzedaż płyt załamała się. Jednak umożliwienie sprzedaży cyfrowej zahamowało ten proces w 2006 roku, a w latach 2010-2012, dzięki nośnikom cyfrowym, całkowita sprzedaż płyt na świecie zaczęła nawet rosnąć – odpowiednio o 2% i 1% [60-62].

Dominacja czterech wielkich koncernów ma rzecz jasna wpływ na geografie rynku, bo największy wkład do globalnej sprzedaży dają Stany Zjednoczone, Wielka Brytania i Japonia. Natomiast tak duże kraje jak Rosja, Chiny i Brazylia mają mały wkład z powodu piractwa i małej siły nabywczej konsumentów (Rys. 4.3). Co ciekawe, 30 największych rynków lokalnych stanowi 95% ogółu światowej sprzedaży.

#### **4.2 Założenia i analogie z rynkami finansowymi. Analiza sprzedaży płyt.**

Aby zastosować analogię z wcześniej omówionymi rynkami finansowymi, musiałem dostosować dane z rynku fonograficznego do postaci odpowiedniej dla analizy statystycznej. Dlatego założyłem, że – analogicznie do akcji spółek na giełdzie lub kursów wymiany walut – przedmiotem obrotu na rynku fonograficznym są artyści, których wartość (notowania) mierzone są cotygodniową sprzedażą ich płyt długogrających, czyli albumów (winyle, CD, mp3 itd.). Założenie to jest uzasadnione, ponieważ na rynku cena płyty jest na ogół stała, ustalona z góry przez wytwórnię i dystrybutora. W szczególności stała cena sprzedaży dotyczy płyt nowych i premierowych, których przeceny w sklepach są dokonywane dopiero, gdy płyta spadnie już z list przebojów (a więc jej sprzedaż łączy się do zera i nie jest ujmowana w moich danych).

W moich badaniach biorę pod uwagę cotygodniowe dane o sprzedaży albumów na świecie w okresie od września 2003 do września 2013 (źródło: [musiccharts.net](http://musiccharts.net)). Do analizy powiązań między artystami wybrałem trzydziestkę najpopularniejszych wykonawców (Rys. 4.2), którzy w latach 2003-2013 sprzedali co najmniej 11 milionów płyt. Najpopularniejsi z nich – Eminem i Black Eyed Peas – osiągnęli sprzedaż prawie 30 milionów. W rzeczywistości jednak artyści często sprzedają sporo płyt występując solo, poza macierzystym zespołem. Z tego powodu do wyników sprzedaży Black Eyed Peas dodałem również sprzedaż solowych płyt członków tego zespołu: Fergie i Will I Ama. Podobnie w przypadku The Beatles uwzględniłem sprzedaż płyt Johna Lennona i Paula McCartneya, a do sprzedaży płyt Beyonce dodałem także te, które nagrała ze swoim macierzystym zespołem Destiny's Child. Ma to uzasadnienie, bowiem konsumenci identyfikują danego artystę i kupują jego muzykę, nawet jeśli występuje on pod innymi nazwami. Podobne zjawisko wystąpiło też na rynkach finansowych, gdy przed 2002 rokiem waluta euro w praktyce funkcjonowała jako waluty narodowe, a spółki należące do



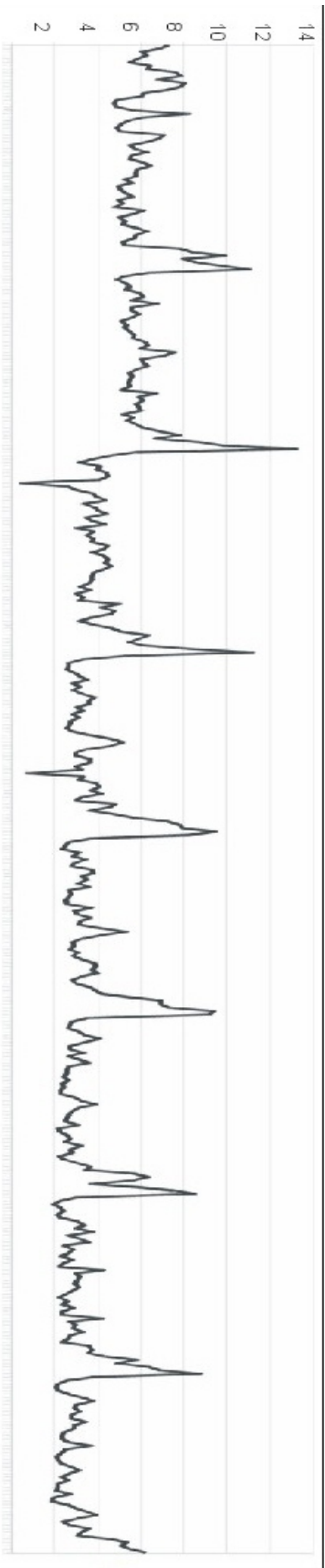
głównych indeksów giełdowych też czasem zmieniały swoje nazwy. Działo się to jednak o wiele rzadziej niż u wykonawców, którzy często wydają płyty pod inną nazwą ze względów artystycznych lub prawnych.

Oczywiście zdarza się, że popularni artyści mają więcej niż jedną płytę na liście najlepiej sprzedawanych albumów w tym samym czasie. Wówczas notowania artysty oznaczają sumę sprzedaży jego notowanych płyt w danym tygodniu. Staje się tak, gdy w okolicach premiery, na fali promocji konsumenci sięgają po wcześniejsze tytuły swojego idola. Na rynku fonograficznym, co jeszcze raz pragnę podkreślić, walorami są artyści, a nie pojedyncze tytuły.

<b>Rynek giełdowy</b>	<b>Rynek fonograficzny</b>
Notowania akcji na giełdzie	Notowania artysty na listach sprzedaży
Stopa zwrotu	Cotygodniowa zmiana sprzedaży płyt
Korelacje między akcjami	Korelacje między artystami
Odległość między akcjami	Odległość między artystami
Główny indeks giełdowy	Grupa najpopularniejszych artystów
Sektory i podsektory gospodarki	Gatunki muzyczne

**Tab. 4.1:** Analogie pomiędzy rynkiem fonograficznym a rynkami akcji.

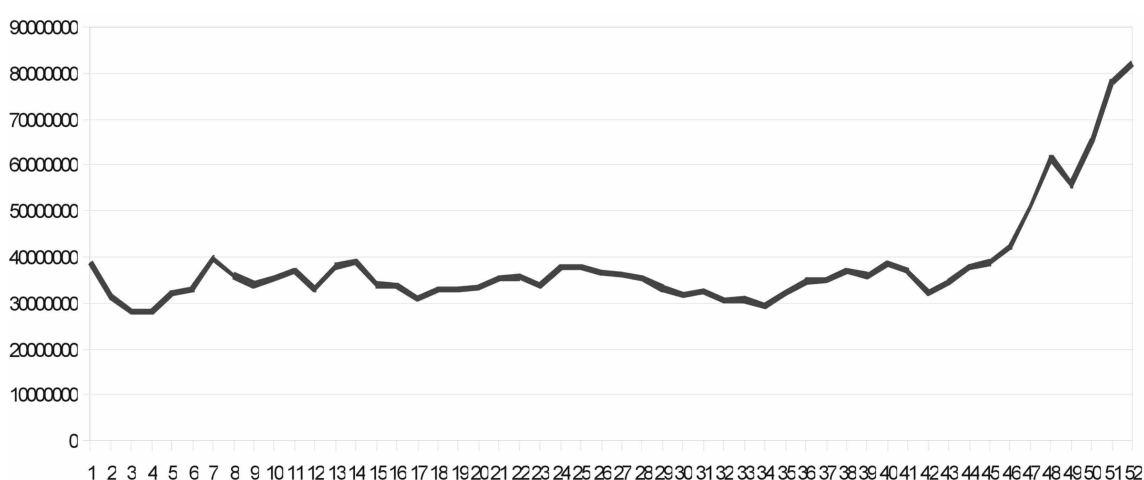
Analogicznie jak we wzorze (1.3) możemy określić współczynnik korelacji między artystami, jednak zamiast stóp zwrotu posługujemy się zmianą  $Y_t$  (wzrostem lub spadkiem) sprzedaży płyt w danym tygodniu w stosunku do sprzedaży w tygodniu wcześniejszym. Oczywiście w fachowych pismach branżowych, takich jak Billboard, wyraża się tę wielkość na ogół w procentach, jednak aby ujednoczyć moje analizy różnych rynków, przyjąłem stosować zmianę logarytmiczną (wzór 1.2). Podobnie jak na rynkach finansowych, percepcja zmiany sprzedaży na rynku fonograficznym ma charakter logarytmiczny, a nie liniowy, dlatego używanie logarytmu na wszystkich tych rynkach ma uzasadnienie.



**Rys. 4.4** Cotygodniowa sprzedaż 50 najpopularniejszych płyt [w milionach egzemplarzy] od września 2003 do września 2012

Rysunek 4.4 pokazuje wykres cotygodniowej skumulowanej sprzedaży płyt wszystkich 718 artystów, którzy umieścili swoje albumy na liście najlepiej sprzedawanych płyt na świecie w latach 2003-2012. Obserwujemy na nim wieloletni trend spadkowy sprzedaży oraz sezonowość, które to zjawiska warto przeanalizować dokładnie.

Sezonowość sprzedaży płyt obejmuje nie tylko globalne maksimum sprzedaży w tygodniu poprzedzającym święta Bożego Narodzenia (51. tydzień roku), ale także mniejsze, lokalne maksima w okolicach Dnia św. Walentego oraz wiele mniejszych maksimów, które powtarzają się regularnie. Aby je zlokalizować, przygotowałem wykres uśrednionej sprzedaży, obejmujący cały sezon od pierwszego tygodnia stycznia do ostatniego tygodnia grudnia, zakładając, że rok ma 52 tygodnie.

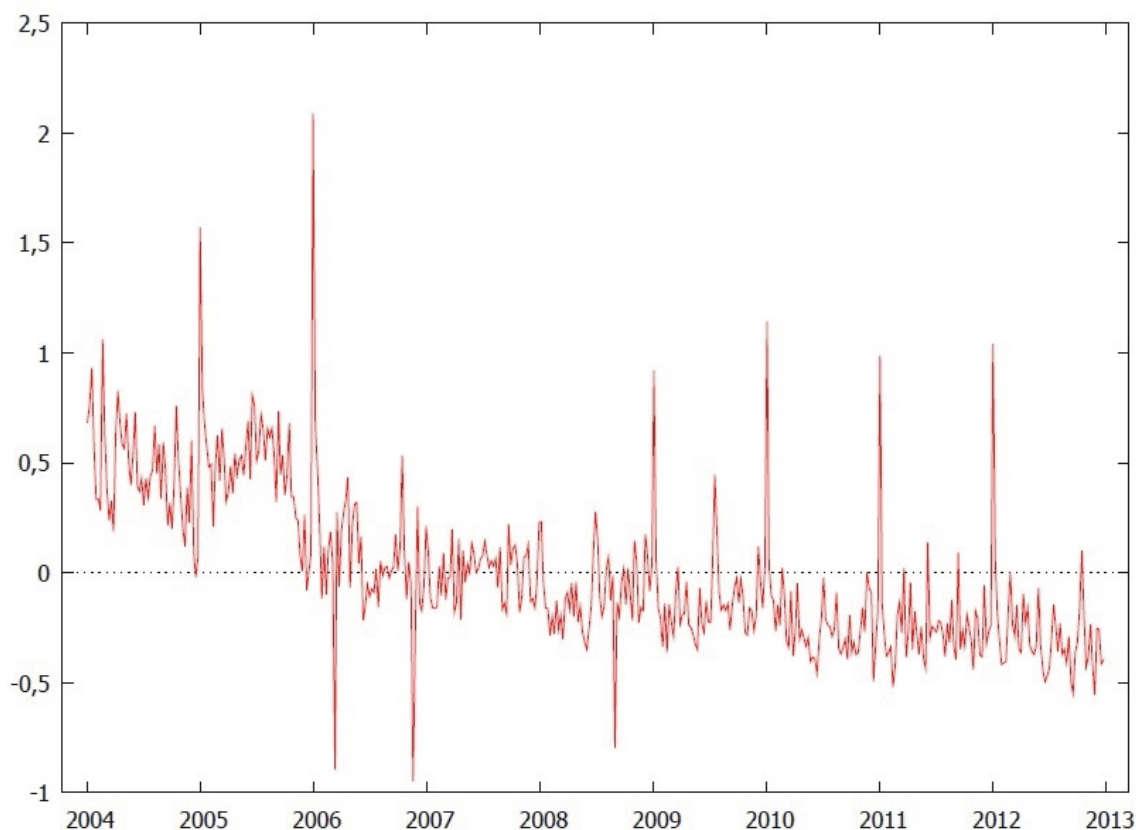


Rys. 4.5 Sezonowo uśredniona sprzedaż płyt od stycznia do grudnia na podstawie danych z sezonów 2003-2012.

Rysunek 4.5 pokazuje, że sprzedaż płyt w sezonie osiąga maksimum na poziomie 10 milionów egzemplarzy w ostatnich tygodniach grudnia. Liczba ta przerasta typową sprzedaż w pozostałych tygodniach nawet trzykrotnie. Oznacza to, że pod koniec roku zawsze włącza się nowa grupa konsumentów, którzy kupują płyty „od święta” - co będzie miało później znaczenie dla hierarchii artystów na rynku.

Ciekawa jest też analiza sprzedaży w latach 2003-2013 po odjęciu trendu, czyli różnicy między sprzedażą, a sprzedażą sezonowo uśrednioną z Rys. 4.5. Przedstawiłem ją na Rys. 4.6, gdzie zauważamy wyraźny spadek sprzedaży w latach 2003-2005 oraz zahamowanie tego spadku od roku 2006. Przyczyną tego zjawiska jest internetowe piractwo, które właśnie w 2006 roku zostało ograniczone przez udostępnienie sprzedaży płyt na nowe nośniki cyfrowe, takie jak mp3, za pośrednictwem sklepów internetowych w rodzaju iTunes, co wyszło naprzeciw potrzebom konsumenta. Wykres pozbawionej trendu

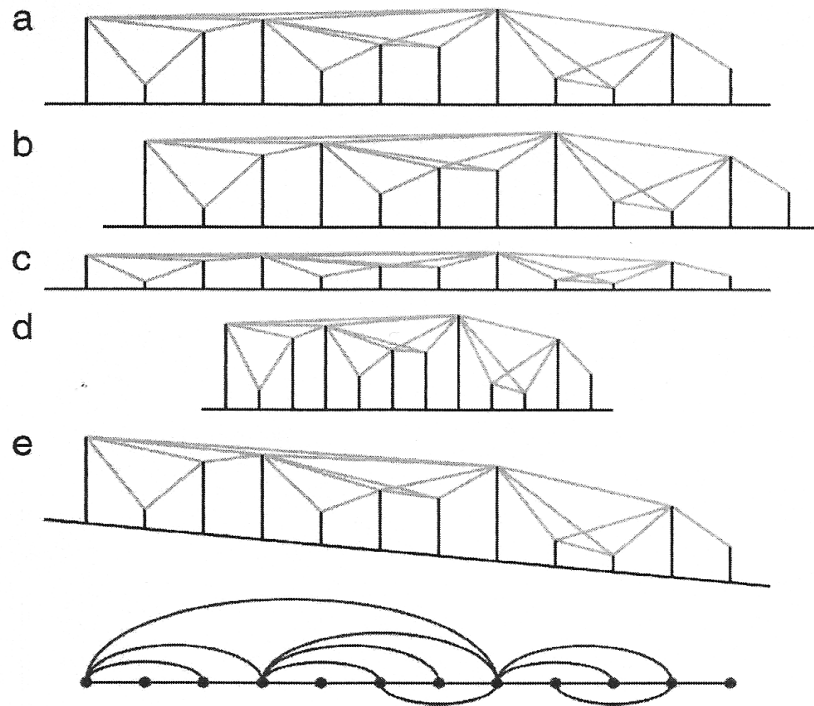
sprzedaży posiada też maksima w odstępie roku, wynikające z faktu, że rok ma 365 lub 366 dni kalendarzowych, a kolejne dni odpowiadających sobie tygodni w różnych latach są różne, np. 22 grudnia 2012 jest sobotą, a 22 grudnia 2009 – wtorkiem. Najważniejsze dla rynku jest to, że pozbawiona trendu sprzedaż od 2006 roku utrzymuje się na mniej więcej stałym i stabilnym poziomie.



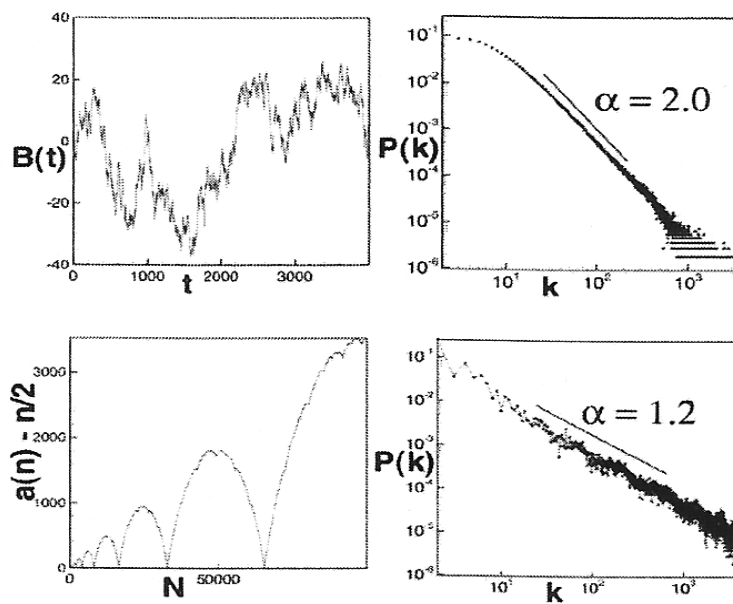
Rys. 4.6 Pozbawiona trendu sprzedaż płyt od września 2003 do grudnia 2012.

Ponieważ dane o sprzedaży mają strukturę sezonową, do ich analizy użyłem też techniki grafów widzialności (*visibility graphs*). Polega ona na przedstawieniu danych w formie wykresu słupkowego, a następnie połączeniu w grafie jako sąsiadów tych słupków, które „widzą” siebie nawzajem (Rys. 4.7). Przedstawienie serii czasowej za pomocą grafu widzialności ma tę zaletę, że technika ta jest niezmiennicza względem wielu przekształceń, takich jak przesunięcie (Rys. 4.7b), przeskalowanie w kierunku prostopadłym (Rys. 4.7c) i równoległym (Rys. 4.7 d), czy dodanie trendu do danych (Rys. 4.7e). Metoda ta w odniesieniu do danych sezonowych oraz zupełnie losowych (Rys. 4.8) jest dobrze znana [64], dlatego zastosowałem ją w tej pracy do serii czasowej cotygodniowych danych o globalnej sprzedaży płyt.

Okazało się, że badana przeze mnie seria przedstawiona na Rys. 4.4, mimo że złożona z sezonowego trendu (Rys. 4.5) i fluktuacji (Rys. 4.6), wykazuje inne własności niż te dla przedstawionych na Rys. 4.8 wygenerowanych sztucznie sezonowych i losowych serii czasowych. Funkcja gęstości prawdopodobieństwa rozkładu krotności węzłów  $P(k)$  przyjmuje wartość zero, gdy  $k=1$  lub  $k=2$ . Oznacza to, że w serii czasowej sprzedaży

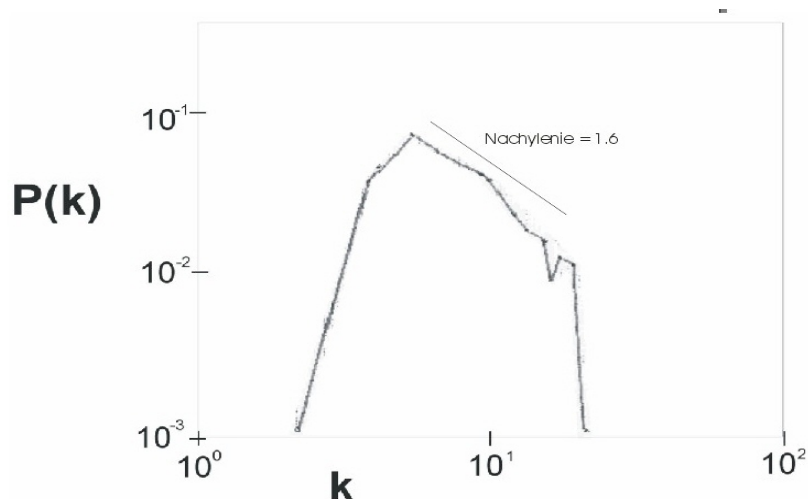


Rys. 4.7 Niezmienniczość grafu widzialności względem wybranych transformacji serii czasowej [64].



Rys. 4.8 Serie czasowe – losowa z rozkładu Poissona (część górna) i sezonowa (część dolna) – wraz z wykresami gęstości prawdopodobieństwa  $P(k)$  rozkładu krotności  $k$  węzłów w skali podwójnie logarytmicznej [64].

płyty nie istnieją tygodnie odstające sprzedażą od swoich sąsiadów z poprzedniego i następnego tygodnia na minus na tyle, aby stracić widzialność ze słupkiem maksymalnej sprzedaży pod koniec roku. Dzięki temu rynek fonograficzny jest stabilny i w dużym stopniu przewidywalny: płyty będą sprzedawać się zawsze i to bez gwałtownych zmian. Przyczyną tego mogą być stabilne dochody konsumentów, otrzymywane regularnie w odstępach tygodniowych lub miesięcznych.



**Rys. 4.9** Analogiczna jak na Rys. 4.8 gęstość rozkładu prawdopodobieństwa  $P(k)$  krotności węzłów  $k$  dla serii czasowej cotygodniowej sprzedaży płyt z Rys. 4.4.

Możliwe, że dlatego właśnie maksimum funkcji  $P(k)$  obserwujemy dla krotności  $k=5$ , a sama funkcja jest zupełnie inna niż dla generowanych rozkładów sezonowych i losowych (Rys. 4.9). Dla  $k > 5$  rozkład prawdopodobieństwa  $P(k)$  w skali logarytmicznej maleje liniowo przez następnych kilka wartości  $k$ , a współczynnik nachylenia  $\alpha$  wynosi 1,6, czyli dokładnie pomiędzy 1,2 a 2,0 charakterystycznymi dla serii sezonowych i losowych (Rys. 4.8).

W praktyce rynek fonograficzny jest więc przewidywalny, gdyż po usunięciu trendu (Rys. 4.6) i unormowaniu sygnału tak, aby średnia sprzedaż wynosiła 1 (oryginalne dane przedstawione na Rys. 4.4 zostały podzielone przez trend sezonowy z Rys. 4.5), otrzymamy odchylenie standardowe 0,36, medianę 0,91, kurtozę 2,66 i skośność równą 1,16, co jest efektem spadku sprzedaży płyt w latach 2003-2006 i późniejszego jej ustabilizowania się. Jednak i tu na sprzedaż mają wpływ zdarzenia losowe, np. śmierć artysty. Zapotrzebowanie na nagrania zmarłych artystów, takich jak Michael Jackson (2009) czy Amy Winehouse (2011), było bowiem tak duże, że wywindowało ich stare płyty na sam szczyt list sprzedaży tuż po ich śmierci. Było to możliwe, bo wielkie

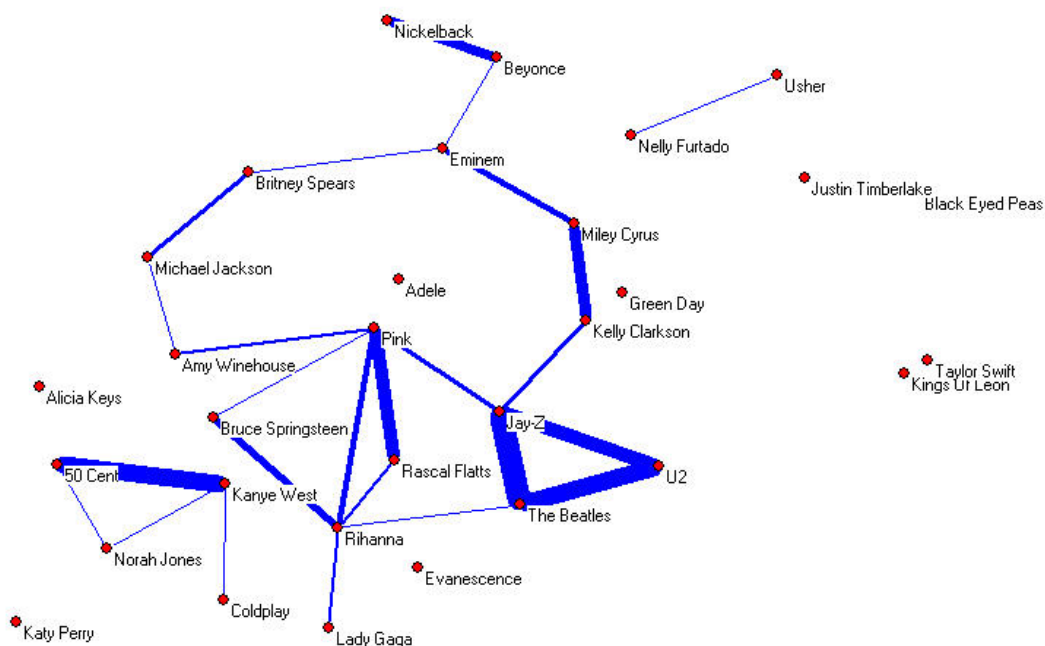
koncerny (Rys. 4.1) mają dziś technologie pozwalające tłoczyć setki tysięcy nowych egzemplarzy w ciągu tygodnia, a zatem zdążyć na najbliższe notowanie cotygodniowej listy sprzedaży. Z pomocą przychodzi też cyfryzacja i możliwość nieograniczonego powielania plików muzycznych. Dla porównania: w roku 1977, gdy umierał Elvis Presley, czy w roku 1980, gdy zastrzelono Johna Lennona, cykl produkcyjny płyt winylowych trwał dłużej i, aby sprostać zapotrzebowaniu, Wielka Brytania musiała zlecać pracę tłoczniom we Francji, a i tak zmarli artyści dotarli na pierwsze miejsca z trwającym parę tygodni poślizgiem [59]. Ten technologiczny efekt jest pokrewny efektom obserwowanym np. na rynkach akcji, gdzie czas reakcji rynku na wydarzenie uległ znacznemu skróceniu na przestrzeni ostatnich dziesiątków lat (np. efekt Eppsa).

### 4.3 Hierarchiczna struktura rynku fonograficznego.

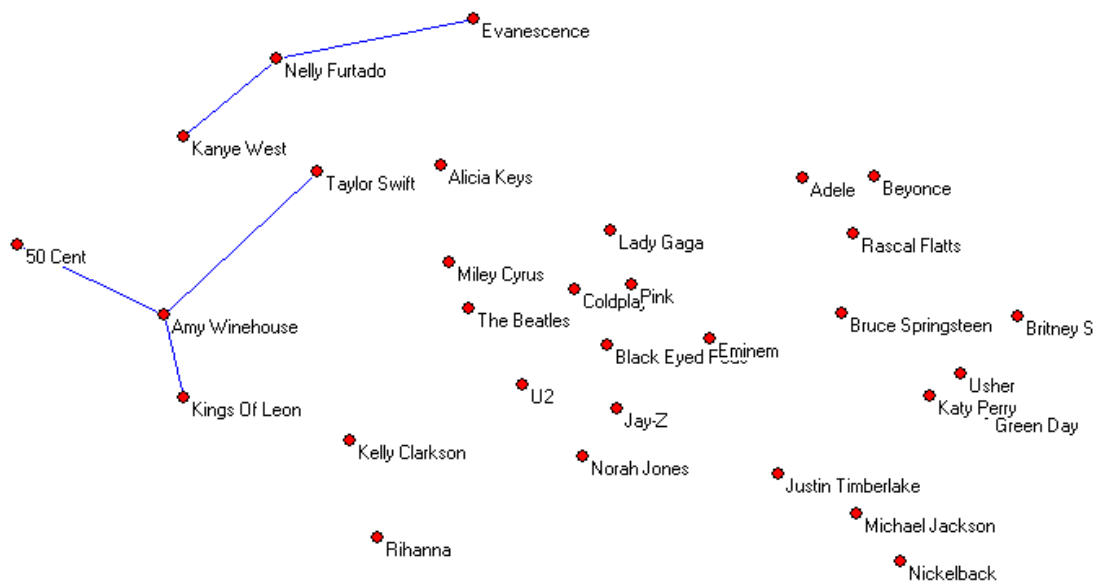
Czy zatem istnieje istotna różnica między rynkiem giełdowym, walutowym i towarowym? Więcej danych statystycznych pokażę pod koniec rozdziału, jednak nawet prosta analiza korelacji między walorami na tych rynkach – to jest liczby korelacji silnych, słabych i ujemnych – pokazuje, że rynek fonograficzny zdecydowanie różni się od rynków finansowych ze względu na dominujące na nim korelacje ujemne [65]. Oznacza to, że jeden artysta zyskuje wysoką sprzedaż kosztem drugiego, czyli – jak na każdym rynku towarowym – występuje tu zjawisko konkurencji między produktami (artystami).

**Tab. 4.2.** Liczba skorelowanych par walorów na rynkach akcji, walutowych i na rynku fonograficznym.

Skorelowane pary	silnie	słabo	ujemnie
DJIA	9	426	0
DAX	205	119	1
WIG 20	1	188	1
FOREX	Różna w zależności od waluty bazowej	Różna w zależności od waluty bazowej	Różna w zależności od waluty bazowej
Rynek fonograficzny	3	72	375



Rys 4.10 Istotne ( $p = 0,05$ ) dodatnie korelacje między artystami [66]



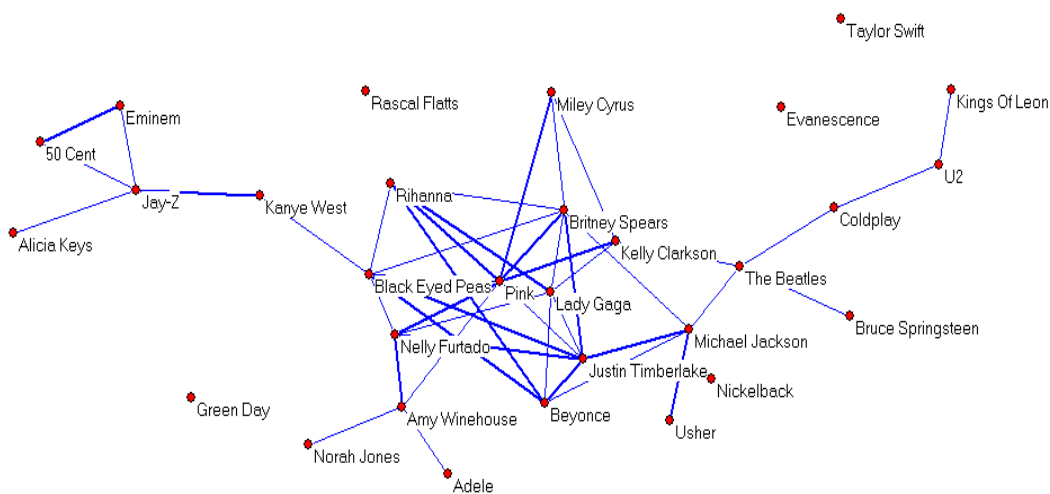
Rys. 4.11 Istotne ( $p = 0,05$ ) ujemne korelacje między artystami [66].

Ponieważ na rynku płytowym dominują korelacje ujemne (Tab. 4.2), postanowiłem sprawdzić, czy ich przyczyną jest jedynie konkurencja między produktami. Okazuje się, że spośród 375 ujemnie skorelowanych par znacząco ujemnych jest jedynie kilka (Rys. 4.11), przy czym są one skupione głównie wokół piosenkarki Amy Winehouse. Aby znaleźć przyczynę tego efektu, warto sięgnąć do historii rynku i cofnąć się do 2011 roku, gdy Amy Winehouse zmarła. Było to w lipcu, a więc w czasie, gdy wytwórnie unikają



premier płytowych, a sprzedaż jest niska. Wówczas to śmierć artystki wywindowała jej nagrania z powrotem do czołówek list najlepiej sprzedawanych płyt, kosztem innych artystów. Śmierci Amy Winehouse nie towarzyszyła żadna istotna premiera płytowa, a więc sprzedaż jej płyt gwałtownie rosła w sytuacji, gdy reszta rynku miała powolny spadek sprzedaży. W tym wypadku śmierć artystki wyprowadziła rynek ze stanu równowagi i stąd pojawienie się tylu ujemnych korelacji. Nieco inny efekt wystąpił, gdy w 2009 roku umierał Michael Jackson – a więc największa gwiazda ostatnich 40 lat. Był wtedy czerwiec, a zatem jego śmierć nałożyła się na przedwakacyjną, zwiększoną podaż premier płytowych. Stąd w otoczeniu Jacksona brak jest istotnych ujemnych korelacji, pomimo że w pierwszym tygodniu po śmierci „króla popu” sprzedano kilka milionów jego płyt.

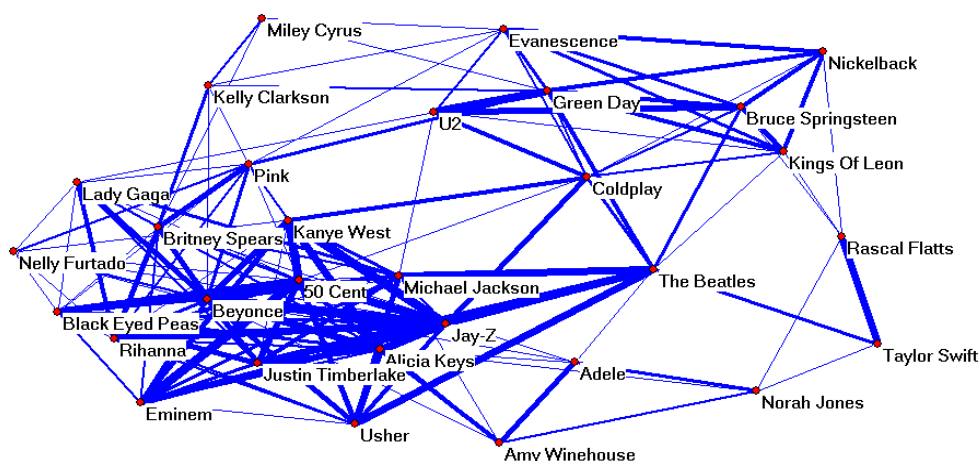
Warto dodać, że zwiększona wartość artysty po śmierci to cecha osobliwa dla rynku fonograficznego. Dla porównania, na rynku akcji wycofanie spółki z obrotu giełdowego powoduje raczej spadek ceny jej akcji, czyli efekt całkowicie odwrotny.



**Rys. 4.12** Sieć podobieństw estetycznych między artystami obliczona na podstawie danych serwisu allmusic.com [66]. Grubość krawędzi odzwierciedla siłę podobieństwa określoną w skali 1-10.

Korelacje między artystami można wykorzystać do określenia podobieństwa estetycznego wykonawców, tak jak zrobił to serwis allmusic.com, biorąc pod uwagę ankiety swoich użytkowników (Rys. 4.12). Tak określona subiektywna ocena podobieństwa jest zbieżna z klasyfikacją, której co roku dokonuje branżowy tygodnik Billboard, dzieląc przemysł

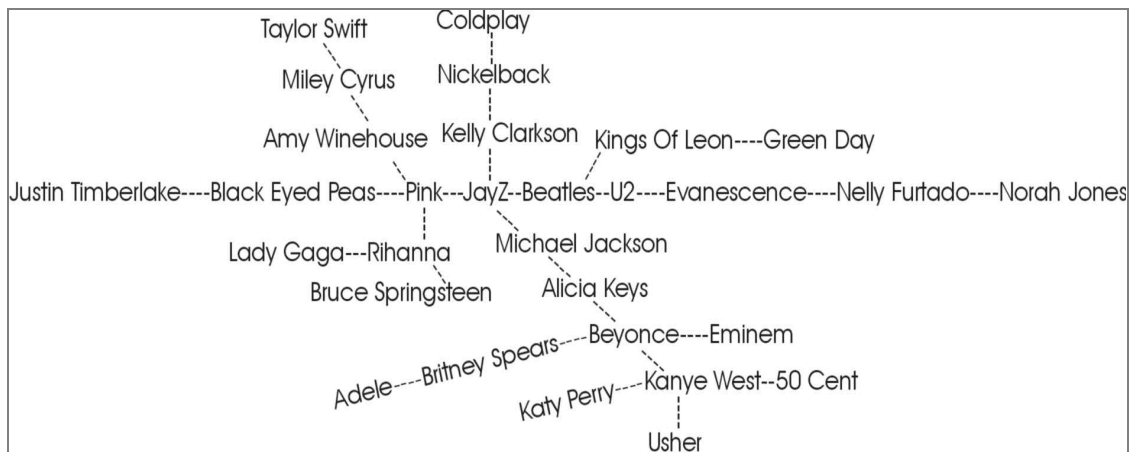
płytowy na sektory gatunkowe [65-67] (dla rynku akcji klasyfikacji tego typu dokonuje np. Forbes, grupując spółki według sektorów gospodarki).



**Rys. 4.13** Podobieństwa między artystami według mojej własnej oceny estetycznej [66]. Grubość krawędzi odzwierciedla siłę podobieństwa określoną w skali 1-10.

Aby mieć pewność co do tak zdefiniowanego podobieństwa estetycznego między artystami, ja również przesłuchałem ich nagrania i dokonałem oceny ich podobieństwa według własnego gustu w skali 1-10. Odpowiednią sieć przedstawiłem na Rys. 4.13. Dopiero teraz mogę sprawdzić, czy na podstawie samych danych ekonomicznych o cotygodniowej sprzedaży płyt da się podobieństwo estetyczne między artystami odtworzyć i na podstawie wyróżnionych klastrów w sieci rozróżnić gatunki muzyki. Obliczyłem w tym celu współczynnik gronowania (klasteryzacji)  $C$  według wzoru (0.18), traktując wskazane wyżej sieci artystów w sposób binarny (tj. elementy macierzowe grafu mają wartość 1, gdy artyści są połączeni, a zero w przeciwnym wypadku). Okazało się, że największy współczynnik gronowania ma sieć artystów oparta na podobieństwie estetycznym według mojej oceny ( $C = 0,607$ ) oraz oceny krytyków muzycznych ( $C = 0,446$ ), a najmniejsze – sieć oparta na korelacjach trajektorii sprzedaży, czyli podobieństwie ekonomicznym ( $C = 0,201$ ) oraz klasyfikacji serwisu allmusic.com ( $C = 0,235$ ).

Dzięki definicji współczynnika korelacji możemy – identycznie jak na rynkach finansowych – określić odległość między „walorami”, czyli artystami, stosując wzór (1.5), w którym  $i, j$  oznaczają tym razem artystów. Dla tak określonej metryki możemy zbudować drzewo minimalnej rozpiętości (Rys. 4.14) i towarzyszącą mu hierarchię artystów w przestrzeni ultrametrycznej (Rys. 4.15).



**Rys. 4.14** Drzewo MST uzyskane dla 30 najpopularniejszych artystów na podstawie sprzedaży ich płyt w latach 2003-2011 [65].

Na rynkach finansowych drzewo MST odzwierciedlało strukturę rynku, czyli sektory i podsektory gospodarki, co zostało omówione w pierwszym rozdziale. Czy zatem rynek towarowy, jakim jest fonografia, także pokaże stylistyczne związki między artystami? Okazuje się, że nie do końca. Drzewo z Rys. 4.14 otrzymane dla „portfela” 30 najpopularniejszych artystów odzwierciedla główne style muzyczne, jednak z wyjątkiem muzyki pop. U góry i po prawej stronie drzewa dominują artyści rockowi (Coldplay, Nickelback, Kelly Clarkson, Kings Of Leon, Green Day, Evanescence, U2). W dolnej części drzewa obserwujemy sektor związany z rapem (50 Cent, Kanye West, Usher, Eminem), z którym sąsiadują artyści soul i r'n'b (Adele, Beyonce, Britney Spears, Alicia Keys, Michael Jackson). Brakuje natomiast sektora muzyki pop. Zamiast niego w środku drzewa sąsiadują ze sobą gwiazdy różnych gatunków: rocka (The Beatles, Bruce Springsteen), rapu (Jay-Z), r'n'b (Rihanna) i popu (Lady Gaga).

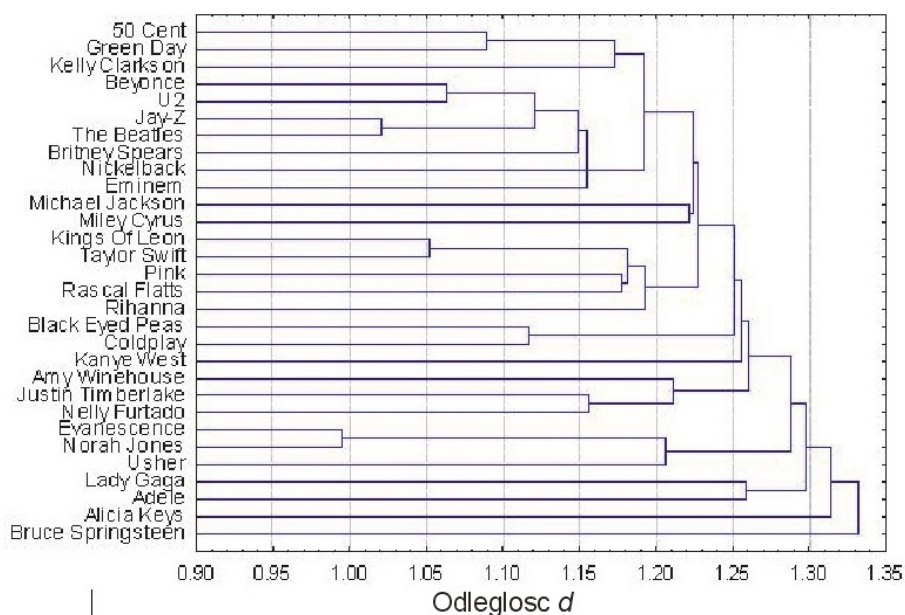
O ile zatem na rynkach finansowych można rozpoznać sektory gospodarcze po drzewie minimalnej rozpiętości, to dla rynku fonograficznego na podstawie sprzedaży płyt drzewo MST rozróżnia co prawda gatunki muzyczne, jednak dla muzyki pop nie istnieje osobny klaster.

Warto się w związku z tym zastanowić, czy wynik ten stawia pod znakiem zapytania estetyczną definicję muzyki pop. Zgodnie bowiem z definicją, gatunek „pop” jest skrótem od słowa „popularny” [59], co oznacza, że tak naprawdę to wszyscy artyści, mimo tworzenia estetycznie różnych odmian muzyki, należą do kręgu pop i próba tworzenia estetycznie nowego gatunku mija się z celem. Klasyczne gatunki, takie jak rock, soul czy

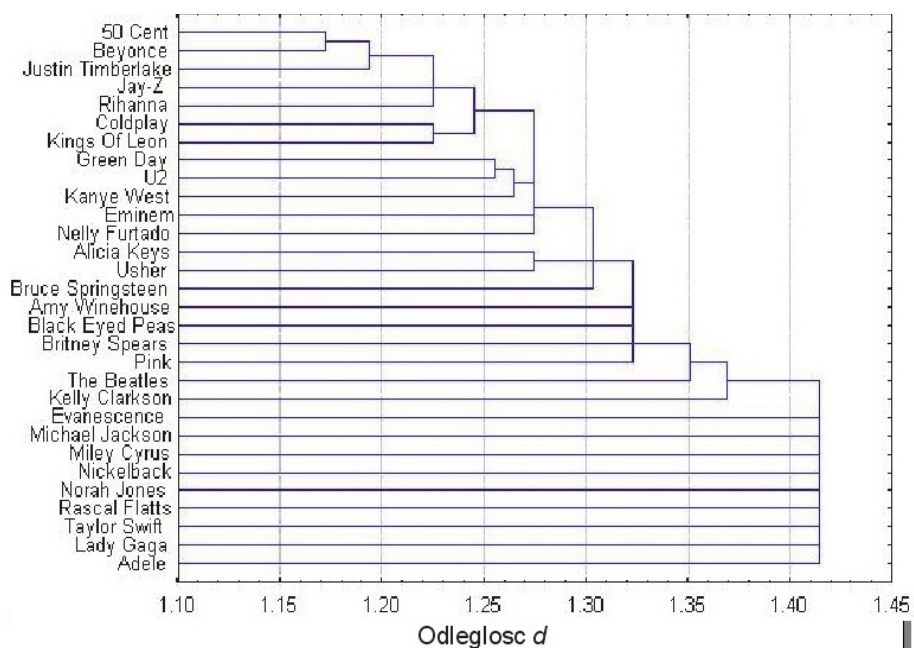
r'n'b zostały zdefiniowane 50-60 lat temu, a ich granice estetyczne praktycznie się nie zmieniły do dzisiaj, podobnie jak granice rapu, który istnieje w przemyśle płytowym od 35 lat [67]. Tymczasem pop oznacza dziś zupełnie inną muzykę niż 30, 40 czy 50 lat temu, a wyrazem tej różnicy jest zupełnie inna estetyka utworów zaliczanych do tego stylu w różnych okresach czasu. Tak więc odkrycie jedynie na podstawie drzewa MST, które gatunki muzyczne istnieją rzeczywiście – w sensie ekonomicznym dla konsumenta – a które nie, jest wynikiem nietrywialnym.

Dla uzupełnienia prezentuję diagram hierarchiczny artystów na rynku fonograficznym (Rys. 4.15), odpowiadający drzewu MST (Rys. 4.14), a także diagram podobieństwa między artystami na podstawie opinii krytyków muzycznych (o najlepszych płytach roku 2003, 2004, 2005, 2006, 2007 – pochodzą z podsumowań rocznych takich pism, jak New Musical Express, Village Voice, Rolling Stone, Spin, Uncut, Q, Vox itd.). Widać, że rynek muzyczny postrzegany oczami krytyków (Rys. 4.16) różni się nieco od tego, który wynika z ekonomii sprzedaży płyt, bo płyty o najlepszych recenzjach nie zawsze sprzedają się najlepiej [59].

Ostatecznie więc estetyczne podobieństwo między artystami (wyrażone przez recenzje krytyków, opinie użytkowników serwisu allmusic.com oraz moją własną ocenę) ma bardzo słabą korelację z siecią artystów stworzoną na podstawie ekonomicznych danych o sprzedaży płyt. Pokazuje to Tab. 4.3



**Rys. 4.15** Hierarchiczna struktura rynku fonograficznego dla 30 najpopularniejszych artystów na podstawie sprzedaży płyt w latach 2003-2011 [65].



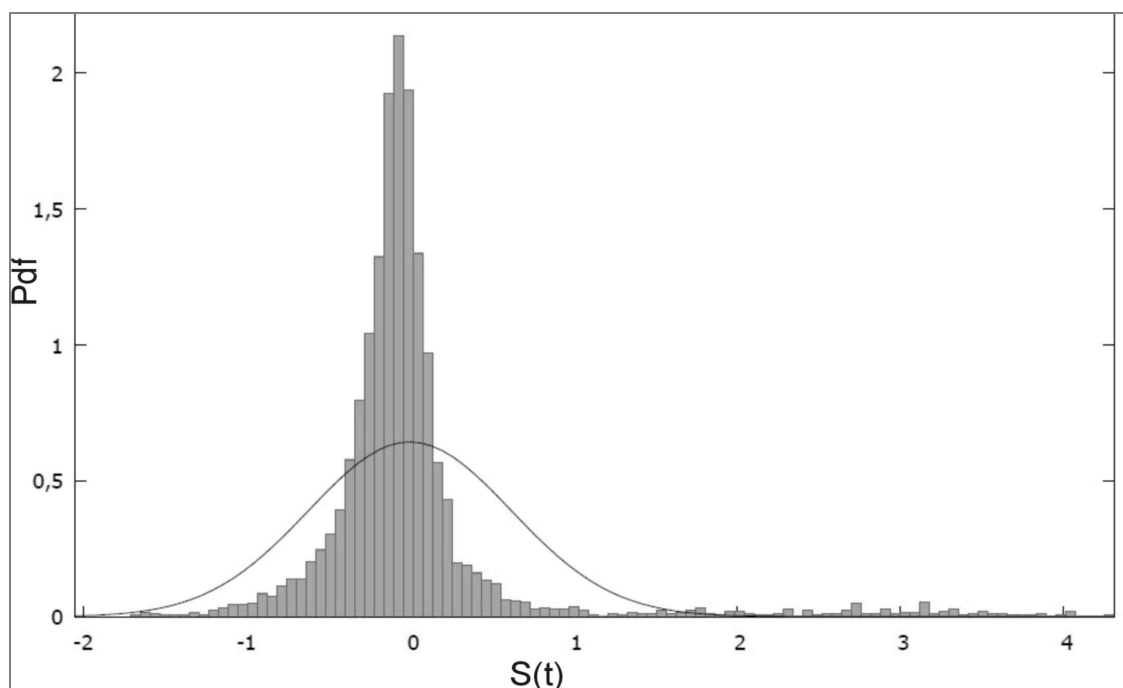
Rys. 4.16 Diagram podobieństwa między artystami na podstawie podsumowań rocznych według opinii krytyków z lat 2003, 2004, 2005, 2006, 2007 [66].

Tabela 4.3 Współczynniki korelacji między przedstawionymi wcześniej sieciami artystów. Kolorem wyróżniłem korelacje statystycznie znaczące ( $p = 0.05$ ) [66].

Networks	2003-2011				
	expert (Buda)	experts	all_music	market_corr	market_sq
expert (Buda)	1.00	<b>0.79</b>	<b>0.34</b>	-0.01	0.02
experts	<b>0.79</b>	1.00	<b>0.41</b>	-0.01	-0.01
all_music	<b>0.34</b>	<b>0.41</b>	1.00	0.02	0.02
market_corr	-0.01	-0.01	0.02	1.00	<b>0.93</b>
market_sq	0.02	-0.01	0.02	<b>0.93</b>	1.00

#### 4.4 Statystyczne własności rynku fonograficznego

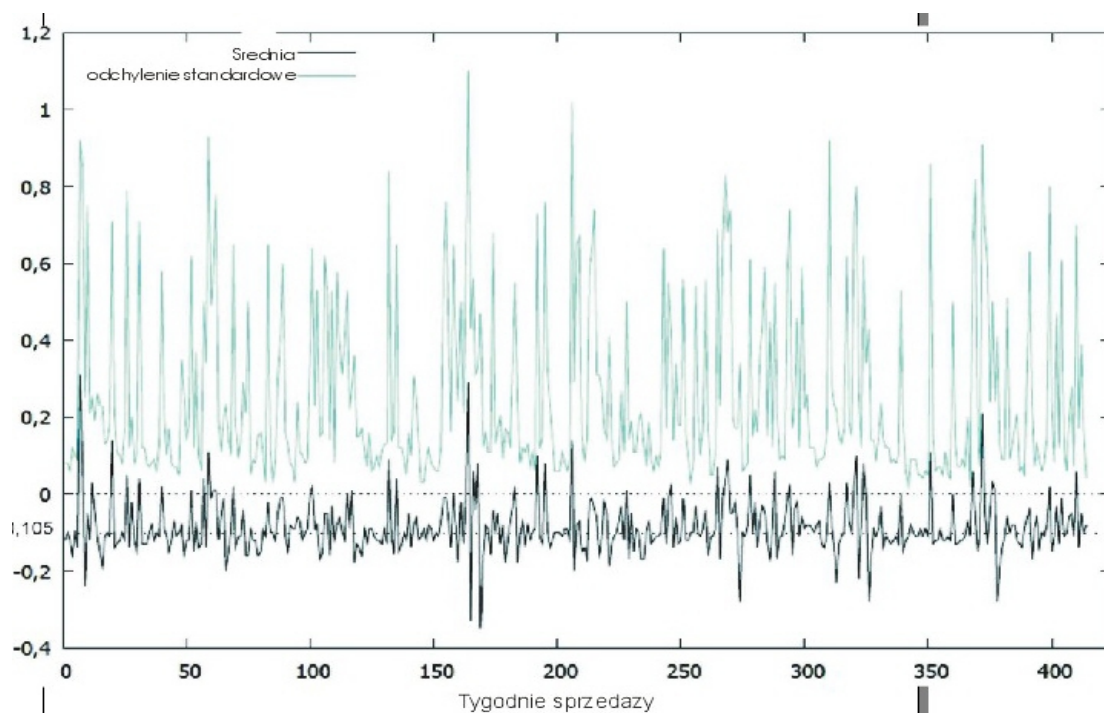
Ciekawym zagadnieniem jest problem, dlaczego korelacje między artystami różnią się tak mocno od korelacji między akcjami czy walutami. Aby odpowiedzieć na to pytanie, warto sięgnąć do podstawowych własności statystycznych rynków fonograficznych i finansowych.



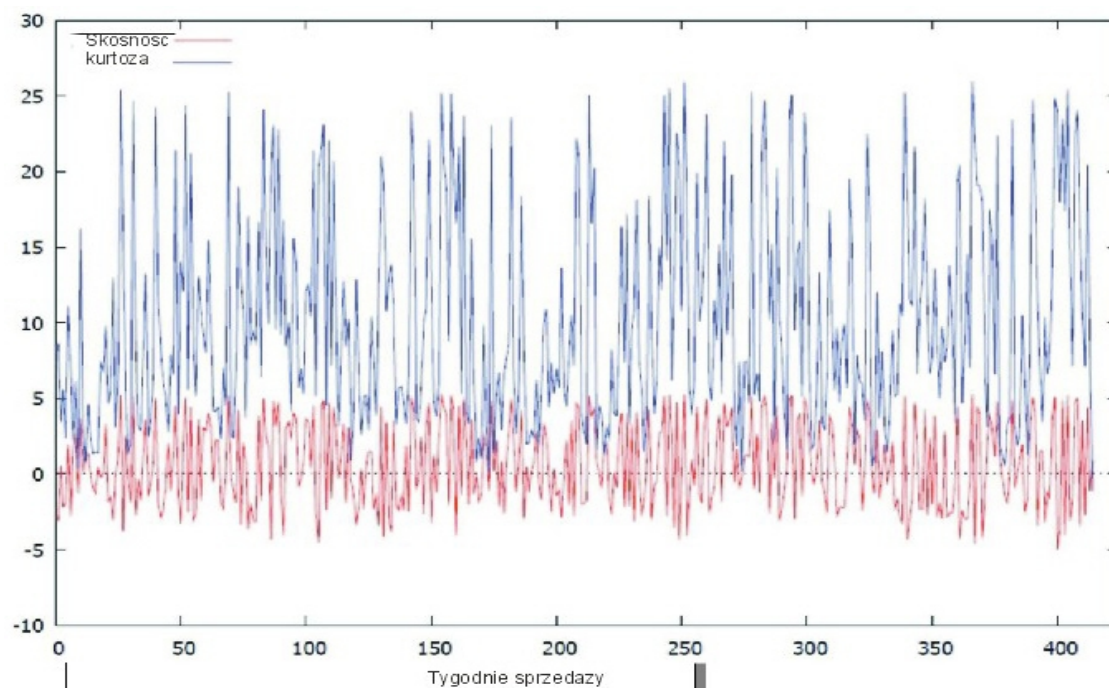
**Rys. 4.17** Gęstość prawdopodobieństwa jako funkcja różnicy logarytmów tygodniowej sprzedaży płyt uzyskana dla 30 najpopularniejszych artystów w okresie od 2003 do 2011 roku. Gładka linia jest wykresem gęstości prawdopodobieństwa rozkładu normalnego dla identycznej wariancji, jak z danych doświadczalnych [65].

Najbardziej ogólny stochastyczny model dynamiki cen walorów zakłada, że  $P_i(t)$  jest procesem dyfuzyjnym przy założeniu, że jego wartości mają rozkład Gaussa. Model ten, znany jako geometryczny ruch Browna, wyjaśnia w pierwszym przybliżeniu zachowanie obserwowanych danych doświadczalnych [35]. Jednak od 1963 roku obserwujemy, że rozkład ten jest bardziej leptokurtyczny od rozkładu Gaussa [36-38]. Z hipotezy Mandelbrota-Levy'go wynika, że  $\ln P_i(t)$  podlega nieciągłej ewolucji w czasie [36], a  $S(t) = \ln P_i(t+1) - \ln P_i(t)$  charakteryzuje się niegaussowskim skalowaniem i rozkładem, którego drugi i kolejne momenty są nieskończone.

Zjawisko to zaobserwowałem także na rynku fonograficznym, gdzie silnie leptokurtyczny rozkład jest charakteryzowany przez węższe i wyższe maksimum oraz grubszy ogon niż dla rozkładu normalnego (Rys. 4.17). Z drugiej strony taki rozkład ma tylko jeden gruby ogon - po prawej stronie - gdyż duża dodatnia wartość  $S(t)$  bierze się z dużych skoków sprzedaży płyt w chwilach premier nowych albumów. Z tego powodu rynek fonograficzny jest bardziej przewidywalny niż rynek finansowy, bo sprzedaż wiąże się z cyklem życia produktu, zależy od jakości płyty i jej promocji. Nawet tu jednak obserwujemy zjawiska całkowicie losowe. Niespodziewana śmierć Michaela Jacksona i Amy Winehouse silnie wpłynęła na sprzedaż ich płyt, zwiększając ją do kilku milionów egzemplarzy miesięcznie.



**Rys. 4.18** Średnia logarytmiczna zmiana cotygodniowej sprzedaży płyt 30 najpopularniejszych artystów (dolna krzywa) oraz jej odchylenie standardowe (górną krzywą) w okresie 2003-2011 [65].

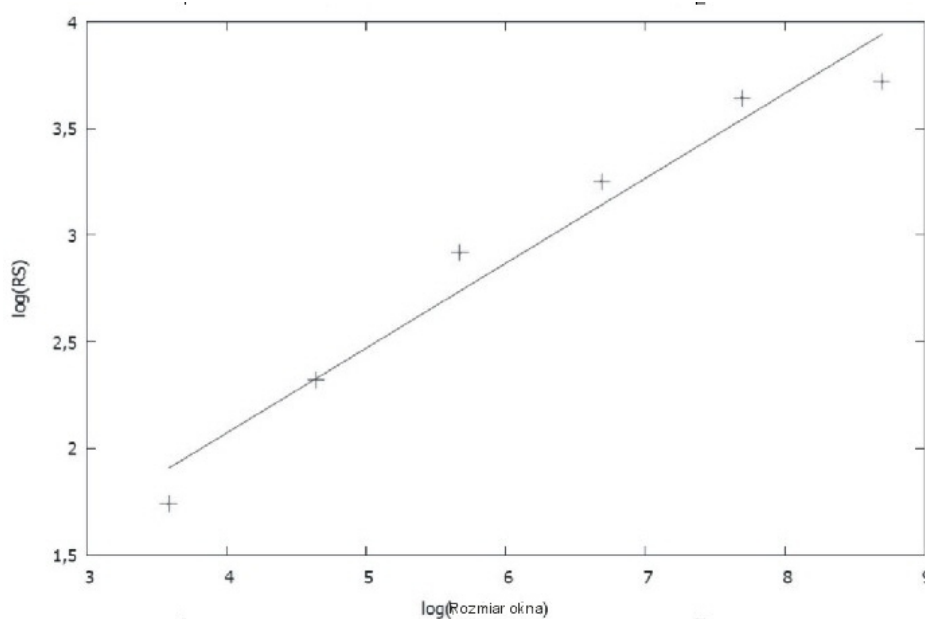


**Rys. 4.19** Skośność (dolna krzywa) i kurtoza (górną krzywą) logarytmicznej zmiany cotygodniowej sprzedaży płyt 30 najpopularniejszych artystów w okresie 2003-2011 [65].

Wykres logarytmicznej zmiany sprzedaży płyt jest na ogół ujemny i oscyluje wokół wartości:  $-0,105$  (Rys. 4.18). Wynika to z faktu, że sprzedaż albumu z czasem maleje i jest zwykle najwyższa w pierwszym tygodniu tuż po premierze. Między wrześniem 2003 a grudniem 2011 trzydziestu najpopularniejszych artystów wydało 203 albumy, dlatego

wyższe momenty, takie jak: odchylenie standardowe (od logarytmicznej zmiany cotygodniowej sprzedaży płyt) i kurtoza zmieniały się gwałtownie z czasem (Rys. 4.18, Rys. 4.19).

Więcej informacji na temat przygotowanego wspólnie z Andrzejem Jarynowskim stochastycznego modelu rynku fonograficznego przedstawię w następnym rozdziale, w którym wyjaśnię, jakie parametry kierują zachowaniem przedstawionym na Rys. 4.4, oraz co stoi za cyklem życia artysty jako produktu (zob. Rys. 5.1 w Rozdz. 5). Wykresy te pokazują bowiem, że sprzedaż płyt gwałtownie rośnie w chwili premiery, a następnie opada w sposób podobny do wykładniczego, a zatem punktem wyjścia dla modelu może być powracający do średniej ruch Browna i zjawiska analogiczne do trzęsień ziemi [69].



**Rys. 4.20** Zakres skalowania użyty do obliczenia wykładnika Hursta dla serii logarytmicznej zmiany cotygodniowej sprzedaży płyt 30 najpopularniejszych artystów w okresie 2003-2011 [65].

Premiera nowego albumu, zwłaszcza firmowanego przez uznanego i popularnego artystę może powodować na rynku silne zaburzenie, które może się odbić nie tylko na ogólnej liczbie sprzedanych na rynku płyt, ale również może powodować zmiany sprzedaży płyt innych wykonawców (najczęściej negatywne z uwagi na dominację ujemnych korelacji, co jest widoczne w Tab. 4.2). Taka perturbacja – dopływ nowego towaru na rynek – może więc przypominać efekt po dolaniu cieczy do zbiornika, kiedy w cieczy tej pojawiają się



turbulencje. Czy zatem następujące po sobie premiery płytowe ustawiają rynek płytowy w stanie podobnym do płynu w przepływie turbulentnym? Zagadnienie to samo w sobie jest skomplikowane, więc ograniczyłem się do zmierzenia długozasięgowej pamięci średniej logarytmicznej zmiany cen  $P(t)$  poprzez obliczenie wykładnika Hursta  $H$ . Wykładnik ten jest miarą persystencji sygnału, a tradycyjna jego definicja opiera się na metodzie przeskalowanego zakresu R/S:

$$E\left[\frac{R(n)}{S(n)}\right] \sim n^H \quad (4.1)$$

gdzie  $E[\dots]$  oznacza wartość oczekiwaną,  $n$  jest długością analizowanego fragmentu sygnału,  $R(n)$  określa zakres jego wartości, a  $S(n)$  – jego odchylenie standardowe. W analizie danych empirycznych wartość wykładnika  $H$  oblicza się z regresji potęgowej funkcji  $n^H$ . Alternatywną metodą określenia wykładnika Hursta jest metoda fluktuacji detrendowanych (ang. *detrended fluctuation analysis*, DFA), w której bada się fluktuacje sygnału po odjęciu od niego trendu  $\Omega$  (najczęściej w postaci wielomianowej). Niech  $S^*(i)$  oznacza całkowity profil sygnału wyjściowego  $S(i)$  w postaci serii czasowej ( $i=1, \dots, T$ ):

$$S^*(i) = \sum_{j=1}^i (S(j) - \bar{S}(j)) \quad (4.2)$$

Sygnał w postaci profilu jest dzielony na  $K$  rozłącznych okien o długości  $s$ , a następnie obliczana jest funkcja fluktuacji:

$$F_2(s) = \frac{1}{K} \left[ \sum_{k=1}^K \frac{1}{s} \sum_{l=1}^s [S^*((k-1)s + i) - \Omega(i)]^2 \right]^{1/2} \quad (4.3)$$

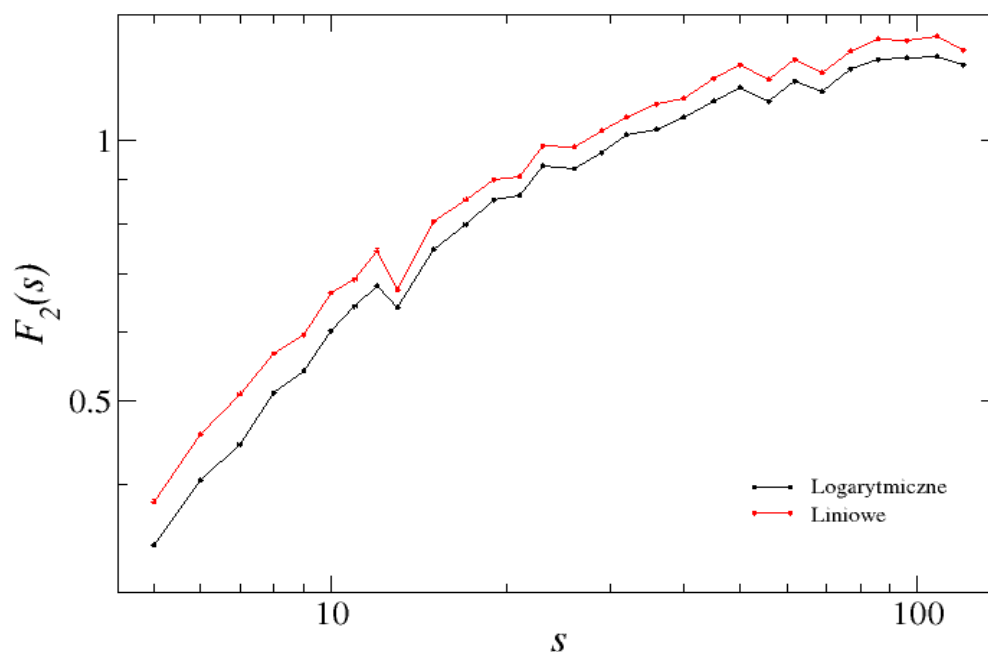
która dla potęgowego sygnału skaluje się analogicznie jak we wzorze (4.1):

$$F_2(s) \sim s^H \quad (4.4)$$

Wartość wykładnika  $H$  można wyliczyć, podobnie jak dla metody R/S – z regresji.

Na Rys. 4.20 pokazałem wyniki uzyskane metodą R/S dla serii czasowej średniej tygodniowej zmiany sprzedaży płyt 30 najpopularniejszych artystów z lat 2003-2011 (Rys. 4.18). Do punktów empirycznych dopasowana została funkcja potęgowa z wykładnikiem  $H=0,39 \pm 0,04$ . Trzeba jednak zauważyć, że przy stosunkowo słabej statystyce (krótki

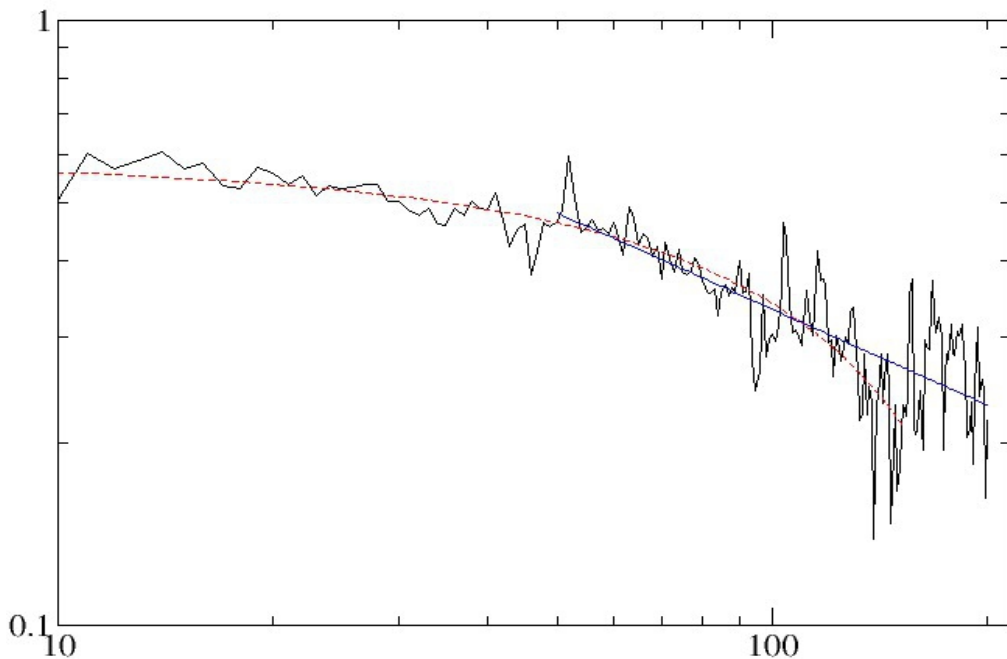
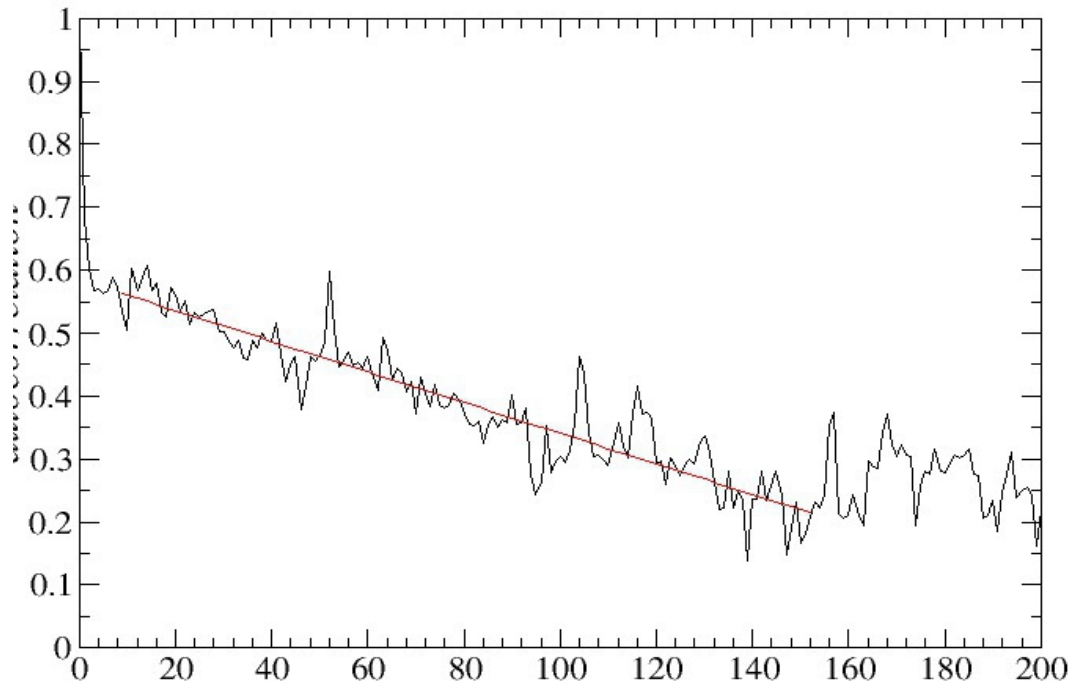
sygnał o długości 500 tygodni i mała liczba artystów) ułożenie punktów na wykresie może nie określać linii prostej, lecz krzywą wklęsłą i wówczas wyznaczenie wykładnika  $H$  byłoby wątpliwe. Uzyskany wynik jest podobny do swojego odpowiednika dla rynku energii [70] i zbliżony do wyliczeń teorii Kołmogorowa, który wykazał, że dla płynu w stanie turbulentnym  $H=0,33$ . Sama turbulencja jest jednak zjawiskiem wieloskalowym i, choć jej obecność jest rozważana także na rynkach finansowych [71], to potwierdzenie jej istnienia na rynkach towarowych wymaga odrębnego badania.



**Rys. 4.21** Funkcja fluktuacji w zależności od okna czasowego  $s$  dla serii cotygodniowych liniowych (górna linia) i logarytmicznych (dolna linia) przyrostów miejsc na brytyjskiej liście sprzedaży singli dla zespołu The Beatles w latach 1962-2012.

Wobec tego, aby mieć szerszy wgląd w sytuację, wybrałem dwa wyjścia. Po pierwsze, podobnie jak w Rozdz. 2 w przypadku długiej serii danych dla pary IBM-GE, wybrałem najstarszego artystę, obecnego na listach sprzedaży od ponad 50 lat. Jest nim zespół The Beatles, który na brytyjskiej liście najlepiej sprzedawanych singli pojawił się już w 1962 roku i jako jedyny do dziś nie utracił popularności. A ponieważ w przeszłości nie mierzono dokładnie cotygodniowej listy sprzedanych egzemplarzy, badałem serię czasową stworzyłem jedynie w oparciu o miejsca The Beatles na liście najlepiej sprzedawanych singli w Wielkiej Brytanii w latach 1962-2012. Przy czym brałem pod uwagę również solowe nagrania członków tego zespołu (Paul McCartney, John Lennon, George Harrison,

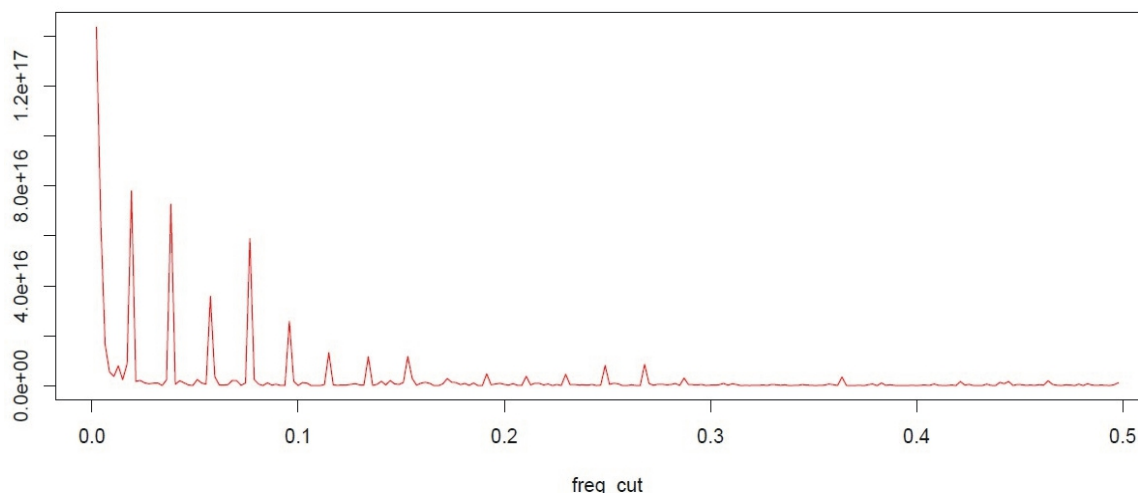
Ringo Starr), a w sytuacji obecności wielu nagrań na liście jednocześnie sumowałem ich wyniki. W efekcie otrzymałem znacznie dłuższą serię czasową, która jednak nie okazała się fraktalna. Obliczona metodą DFA funkcja fluktuacji w zależności od okna czasowego  $s$  okazuje się wyraźnie wklęsła, zarówno dla liniowych, jak i logarytmicznych przyrostów miejsc na listach sprzedaży. Oznacza to, że wykładnik Hursta jest nieokreślony (Rys. 4.21).



**Rys. 4.22** Funkcja autokorelacji w zależności od okna czasowego  $s$  dla serii cotygodniowej sumy 50 najlepiej sprzedawanych płyt w czasie 2003-2012. Na dolnym wykresie log log zaznaczona prosta ma nachylenie  $\nu = 0,52$ .

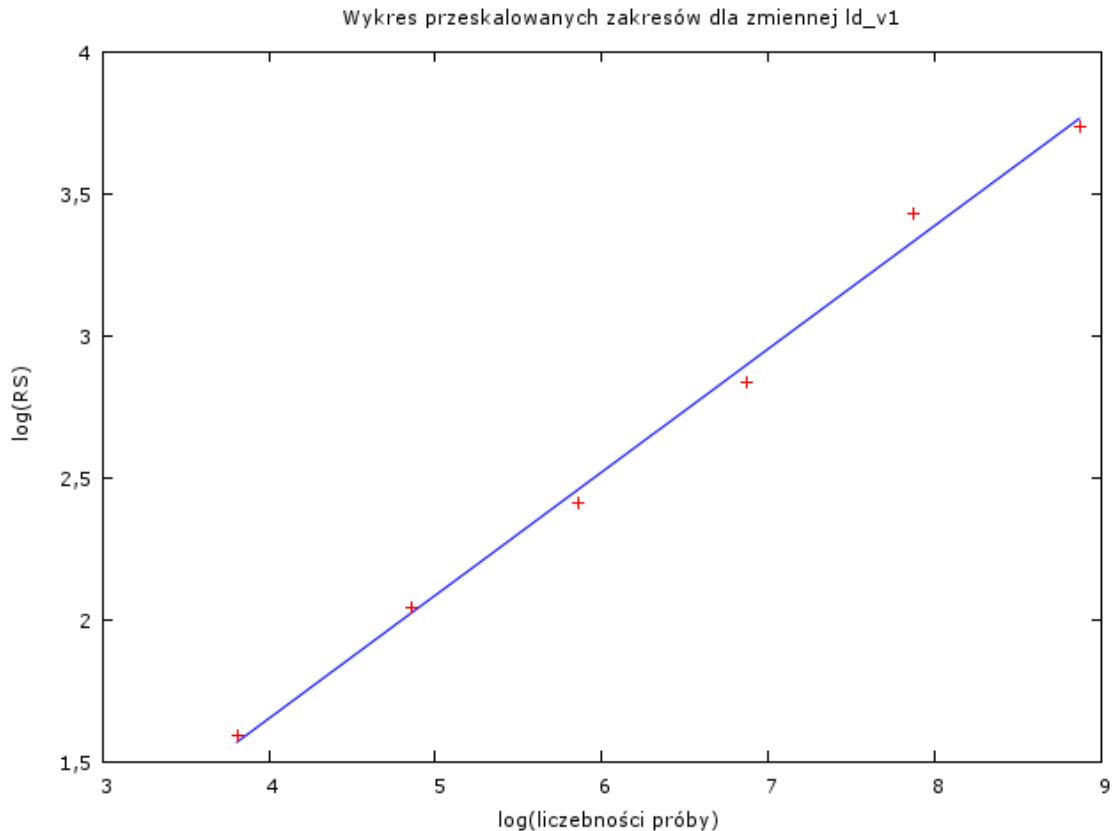
Warto jednak dodać, że branie pod uwagę jedynie liniowych zmian miejsc na listach sprzedaży ma wady, ponieważ różnice w sprzedaży między pierwszym a drugim miejscem są na ogół znacznie wyższe niż np. między 75 a 76 miejscem. Ponadto nie uwzględniamy efektu sezonowości, z powodu którego płyta z pierwszego miejsca w okresie przedświątecznym ma na ogół 3 razy wyższą sprzedaż od płyty, która zajmowała 1 miejsce wiosną czy latem.

Drugą próbą było zastosowanie funkcji autokorelacji do serii, w której sygnałem są cotygodniowe dane o sprzedaży płyt 2003-2012, ale dla sumy wszystkich płyt sprzedanych w danym tygodniu na świecie (w praktyce bierzemy pod uwagę jedynie 50 najlepiej sprzedawanych płyt, bo one stanowią większość globalnej sprzedaży). Okazało się, że zanik autokorelacji ma charakter liniowy w zakresie opóźnień od 10 do 150 tygodni, a później wartość funkcji stabilizuje się. Gdyby w skali podwójnie logarytmicznej połączyć te obszary w jeden i dopasować funkcję potęgową, wówczas otrzymalibyśmy potęgowy zanik funkcji autokorelacji z wykładnikiem  $\nu=0,52$ , co dowodzi istnienia jedynie słabych korelacji długookresowych (Rys. 4.22). Warto zauważyć, że na górnym wykresie występują okresowe maksima dla opóźnień wynoszących 52, 104 i 156 tygodni, co odzwierciedla sezonowość badanej serii. Zależność tę potwierdza także wykres transformaty Fouriera dla tego sygnału (Rys. 4.23).



..

**Rys. 4.23** Wykres transformaty Fouriera dla sygnału cotygodniowej sumy sprzedaży 50 najpopularniejszych płyt za dany tydzień na podstawie serii danych od września 2003 do grudnia 2013.



**Rys. 4.24** Zakres skali przy obliczaniu wykładnika Hursta dla serii logarytmicznej zmiany cotygodniowej sumy sprzedaży 50 najpopularniejszych płyt za dany tydzień w okresie 2003-2011

Nieco bardziej dokładny wynik otrzymałem, badając serię złożoną z tygodniowych zmian logarytmu sumy sprzedaży wszystkich 50 najlepiej sprzedawanych płyt w danym tygodniu od września 2003 do grudnia 2012, ale po wcześniejszym usunięciu trendu (Rys. 4.6) i unormowaniu sygnału tak, aby średnia sprzedaż w danym tygodniu wynosiła 1. Wówczas dla przyrostów logarytmu sprzedaży (czyli odpowiednika stopy zwrotu) wyliczony tradycyjną metodą wykładnik Hursta wynosi  $0,43 \pm 0,01$  (Rys. 4.24). Wynik ten, podobnie jak uzyskana dla 30 artystów wartość 0,39 (Rys. 4.20), jest zbliżony do wartości 0,41, uzyskanej na innym rynku towarowym – cen energii [38].

Funkcja autokorelacji zmian logarytmu sprzedaży pozwala określić wzajemną niezależność kolejnych skoków sprzedaży płyt. Dla stacjonarnych procesów stochastycznych charakterystyka ta jest równoważna gęstości spektralnej samego przebiegu sygnału w czasie, co przedstawię w następnym rozdziale (Rys. 5.13). Rzeczywiście, dla niektórych procesów gęstość spektralna  $S(f)$  jest odwrotnie proporcjonalna do kwadratu częstotliwości  $f$ . Tak jest np. w przypadku błędzenia losowego.

Na rynkach akcji funkcja autokorelacji zmian logarytmu ceny jest funkcją szybko zanikającą, odznaczającą się czasem korelacji dużo krótszym niż czas dziennej aktywności giełdy [72]. Z definicji wykładnika Hursta wynika, że dla nieskorelowanych (niezależnych) zmiennych jego wartość jest równa 0,5. Uzyskany więc dla całkowitej tygodniowej sprzedaży płyt wykładnik  $\nu=0,53$  wskazuje na istnienie bardzo słabych dodatnich korelacji długookresowych, co oznacza słabą persystencję.

Podobne wartości otrzymano wcześniej podczas badania rynków walutowych [73] oraz rynków akcji [74]. Jednak w przypadku sprzedaży płyt seria czasowa z lat 2003-2012 jest zbyt krótka, ze względu na dostępność dokładnych danych obejmując mniej niż 500 tygodni. Dlatego warto takie badanie przeprowadzić w przyszłości, gdy będą już dostępne dłuższe (i, być może, częściej próbkowane) serie czasowe i analiza będzie mogła zostać rozszerzona na takie zagadnienia, jak np. rodzaj dyfuzji związanej z procesem sprzedaży płyt, podobnie, jak to się robi w przypadku rynków finansowych [72].

## **.Rozdział 5. Modelowanie i dynamika rynku fonograficznego. Cykl życia produktu.**

### **5.1 Cykl życia produktu w przemyśle płytowym. Dynamika sprzedaży.**

Jak pokazałem w poprzednich rozdziałach, rynek fonograficzny jako rynek towarowy jest bardziej przewidywalny od rynków finansowych. Przewidywalność rynków finansowych silnie ogranicza hipoteza efektywnego rynku, według której każda informacja, która dostępna jest dla uczestników rynku, natychmiast (w przeciągu co najwyżej minut) znajduje swoje odbicie w cenie [57]. Na rynku fonograficznym tak rozumiana efektywność jest oczekiwana w mniejszym stopniu, choć napływające na ten rynek informacje branżowe także mają duży (i do pewnego stopnia ograniczony w czasie) wpływ na sprzedaż. Ponadto analizowana w poprzednim rozdziale sezonowość sprzedaży płyt oraz istniejąca baza fanów konkretnego artysty, którzy zawsze kupują jego nowe albumy, pozwala przewidywać dynamikę sprzedaży, przynajmniej pod względem jakościowym. To otwiera drogę do opisu dynamiki sprzedaży płyt przy wykorzystaniu odpowiednio zaprojektowanych modeli.

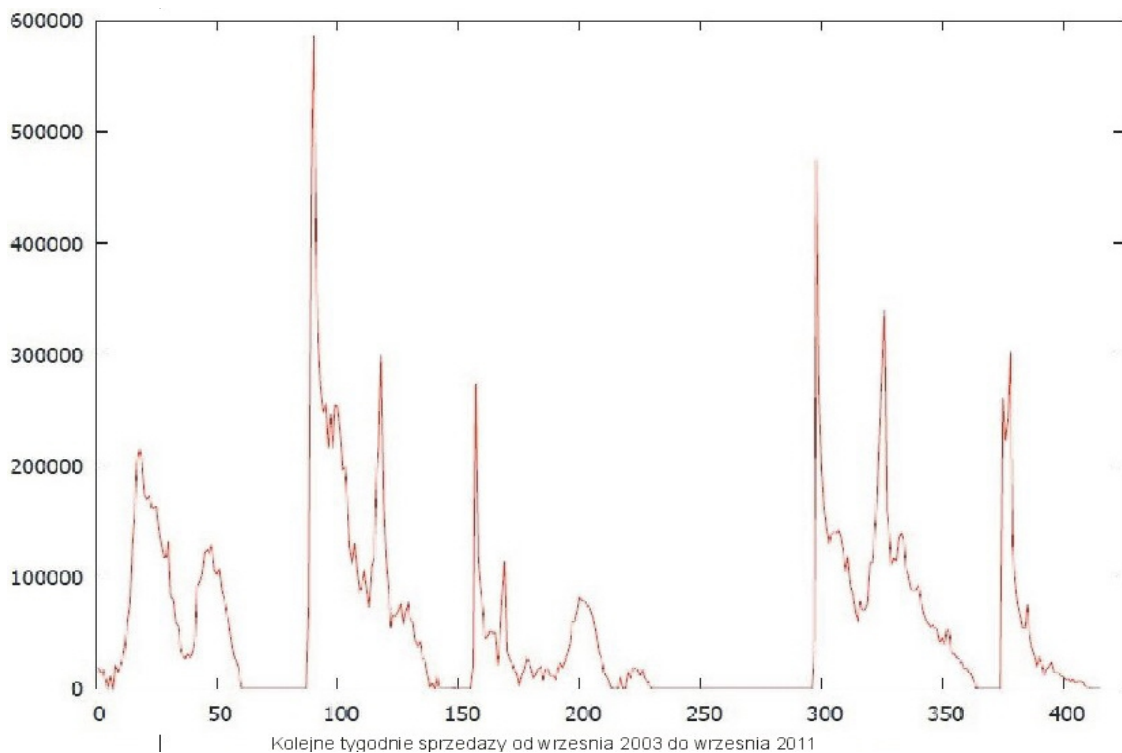
Zjawiska zachodzące na rynkach finansowych są zazwyczaj modelowane przy wykorzystaniu modelowania agentowego [75], modelowania przy pomocy procesów stochastycznych [76], a niekiedy także metodami analitycznymi (np. [77]). Modelowanie agentowe stanowi kolejny etap moich badań, natomiast modelowanie stochastyczne jest przedmiotem tego rozdziału.

W przypadku modelowania cen na rynkach finansowych najczęściej wykorzystywane są procesy heteroskedastyczne (ARCH/GARCH i modele pokrewne [78]), procesy skokowe typu przelotów Levy'ego, czy procesy kaskadowe [79], dzięki którym możliwe jest odtworzenie wielu własności statystycznych ruchów cen, w tym także tych, tworzących tzw. finansowe fakty stylizowane (grube ogony rozkładów stóp zwrotu, długookresowa pamięć, multifraktalność). Procesy te jednak stosunkowo słabo spisują się w przypadku prób prognozowania, nawet jeśli efektywność rynku może być nie do końca zapewniona [80-81].

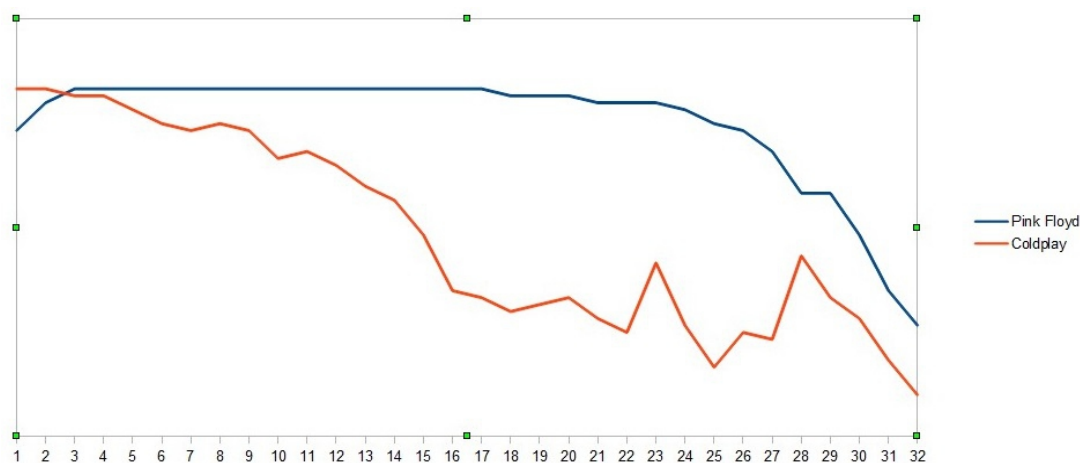
Obserwacje empiryczne – widoczne choćby na Rys. 5.1 – pokazują, że współcześnie sprzedaż płyt danego wykonawcy osiąga maksimum dokładnie w tygodniu premiery jego nowej płyty. Dotyczy to artystów o statusie gwiazdorskim, ponieważ w przypadku artystów dopiero pozyskujących fanów krzywe sprzedaży osiągają maksimum dopiero po ich wypromowaniu. W przeszłości jednak nawet uznane gwiazdy, jak np. Pink Floyd, miały trajektorie sprzedaży swoich płyt podobną do debutantów, gdyż informacja o premierze (najczęściej w prasie i radio) rozprzestrzeniała się wolniej niż obecnie przez internet. Odzwierciedla to porównanie trajektorii sprzedaży w USA płyt *The Wall* Pink Floyd oraz *Viva La Vida* grupy Coldplay (Rys. 5.2). Jest to analogiczne do obserwowanego na rynkach akcji efektu Eppsa, czyli wzrostu siły korelacji pomiędzy akcjami wraz ze spadkiem częstotliwości próbkowania danych [49, 82].

W praktyce premiery płytowe odbywają się w pierwszy dzień roboczy nowego tygodnia – a więc w poniedziałki, lub wtorki – sprzedaż za dany tydzień tradycyjnie zliczana jest bowiem od poniedziałku do soboty (lub w niektórych krajach do niedzieli) i publikowana w sobotę następnego tygodnia. Oczywiście na części rynków lokalnych wyniki sprzedaży znane są już wcześniej. Na przykład w Wielkiej Brytanii zliczone są już kilka godzin po zakończeniu tygodnia handlowego, a dla wybranych wykonawców możliwe jest nawet oszacowanie sprzedaży ich płyt dzień po dniu. Dane te są jednak tajemnicą handlową wytwórni i nie są publikowane, bo służą głównie kierowaniu promocji w taki sposób, aby całkowita sprzedaż była największa. Znane są jednak nawet otwarte wojny handlowe, jak np. ta między The Rolling Stones a The Beatles (1967) oraz między Blur a Oasis (1995). Wojny takie najczęściej dotyczą singli, natomiast jeśli chodzi o sprzedaż albumów, to największe koncerty (EMI, Universal, Warner, Sony BMG), dążąc do maksymalnej sprzedaży, wolą jednak ustalać daty premier płytowych odgórnie, nawet w porozumieniu między sobą, często z wielomiesięcznym wyprzedzeniem.





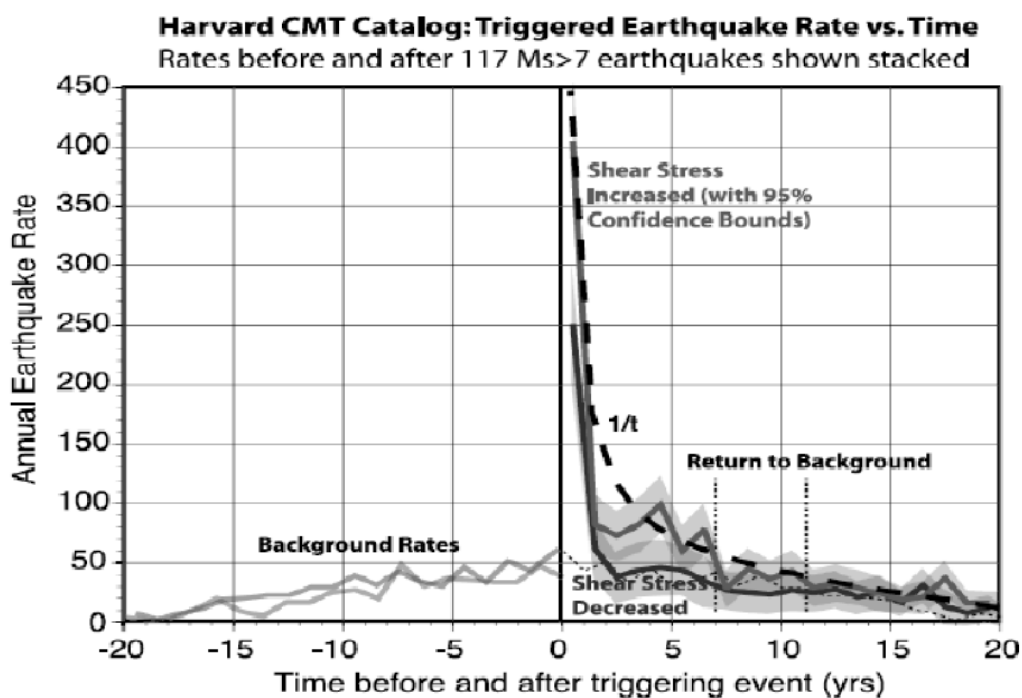
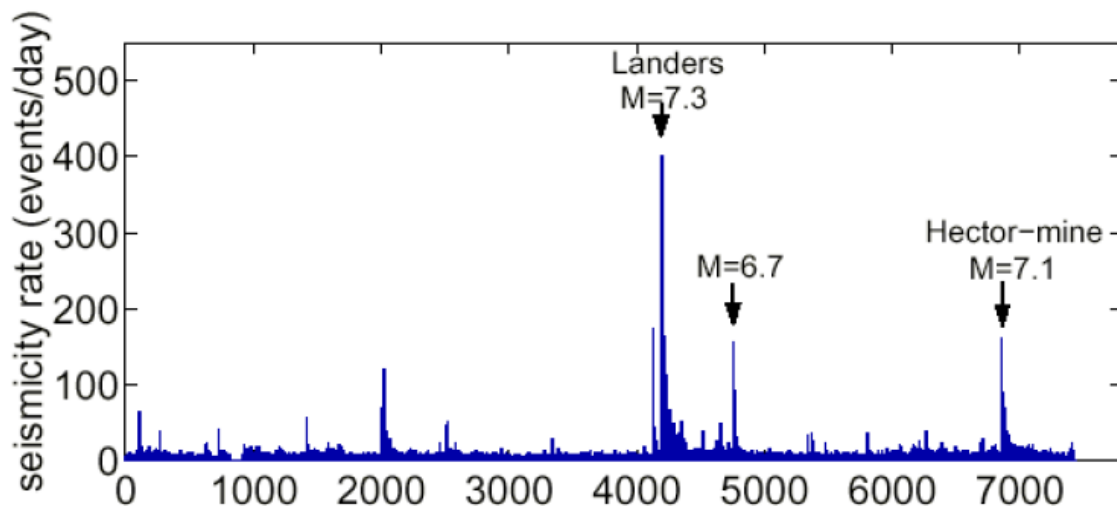
**Rys. 5.1** Cykl życia produktu: sprzedaż płyt zespołu Black Eyed Peas od września 2003 do września 2011 [65].



**Rys. 5.2** Cykl życia produktu: porównanie trajektorii obecności na listach sprzedaży płyt *The Wall* (1979) zespołu Pink Floyd (górną linią) oraz *Viva La Vida* (2008) grupy Coldplay (dolną) w kolejnych tygodniach od ich premiery. Każda z nich osiąga maksimum, czyli miejsce 1. listy sprzedaży w USA w innym odstępie od premiery.

Premierze albumu najczęściej towarzyszy promocyjny singiel, którego obecność w mediach podnosi sprzedaż, przez co łatwiej ją kontrolować, gdyż tuż po premierze obserwujemy spadek sprzedaży, podobny do wykładniczego (Rys. 5.1). Oczywiście wytwórnie chcą mu zapobiec, dlatego wydają kolejne single wykrojone z albumu i w ten sposób ponownie podnoszą, choć na krócej, sprzedaż. To również widać na Rys. 5.1 w kształcie opadania krzywej po premierach płytowych.

Podobne zjawisko gwałtownego skoku wartości sygnału i proces powracania jej do średniej wraz z pojawiającymi się później wstrząsami wtórnymi obserwujemy także w trzęsieniach ziemi (Rys. 5.3). Wstrząsy te na rynku fonograficznym są związane z wydawaniem kolejnych singli podnoszących sprzedaż płyty).

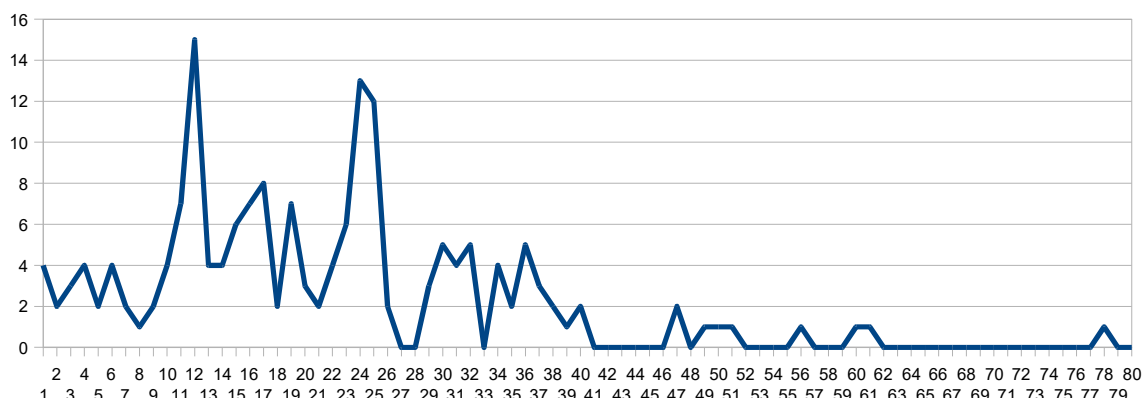


**Rys. 5.3** Analogia pomiędzy sprzedażą płyt a trzęsieniami ziemi. Wykres górny: współczynnik sejsmiczności (liczba wstrząsów na dzień) w zależności od czasu. Wykres dolny: zależność liczby wstrząsów od czasu przed i po głównym wstrząsie [69].

Podobieństwo między dynamiką sprzedaży (Rys. 5.1) a dynamiką wstrząsów (Rys. 5.3) obserwujemy nie tylko, jeśli chodzi o gwałtowne skoki w chwili wstrząsu lub premiery płytowej i ich zaniki przerywane wstrząsami wtórnymi. Zdarza się bowiem, że tuż przed

wstrząsem głównym obserwowane są wstrząsy poprzedzające, czyli wstrząsy o małej sile, które pojawiają się (lub nie) przed wstrząsem głównym [69]. Podobne zjawisko zachodzi także na rynkach energii [83] i, oczywiście, na rynku fonograficznym. Wskutek uruchomienia promocji, a jeszcze przed dostarczeniem premierowego materiału na rynek, obserwujemy bowiem wzrost sprzedaży wszystkich poprzednich tytułów danego artysty. Zjawisko to zachodzi jedynie u niektórych artystów (np. nie zachodzi ono u debiutantów), podobnie jak tylko niektóre trzęsienia ziemi są poprzedzone drobnymi wstrząsami. Wstrząsy poprzedzające i wtórne mają związek z budową i reakcjami chemicznymi, które zachodzą wewnątrz ziemi, podobnie jak wtórne skoki sprzedaży są wynikiem mechanizmów promocyjnych funkcjonujących na rynku płytowym.

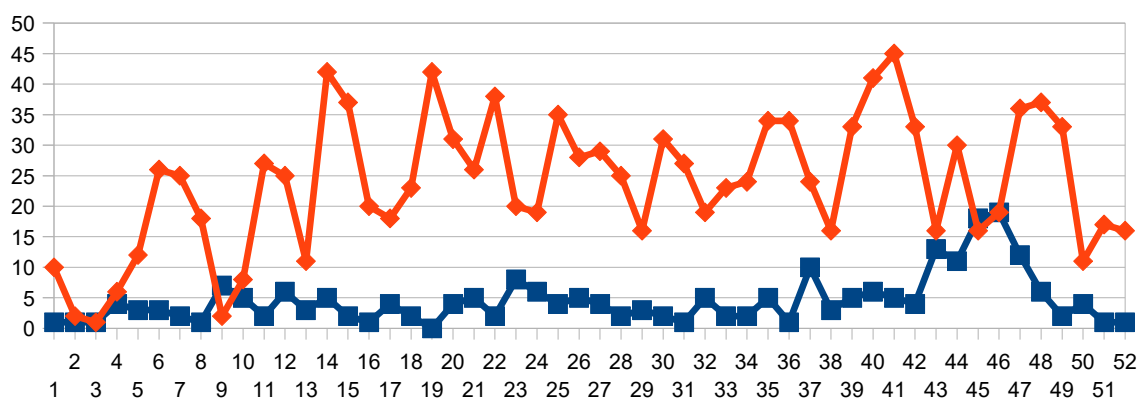
Zjawisko trzęsień ziemi znalazło swój opis i modele w literaturze [69, 84-88], jednak problemem jest przewidzenie, w którym miejscu i czasie główny wstrząs nastąpi. Takiego problemu nie mamy dla rynku płytowego, bowiem daty premiery – a więc maksimum sprzedaży, czyli „wstrząsów głównych” - są ustalane odgórnie przez wytwórnę płytową wiele tygodni przed premierą. Na ogół najpopularniejsi artyści wydają kolejne nowe albumy co 2 lata (Rys. 5.4), jednak 40 lat temu powszechne było wydawanie albumów co roku. Ta historyczna zmiana ma związek z pojemnością nośnika, gdyż do połowy lat 80-tych na rynku dominowała płyta winylowa, która w wersji długogrającej zawierała do 40 minut muzyki. Natomiast wprowadzone 40 lat temu na rynek płyty kompaktowe mogą pomieścić do 80 minut nagrań, co przy założeniu, że artyści są tak samo płodni, w naturalny sposób wydłużyło odstęp między premierami płytowymi [59]. W efekcie wytwórnie, aby skrócić czas oczekiwania na nowy album, wydają w międzyczasie składanki największych przebojów artysty bądź jego nagrania koncertowe – najczęściej pod koniec roku. W ten sposób dostosowują się do sezonowości rynku. Rys. 5.4 przedstawia mierzony w miesiącach rozkład odstępów między kolejnymi premierami płytowymi i wykazuje maksima dla odstępów 12, 24, 36 i 48 miesięcy. Obserwujemy zatem okresowość, będącą wielokrotnością długości sezonu (12 miesięcy). Jedynie odstęp 25 miesięcy występuje porównywalnie często, co uznałem za najczęstszą naturalną przerwę między kolejnymi albumami artysty, wynikającą z procesu twórczego.



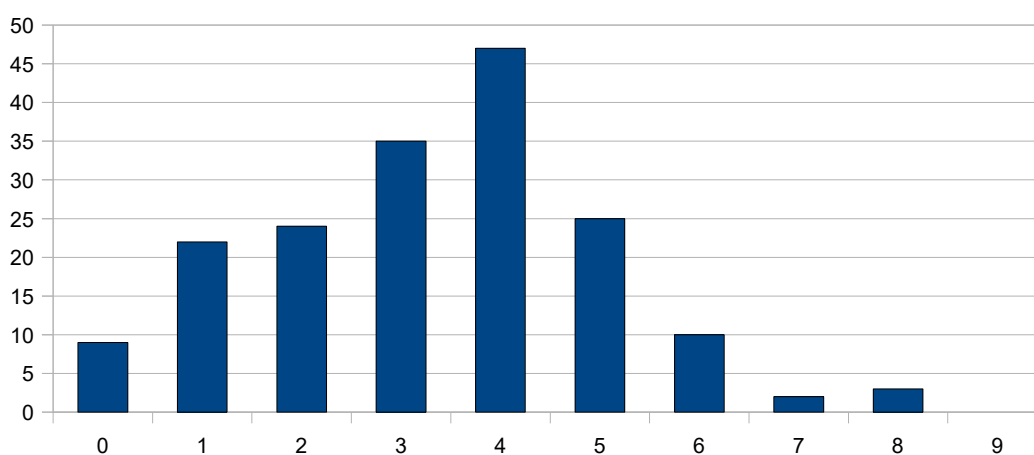
**Rys. 5.4** Rozkład odstępów między premierami płytowymi podany w miesiącach. Pod uwagę wziąłem daty premiery 206 albumów 30 najlepiej sprzedających się artystów z lat 1996-2013.

Równie ciekawie wygląda rozkład premier płytowych w całym sezonie (Rys. 5.5). Okazuje się, że wytwórnie płytowe, aby zwiększyć sprzedaż, planują premiery płyt gwiazd najczęściej w miesiącach od września do listopada, gdzie obserwujemy natężenie premier 3-krotnie większe niż w pozostałej części roku, kiedy ich rozkład jest równomierny. Planowanie to ma związek z faktem, iż kilka tygodni przed świętami Bożego Narodzenia do rynku dołączają się konsumenci, którzy kupują płyty „od święta” lub na prezent, a ich znajomość artystów ogranicza się jedynie do najpopularniejszych gwiazd. Wysokie natężenie gwiazdorskich premier w okresie jesieni jest zatem jedną z przyczyn powstania opisanego w poprzednim rozdziale „sektora gwiazdorskiego” (Rys. 4.10 i 4.14), gdyż wysoka sprzedaż zanotowana w tym samym momencie powoduje korelację między największymi gwiazdami, nawet jeśli należą one do zupełnie innych gatunków muzyki. Co więcej, przynależność do estetycznie różnych gatunków powoduje, że gwiazdy te nie są dla siebie konkurencją, bo ich płyty kupują konsumenci o różnych gustach.

Jeśli chodzi o rozkład premier płytowych dla pozostałych artystów, to jest on zadziwiająco regularny z częstotliwością około 1 miesiąca. Prawidłowość ta może mieć związek z dochodami konsumentów, o których mówiłem przy okazji Rys. 4.9 w poprzednim rozdziale. Na świecie większość nabywców płyt na ogół otrzymuje bowiem wynagrodzenie co miesiąc, a wytwórnie, planując premiery płytowe dla konsumentów, prawdopodobnie biorą ich dochody pod uwagę.



**Rys. 5.5** Rozkład premier płytowych 30 największych gwiazd (dolna linia) i pozostałych artystów (górną linia) w całym sezonie.

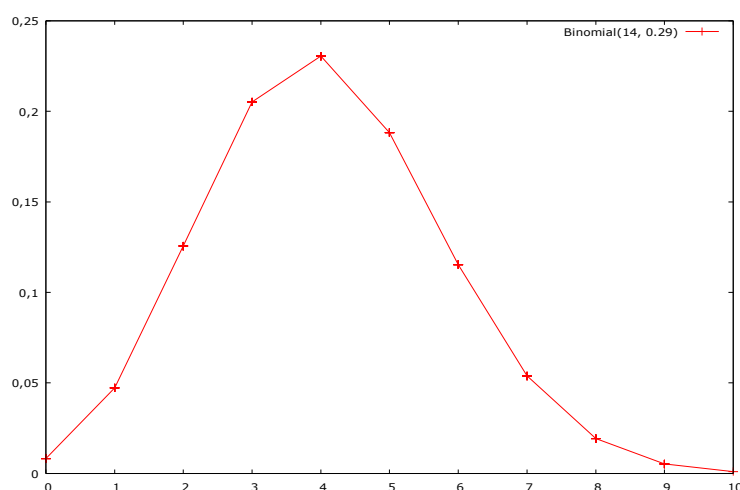


**Rys. 5.6** Rozkład liczby singli przypadających na jeden promowany album na podstawie danych dla 30 najpopularniejszych artystów z lat 1995-2013.

Kolejnym istotnym parametrem jest promocja, która w uproszczeniu sprowadza się do liczby przebojów z danego albumu, które trafiają do rozgłośni radiowych, muzycznych stacji telewizyjnych oraz do Internetu, podnosząc sprzedaż. W praktyce, jeśli pierwszy singiel okaże się przebojem, wytwórnia promuje album następnym – tak długo, aż jest zapotrzebowanie. W momencie, gdy singiel okazuje się chybionym przebojem, płyta spada z listy albumów, a promocja ustaje.

Z tego powodu dla najpopularniejszych artystów rozkład singli przypadających na album przypomina rozkład dwumianowy i osiąga maksimum dla  $k=4$  (Rys. 5.6). W praktyce ten rozkład jest ograniczony przez  $n$ , czyli liczbę utworów na płycie, ponieważ album ma ograniczony czas trwania i może zawierać co najwyżej  $n$  utworów, z których zostaje wybranych  $k$  singli (zakładamy, że zachodzi relacja  $k < n$ ).

Spośród rozkładów dwumianowych odzwierciedlających liczbę singli przypadających na album (Rys. 5.6) najlepiej dopasowany jest ten o parametrach  $p=0,29$  i  $n=14$  (Rys. 5.7). Oznacza to, że liczbę doświadczeń  $n=14$  możemy interpretować jako całkowitą liczbę utworów na albumie CD (wynikającą z pojemności nośnika), a  $p=0,29$  jest prawdopodobieństwem, że piosenka z płyty zostanie wydana na singlu. W efekcie wytwórnie promują albumy tak długo, aż jest zapotrzebowanie, które mija najczęściej po czwartym singlu. Co ciekawe, rozkład dwumianowy (Rys. 5.7) doskonale pokrywa się z rozkładem z Rys. 5.6 także w obszarze  $k > 8$  i zanika do zera pomimo tego, że płyta CD zawiera kilkanaście utworów, bowiem rozkład dwumianowy w tym obszarze przyjmuje wartości bliskie zero. W praktyce oznacza to, że tylko kilka z kilkunastu piosenek może promować album.

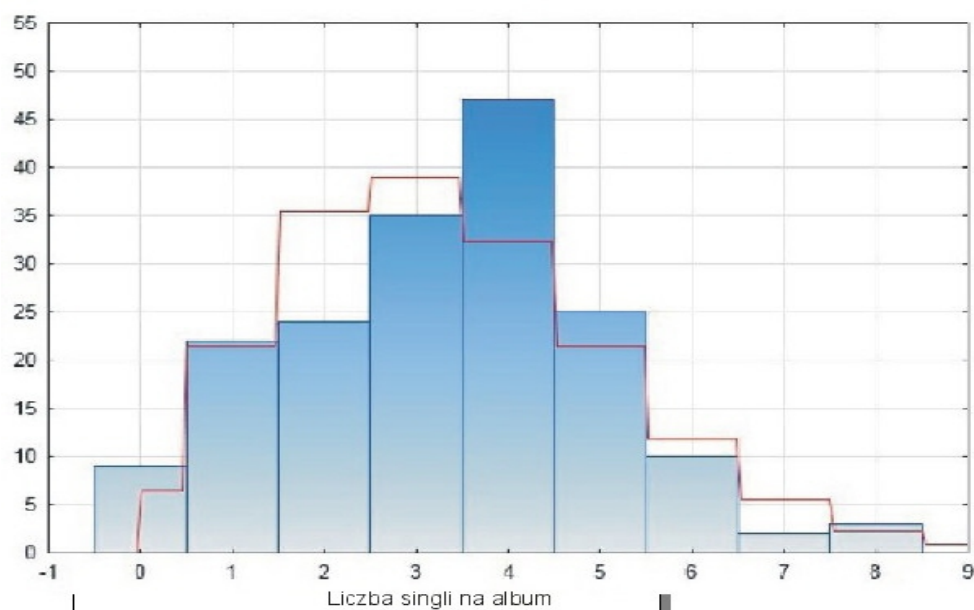


**Rys. 5.7** Gęstość prawdopodobieństwa  $P(k)$  odpowiadająca rozkładowi dwumianowemu o parametrach  $n=14$  i  $p=0,29$ .

Liczba singli przypadająca na jeden album oczywiście wpływa na czas trwania promocji, który zwykle równa się liczbie tygodni, które album danego wykonawcy spędza na liście sprzedaży od chwili premiery. Wytwórnie wychodzą bowiem z założenia, że warto promować te produkty, które się dobrze sprzedają. Stąd płyty niszowych, awangardowych artystów nie są promowane w ogóle, a płyty Michaela Jacksona *Thriller* (1982) i *Bad* (1987) doczekały się, odpowiednio, 7 i 9 przebojowych singli, powodując wzrosty sprzedaży macierzystych albumów i przedłużając ich obecność na listach do kilku lat.

Oczywiście promocja to nie tylko przeboje, ale także trasy koncertowe i szereg innych działań marketingowych. Zwłaszcza że o obecności na listach może decydować jeszcze

jeden parametr: wyjątkowa jakość produktu. W przeszłości zdarzało się więc, że materiał promowany tylko jednym singlem utrzymywał się na listach sprzedaży przez kilkanaście lat – jak rekordzistka *Dark Side Of The Moon* grupy Pink Floyd. Jakość płyt mierzona ocenami recenzentów istnieje w przemyśle płytowym od 1971 roku, gdy amerykański magazyn *Village Voice* jako pierwszy opublikował listę najlepszych płyt roku na podstawie opinii zespołu krytyków. Wcześniej płyty oceniane były przez pojedynczych recenzentów i miały mały wpływ na odkrywanie nowych artystów i sprzedaż. Od 40 lat układaniem najlepszych albumów roku na podstawie ich jakości zajmują się branżowe gazety, takie jak *New Musical Express*, *Rolling Stone* czy *Spin*, a od kilkunastu lat także serwisy internetowe w rodzaju serwisu Pitchfork [59].



**Rys. 5.8** Rozkład singli przypadających na album (słupki pełne) w porównaniu do gęstości prawdopodobieństwa  $P(k)$ , odpowiadającej rozkładowi Poissona o parametrze  $\lambda = 3.3073$  (słupki puste) [89].

Fakt istnienia innych niż single elementów promocji uwzględniłem, zastępując rozkład dwumianowy bardziej adekwatnym rozkładem Poissona o identycznej wartości oczekiwanej (Rys. 5.8). Promocja bowiem to nie tylko  $k$  singli z  $n$  piosenek na albumie, ale także trasy koncertowe, w których otrzymujemy nowe wykonania piosenek z albumu, starsze utwory, wykonania piosenek z repertuaru innych artystów itp. W efekcie liczba  $n$  jest na tyle duża, że bezpiecznie możemy założyć rozkład Poissona zamiast rozkładu dwumianowego.

Wybór ten oczywiście przełoży się na czas trwania promocji, ponieważ zakładam, że pierwszy singiel promocyjny ukazuje się równocześnie z albumem, a kolejne, losowane

zgodnie z rozkładem Poissona single, pojawiają się średnio w odstępie 10 tygodni. Takie uogólnione założenie wynika z reguł przyjętych na największym na świecie rynku płytowym – w Stanach Zjednoczonych, gdzie ze względu na wielki rozmiar kraju wytwórnie płytowe wydają i promują single równocześnie z albumem. Reguły te obowiązują także w innych krajach o dużej powierzchni, takich jak Kanada i Australia, jednak warto zauważyć, że inne wiodące rynki (np. Wielka Brytania) mają odmienną tradycję, gdyż tam singiel promocyjny wydawany jest zwykle na tydzień przed premierą albumu [59].

Zakładam ponadto, że premiera płytowa odbywa się jednocześnie na wszystkich rynkach lokalnych, co dla sprzedaży płyt 30 największych gwiazd jest zgodne z rzeczywistością. Obecnie bowiem, w dobie Internetu, we wszystkich krajach świata obowiązuje ten sam dzień premiery, co najwyżej przesunięty o dobę ze względu na strefę czasową. Tymczasem dawniej, choćby 40 lat temu, występowały spore różnice w dacie premier na rynkach: brytyjskim i amerykańskim nawet dla płyt największych gwiazd, sięgające niekiedy jednego roku [59]. Natomiast dla pozostałych artystów daty premier w różnych krajach są zróżnicowane, gdyż mogą oni mieć status gwiazdorski jedynie lokalnie. Z tego powodu zostaną oni pominięci w zaproponowanym w następnym podrozdziale modelu, który dotyczy rynku globalnego.

## **5.2 Stochastyczne modelowanie krzywej sprzedaży płyt.**

### **5.2.1 Markowowskie przełączanie stanów. Przypadek stacjonarny.**

Własności rynku przedstawione w Rozdz. 4 i Rozdz 5.1 sugerują, aby model dynamiki sprzedaży oprzeć na stochastycznych procesach typu markowowskiego przełączania (ang. *Markov switching regime*, MSR). Procesy tego typu wydają się naturalnymi kandydatami do modelowania gwałtownych skoków nieliniowego zachowania sprzedaży płyt. Model, który przedstawiam poniżej, powstał we współpracy z Andrzejem Jarynowskim [89].

Na podstawie krzywej sprzedaży płyt (Rys. 5.1) można zauważyć, że sprzedaż płyt danego artysty może być scharakteryzowana przez dwa stany: podstawowy, gdy nie trwają żadne



kampanie promocyjne, oraz wzbudzony, gdy artysta wydaje nowy album, któremu towarzyszy promocja. Zakładamy, że artysta w stanie podstawowym sprzedaje płyty w sposób przypadkowy, a dane o sprzedaży przypominają szum. Premiera nowego albumu i promocja sprawiają, że sprzedaż albumu gwałtownie rośnie i reakcja konsumentów ma charakter kolektywny. Zachowanie się krzywej sprzedażowej można opisać przez proces, w którym sygnał, po wytrąceniu go ze stanu równowagi, charakteryzowanego przez średnią wartość  $a$ , ma tendencję do powolnego powrotu do tego stanu. Pożądaną własność posiada proces powracającego do średniej, quasi-geometrycznego błędzenia losowego (ang. *mean-reverting quasi-geometric random walk*, MrqGRW), znanego także pod nazwą powracającego do średniej quasi-geometrycznego ruchu Browna (ang. *mean-reverting quasi-geometric Brownian motion*, MrqGBM).

Niech  $X_t$  oznacza proces stochastyczny odpowiadający krzywej sprzedaży płyt danego artysty. Ruch Browna  $W_t$  jest odpowiedzialny za relatywnie słabe (proporcjonalnie do stałej  $s$  pomnożonej przez aktualną wartość procesu  $X_t$ ) fluktuacje wokół długozasięgowej średniej  $a$ :

$$dX_t = (a - bX_t)dt + sX_t dW_t \quad , \quad (5.1)$$

co w wersji z dyskretnym czasem przy warunku  $dt = 1$  (tydzień) przyjmuje postać (rozwiniecie Eulera):

$$X(t + 1) = X(t) + a - bX(t) + sX(t)N(0, 1) \quad , \quad (5.2)$$

gdzie  $N(0,1)$  jest zmienną losową z rozkładu normalnego.

Mechanizm MSR opisuje przejście ze stanu podstawowego (1) do stanu wzbudzonego (2) związanego z promocją. Możemy zdefiniować macierz przejścia  $\mathbf{R}$  o elementach  $r_{ij}$ , opisujących prawdopodobieństwo przejścia ze stanu (i) do stanu (j) w chwili  $t$ :

$$\{r_{ij}\} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1 - r_{12} & r_{12} \\ 1 - r_{22} & r_{22} \end{bmatrix} \quad . \quad (5.3)$$

Biorąc pod uwagę opisaną w Rozdz. 5.1 analogię kształtu krzywej sprzedaży w reżimie promocji (Rys. 5.1) do kształtu krzywych aktywności sejsmicznej (Rys. 5.3), założyliśmy (na podstawie własności tych drugich krzywych [88]), że czas trwania stanu promocji jest dany przez zmienną losową z rozkładem wykładniczym i wartością oczekiwaną  $1/r_{22}$ .

Proces obejmujący oba stany możemy zatem opisać następującym równaniem:

$$dX_t = (a - bX_t)dt + sX_t dW_t + j(t)(dQ_t + dq_t) \quad (5.4)$$

gdzie  $j(t) = 1$ , jeśli stan (1) przechodzi w (2), oraz  $j(t) = 0$  w przeciwnym wypadku. Jako że wydanie singli powoduje skokowy wzrost sprzedaży albumu, z którego te single pochodzą (Rys. 5.1), konieczne było wprowadzenie we wzorze (5.4) zmiennych, które będą to odzwierciedlać. Pierwsza z tych zmiennych,  $Q_t$  opisuje efekt promocji po wydaniu albumu i jest wolnozaniakająca (w tym przypadku ma ona rozkład lognormalny), natomiast druga zmienna,  $q_t$ , odpowiada za efekt promocyjny związany z wydawaniem singli i jest szybkozaniakająca (postulujemy dla niej rozkład jednorodny).

Powyższy proces można nazwać powracającym do średniej, quasi-geometrycznym błędzeniem losowym ze skokiem (ang. *mean-reverting jumped quasi-geometric random walk*) lub powracającym do średniej dyfuzyjnym ruchem Browna ze skokiem (*mean-reverting jump diffusion quasi-geometric Brownian motion*).

Prawdopodobieństwo  $p$ , że w ciągu jednostki czasu zostanie wydany singiel, jest określone przez rozkład Poissona (Rys. 5.8), odstępy czasowe pomiędzy chwilami publikacji następujących po sobie singli są natomiast określone przez  $1/p$ . Efekt wydania singla (co zachodzi tylko w stanie (2)) modelujemy w ten sposób, że średnia  $a$ , do której proces powraca, staje się sama w sobie zmienną zależną od czasu i ma postać:  $a(t) = u_t + a$ , gdzie  $a$  jest stałym parametrem, natomiast  $u_t$  ma ten sam rozkład co  $q_t$ . Uzmiennienie  $a(t)$  odzwierciedla fakt, że niektóre albumy mogą się sprzedawać na stałym, wysokim poziomie niezależnie od promowania ich singlami (wyróżniamy wówczas, w ramach stanu (2), także stan (2') - „popularność”. Oznacza on, że artysta dzięki swojej płycie zdobył szeroką publiczność, wykraczającą poza swoich stałych fanów. Zakładam, że stan ten można osiągnąć tylko w chwili wydania singla, który staje się przebojem. Można to formalnie zapisać, że  $r_{22} > 0$  wtedy i tylko wtedy, gdy  $i(t) = 1$ . Macierz przejścia  $\mathbf{R}'$  ma wówczas postać:

$$\{r_{ij}\} = \begin{bmatrix} r_{22} & r_{22'} \\ r_{2'2} & r_{2'2'} \end{bmatrix} = \begin{bmatrix} 1 - r_{22'} & r_{22'} \\ r_{2'2} & 1 - r_{2'2'} \end{bmatrix} = \begin{bmatrix} 1 - pp' & pp' \\ p & 1 - p \end{bmatrix} \quad (5.5)$$

gdzie  $p$  oznacza prawdopodobieństwo wydania singla z albumu, a  $p'$  prawdopodobieństwo, że singiel stanie się przebojem. Układ jest w stanie powrócić do średniej  $a(t)=a$ , jeśli zostanie wydany kolejny singiel (z tym samym prawdopodobieństwem  $p$ ) lub jeśli nastąpi koniec promocji (z prawdopodobieństwem  $r_{21}$ ). A zatem średni czas pobytu w stanie (2') wynosi  $r_{22}/p$ .

Mogę teraz zatem napisać ogólną postać równania opisującego analizowany proces:

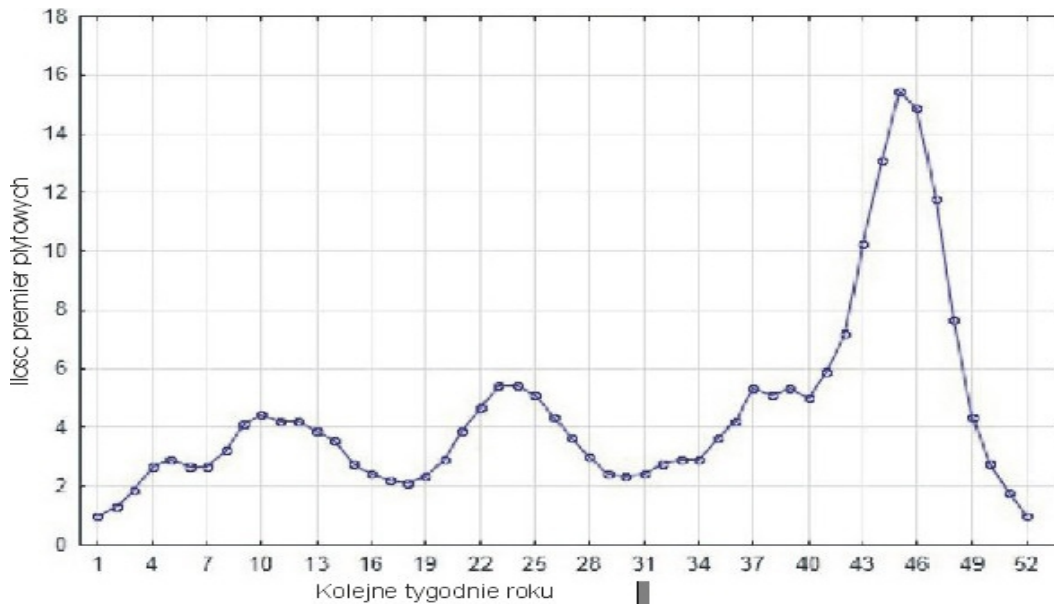
$$dX_t = (a(t) - bX_t)dt + sX_t dW_t + j(t)(dQ_t + dq_t) + i(t)dq_t \quad , \quad (5.6)$$

gdzie:  $i(t)=1$ , jeśli proces znajduje się w stanie (2) z prawdopodobieństwem  $p$ , a w przeciwnym razie  $i(t)=0$ .

### 5.2.2 Markowskie przelączenie stanów. Przypadek niestacjonarny.

Model stacjonarny, choć opisuje dobrze trajektorie sprzedaży, nie uwzględnia jeszcze kompletnych realiów ekonomicznych. Jako jego udoskonalenie proponuję zatem następujące dodatkowe założenia:

- 1) zależność od sezonowości, jeśli chodzi o częstotliwość wydawania albumów – tak aby uwzględnić, że szczyt natężenia premier płytowych przypada na jesień i na miesiąc przed świętami Bożego Narodzenia;
- 2) markowski rozmiar skoku sprzedaży w chwili premiery jest proporcjonalny do całkowitej sprzedaży poprzedniej płyty – tak, aby uwzględnić, że zakupu albumu w chwili premiery dokonują przede wszystkim dotychczasowi fani, posiadający poprzednie płyty artysty.



**Rys. 5.9** Wygładzona funkcja korygująca  $C_y(t)$ , odpowiadająca Rys. 5.5, określająca prawdopodobieństwo wydania albumu w określonej porze roku na podstawie danych empirycznych 30 najpopularniejszych artystów (2003-2013) [89].

Założenie zależności częstotliwości premier płytowych od pory roku wynika z Rys. 5.5, na którym interesuje nas dolna krzywa, uzyskana dla premier płytowych 30 największych gwiazd. Na potrzeby modelu użyjemy jej, ale w postaci wygładzonej celem uniknięcia wpływu lokalnych fluktuacji (Rys. 5.9). Przy użyciu tej funkcji  $C_y(t)$  prawdopodobieństwo wydania albumu, czyli element macierzy przejścia  $r_{12}$  zostaje przeskalowany o odpowiadającą mu wartość funkcji korygującej. Nowozdefiniowane prawdopodobieństwo przejścia zależy zatem od czasu:

$$r_{12}(t) = \frac{r_{12}C_y(t)}{E[C_y(t)]} \quad , \quad (5.7)$$

gdzie  $t$  jest określonym tygodniem danego roku, a funkcja korygująca  $C_y(t)$  jest znormalizowana zgodnie z jej wartością oczekiwaną  $E[C_y(t)]$  (Rys. 5.9).

Kolejnym udoskonaleniem jest wprowadzenie założenia, że liczba sprzedanych płyt w tygodniu premiery jest proporcjonalna do całkowitej sprzedaży poprzedniego albumu. Zamiast dotychczasowej zmiennej  $Q_t$  ze stacjonarnego modelu, wprowadzamy nową zmienną  $Q_t(i)$ :

$$Q_t(i) = \begin{cases} \text{dla } i > 1 & \max\left(\frac{s_t(i-1)}{s_c} + s_s N(0, 1), 0\right) \\ \text{dla } i = 1 & Q_t \end{cases} \quad (5.8)$$

gdzie  $i$  jest numerem kolejnego albumu artysty,  $s_t(i)$  jest całkowitą sprzedażą poprzedniego

albumu,  $s_c$  jest czynnikiem skalującym sprzedaż, a  $s_s$  jest poziomem szumu. Jednocześnie przyjmujemy, że sprzedaż pierwszego (debiutanckiego) albumu artysty jest zmienną losową, tak jak w modelu stacjonarnym. Założenie to ma uzasadnienie w małych budżetach, które koncerty fonograficzne przeznaczają na promocję pierwszych płyt. Debiutanci bowiem, zanim podpiszą kontrakt płytowy, najczęściej promują swoją twórczość na własną rękę i dopiero, jeśli spodobają się publiczności, trafiają pod skrzydła wielkich koncernów płytowych [59].

Z powodu dużej liczby parametrów przeprowadzamy tylko kilka numerycznych procedur estymacji, a model wzbogacamy o metody heurystyczne. Przyjmuję, że jednostką sprzedaży jest tysiąc egzemplarzy [ $1k=1000$ ]. Wartości początkowe macierzy  $\{r_{ij}\}$  zostały zadane z góry – średni czas trwania promocji zdefiniowany jako  $1/r_{22} = p * E(k) = 10 * 3,311 = 33$  tygodnie, gdzie  $k$  jest liczbą singli z jednego albumu,  $1/p=10$  jest średnią wartością odstępu między kolejnymi singlami (w tygodniach). Wartość oczekiwana  $E(k)$  jest oszacowana w oparciu o obserwacje empiryczne (Rys. 5.8), natomiast średni odstęp między kolejnymi albumami artysty w oparciu o Rys. 5.4 oszacowałem na  $d=131$  tygodni. Oznacza to, że czas artysty pozostającego w spoczynku wynosi średnio  $d-1/p=98$  tygodni.

W kolejnym kroku parametry MRqGRW zostały oszacowane metodą największej wiarygodności, stąd:

$$\begin{aligned} a &= 1,16 \\ b &= 267,51 \\ s_c &= 0,063. \end{aligned}$$

Niestety, z powodu uproszczeń modelu i zbyt krótkich serii czasowych, znaczenie takiej estymacji jest symboliczne, gdy artyści w stanie spoczynku mają cotygodniową sprzedaż płyt poniżej tysiąca egzemplarzy i nie są notowani na listach sprzedaży (wówczas przyjmuję, że sprzedaż wynosi zero). Szczytowe wartości określonego rozkładem lognormalnym  $\ln N$  parametru:

$$Q_t = 1000 * \ln N(\bar{Q}_t, \sigma_{Q_t}) \quad (5.9)$$

zostały oszacowane na podstawie empirycznych obserwacji i wynoszą:

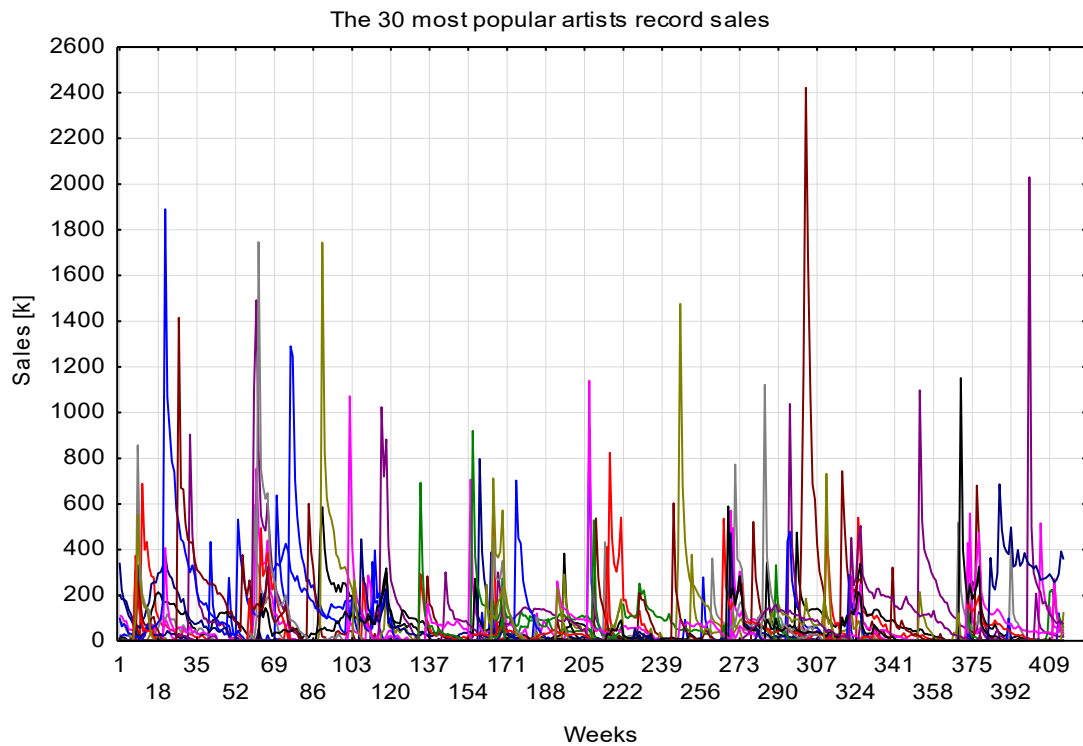
$$\bar{Q}_t = 1,2, \quad \sigma_{Q_t} = 1,7.$$

Wartość  $q_t$  została wybrana arbitralnie z przedziału  $(0,100)$ . Ponieważ w rzeczywistości artysta średnio co trzecim albumem zdobywa większą popularność niż krąg własnych fanów, otrzymujemy:  $1/3 = E(k)*p'$ , co prowadzi do wniosku, że  $p' = 0,1007$ .

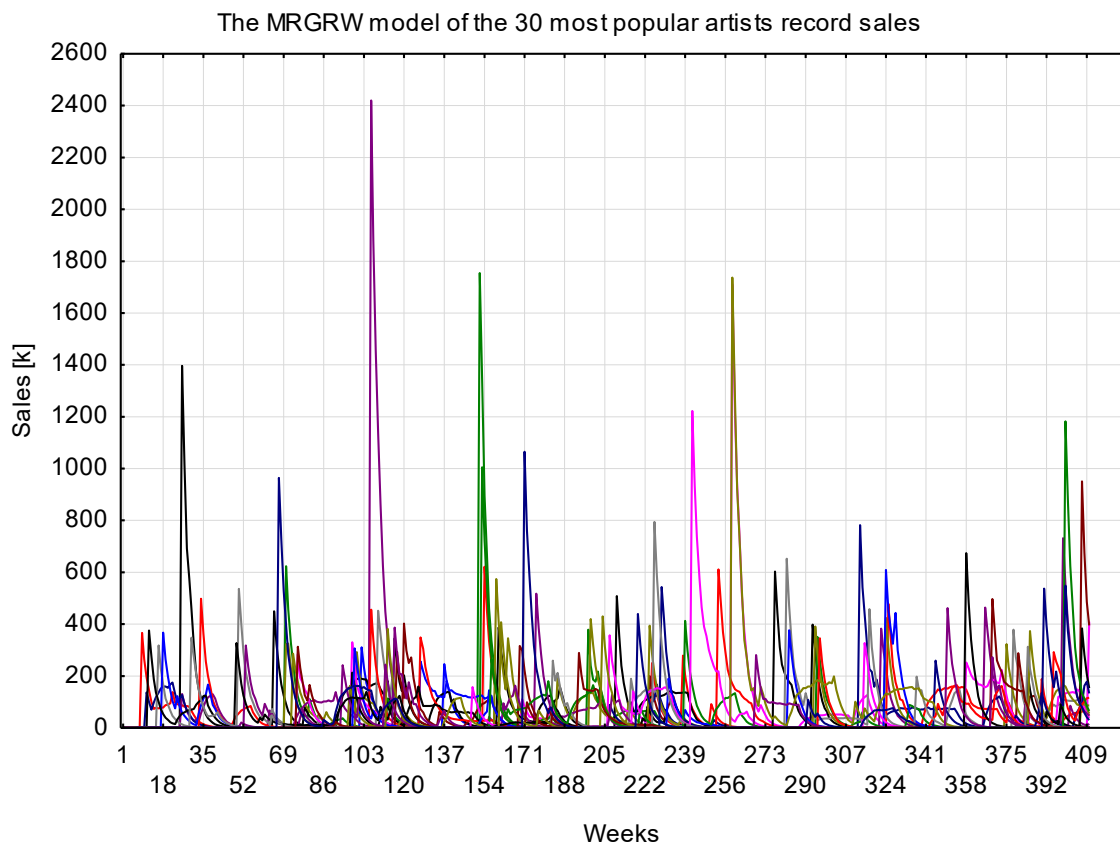
W modelu niestacjonarnym funkcja korygująca datę wydania albumu ze względu na porę roku (Rys. 5.9) powstała poprzez wygładzenie krzywej z Rys. 5.5 dla 30 najpopularniejszych artystów za pomocą średniej ruchomej w oknie czasowym 3 tygodni ze względu na fakt, iż czas sprzedaży wynosi tydzień, czas zliczania globalnej sprzedaży do momentu ogłoszenia wyników również trwa tydzień, a trzeci tydzień wynika z faktu, że rok może obejmować 52 lub 53 cotygodniowe notowania. Podstawą funkcji rozmiaru skoku sprzedaży w chwili premiery jest całkowita sprzedaż poprzedniego albumu wyrażona przez współczynnik skalujący  $s_c$  i poziom szumu  $s_s$ , które zostały dopasowane metodą najmniejszych kwadratów do empirycznego rozkładu następujących po sobie skoków sprzedaży.

### **5.3 Konfrontacja wyników modelowania z rzeczywistością.**

Złożenie 30 wykresów cyklu życia produktu (Rys. 5.1) na podstawie rzeczywistej sprzedaży albumów najpopularniejszych artystów w latach 2003-2011 prowadzi do wykresu przedstawionego na Rys. 5.10. Stacjonarny model sprzedaży na rynku fonograficznym również prowadzi do analogicznego wykresu (Rys. 5.11), podobnie jak model niestacjonarny. Jednak, mimo że wykresy wydają się niemal identyczne, konfrontacja wyników modelowania z rzeczywistością pokazuje ciekawe różnice.

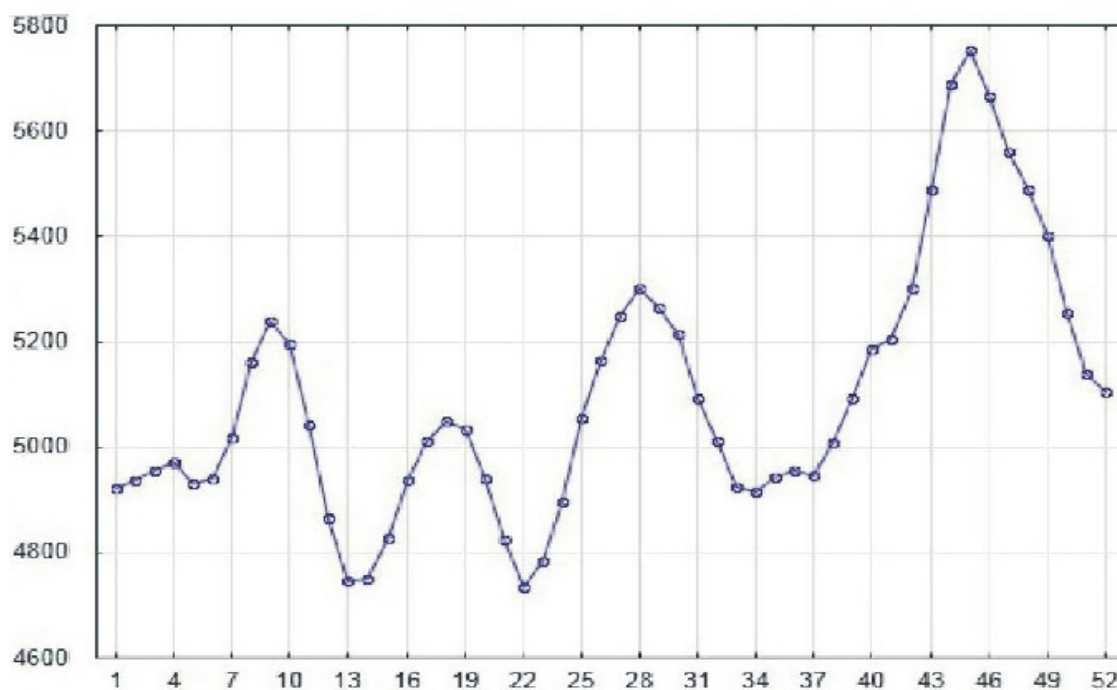


**Rys. 5.10** Złożenie 30 cykli życia produktu na podstawie rzeczywistej sprzedaży albumów najpopularniejszych artystów w latach 2003-2011 [89]. Oś pozioma oznacza kolejne tygodnie (weeks), a pionowa sprzedaż (sales) w tys.egz



**Rys. 5.11** Złożenie 30 cykli życia produktu na podstawie stacjonarnego modelu sprzedaży albumów najpopularniejszych artystów w latach 2003-2011 [89]. Oś pozioma oznacza kolejne tygodnie (weeks), a pionowa sprzedaż (sales) w tys.egz [K].

Pierwsze interesujące porównanie przynosi nam wykres sprzedaży w całym sezonie, uzyskany na podstawie danych modelowanych przez kolejne lata (Rys. 5.12). W wynikach uzyskanych z modelu niestacjonarnego obserwujemy ciągłość krzywej sprzedaży przy przejściu do następnego roku (tygodnie: 52. i 1.), czego nie ma na rynku rzeczywistym, gdzie przełom roku jest związany ze skokowym (praktycznie niemal nieciągłym) spadkiem sprzedaży (Rys. 4.5). Gwałtowność zmiany wartości sprzedaży w tych dwóch tygodniach pozwala na stwierdzenie, że mamy tu do czynienia ze swoistym przejściem fazowym. Ponadto następujące wówczas maksimum sprzedaży w modelu niestacjonarnym jest ledwie o 20% wyższe niż średnia sprzedaż w pozostałej części roku (Rys. 5.12), podczas gdy w rzeczywistości sprzedaż przedświąteczna jest aż 2-krotnie wyższa niż w pozostałych okresach (Rys. 4.5). Pojawia się w związku z tym pytanie o przyczynę tych różnic.



**Rys. 5.12** Średnia sprzedaż wszystkich albumów 30 najpopularniejszych artystów (podawana w tysiącach egzemplarzy) w modelu niestacjonarnym w zależności tygodnia roku [89].

Bazując na wynikach przedstawionych w poprzednim rozdziale, kiedy w hierarchicznej strukturze drzewa minimalnej rozpiętości dla artystów (Rys. 4.14) zamiast sektora wykonawców muzyki pop otrzymaliśmy sektor największych gwiazd różnych gatunków, wskazując, że przyczyną różnic pomiędzy modelem a rzeczywistością jest fakt, że w rzeczywistości przed świętami Bożego Narodzenia do rynku włącza się nowa, zewnętrzna grupa konsumentów, która kupuje muzykę „od święta”. Tymczasem nasz model zakłada,

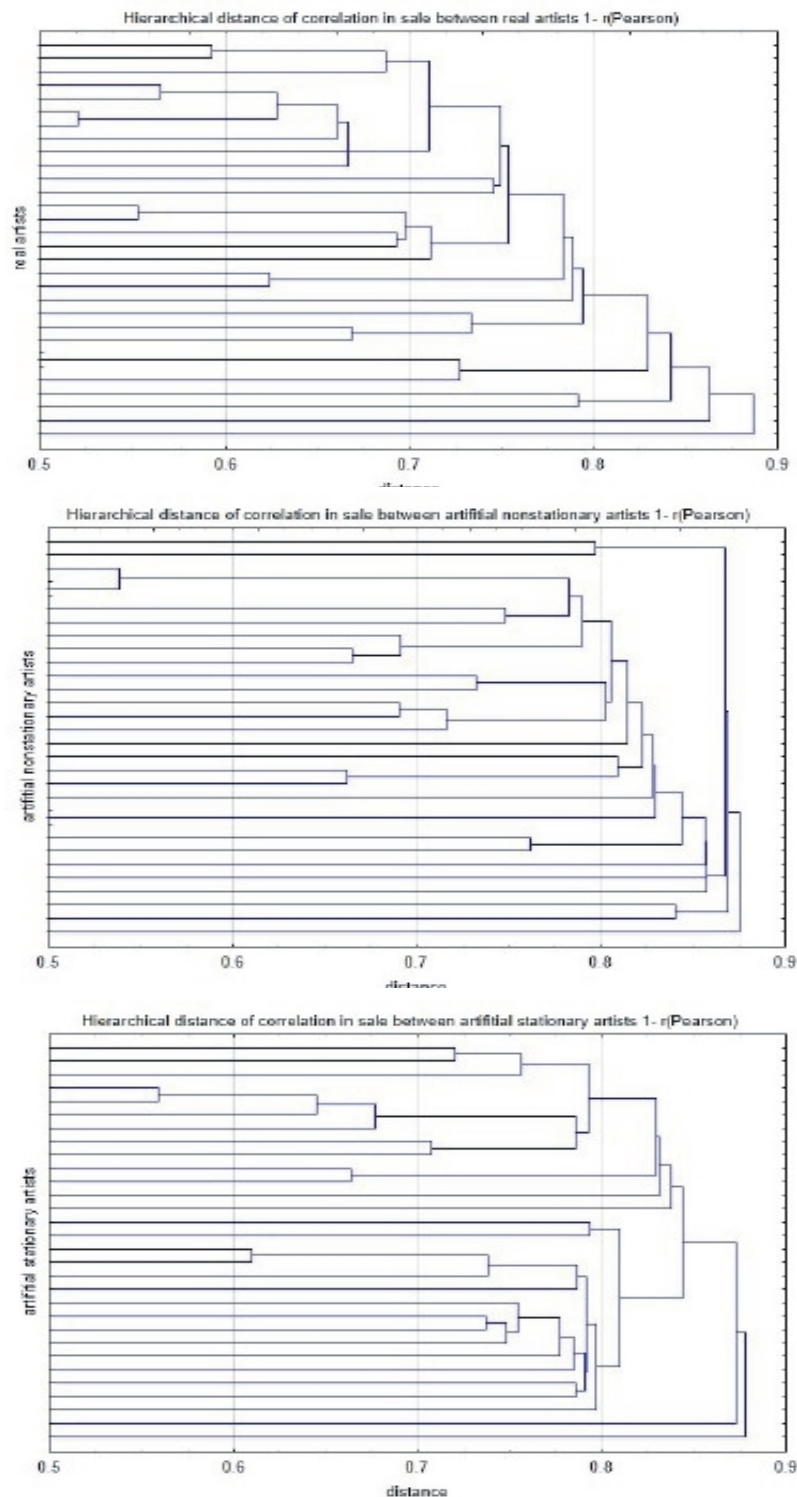


że rynek fonograficzny opiera się na grupie konsumentów kupującej płyty na co dzień. W efekcie przyczyną maksimum przedświątecznej sprzedaży dla modelu niestacjonarnego (Rys. 5.12) jest wysokie natężenie premier płytowych pod koniec roku, a nie sam okres przedświąteczny i związane z nim wzmożone zakupy. Stąd modelowy wzrost sprzedaży jedynie o 20%.

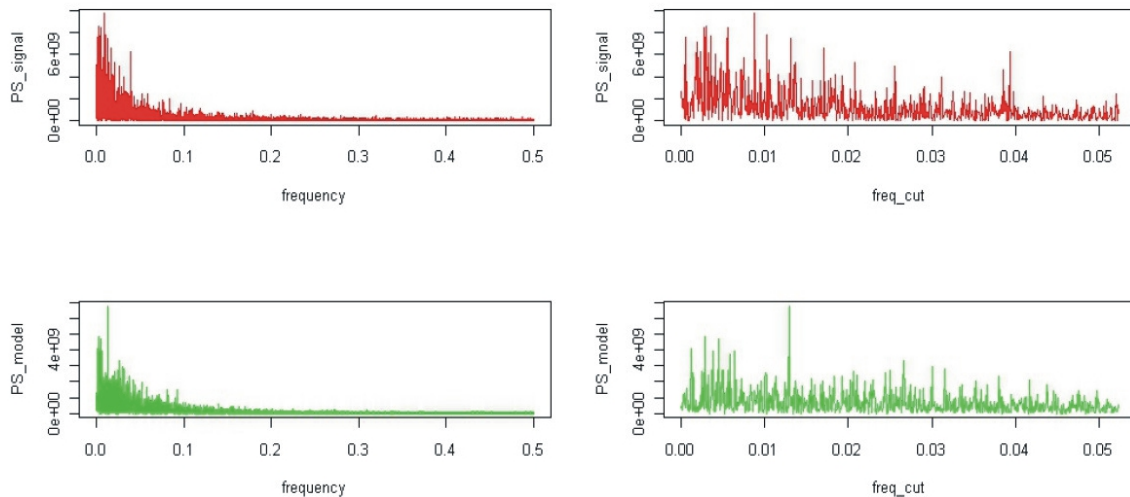
Za pozostałą część wzrostu, który następuje dla rzeczywistych danych, odpowiadają nie premiery płytowe, ale nowi konsumenci, dla których impulsem do zakupów jest właśnie okres przedświąteczny i zakup prezentów, a nie natężenie premier. Wynika stąd, że ta nowa, zewnętrzna grupa konsumentów kupowała płyty przed Świątami nawet, gdyby nie było premier płytowych. Zatem wnioskuję, że jest ona złożona raczej z laików, którzy nie słuchają muzyki na co dzień. Zapewne z tego samego powodu przedstawiony na Rys. 5.5 wykres pokazuje, że wytwórci z premedytacją ustawiają premiery płyt najbardziej znanych gwiazd właśnie przed końcem roku – aby walczyć o tę dodatkową grupę konsumentów, która będzie kupować płyty tylko największych gwiazd. Dlatego w drzewie minimalnej rozpiętości (Rys. 4.14) zamiast sektora muzyki pop mamy sektor gwiazdorski artystów różnych gatunków. Można na podstawie tych rozważań dojść do wniosku, że obserwowany efekt kolektywnych zakupów ma charakter emergentny i jest przesłanką, że rynek płytowy ma pewne cechy układu złożonego.

Kolejną wartą uwagi cechą rynku jest omówiona w poprzednim rozdziale hierarchiczność. O ile występuje ona w rzeczywistych danych (Rys. 4.14, Rys. 4.15) to w modelu wygląda ona inaczej, a lepiej odzwierciedla ją model niestacjonarny (Rys. 5.13). Przyczyną jest fakt, że w rzeczywistości artyści należą do określonych gatunków muzyki, a podobieństwo w trajektorii sprzedaży wynika z podobieństw stylistycznych, które inspirują konsumentów do zakupu, czego w modelu nie zakładamy.

Choć model okazał się tylko częściowo zgodny z rzeczywistością, to jednak dzięki konfrontacji z nią pomógł wyjaśnić, skąd wziął się ten gwiazdorski sektor. Dodam, że model niestacjonarny sprawdza się za to doskonale na tych rynkach, w których nie mamy kulturowego efektu Bożego Narodzenia, jak choćby na trzecim co do wielkości wiodącym rynku – w Japonii. Rozkład światowych premier płytowych jest tam prawie identyczny jak w USA i na całym świecie, ale na przełomie roku sprzedaż płyt maleje łagodnie w sposób ciągły, czyli praktycznie tak, jak w omówionym modelu niestacjonarnym.



**Rys. 5.13** Hierarchiczna struktura artystów na rynku fonograficznego w rzeczywistości (górný diagram) oraz w modelu stacjonarnym (środkowy) i niestacjonarnym (dolny) [89].



**Rys. 5.14** Widmo mocy dla rzeczywistych krzywych sprzedaży (u góry) i ich odpowiedniki dla modelu niestacjonarnego (na dole) [89]

Kolejną różnicę między modelem, a rzeczywistością daje widmo mocy, które dla danych rzeczywistych odzwierciedla sezonowość procesu sprzedaży, nie widziana w przypadku modelu. Oznacza to, że tak jak w przypadku nieco odmiennych wykresów sprzedaży (Rys. 4.5, Rys. 5.12), przyczyną efektu sezonowości jest nie częstotliwość wydawania nowych premier na rynku, ale napływ nowej, zewnętrznej grupy konsumentów, która pojawia się co roku jesienią. Prawdopodobnie także z tego powodu wykładnik Hursta, który obliczyłem dla wygenerowanej przez model niestacjonarny serii danych sprzedaży 30 artystów o długości 400 lat okazał się nieco niższy ( $H=0,31$ ) od wyników rzeczywistych ( $H=0,39$ ) i pokazał większą niż w rzeczywistości antypersystencję.

Zaproponowany model sprzedaży płyt jest jak dotąd jedynym stochastycznym modelem tego typu sprzedaży konsumenckiej na świecie. I choć model z powodu swoich założeń sprawdza się w pełni na razie jedynie dla rynku japońskiego, to dzięki jego zastosowaniu odkryłem, że w pozostałych krajach świata za zjawisko sezonowości na rynku fonograficznym odpowiada nie tyle natężenie premier płytowych jesienią, ale napływ nowej, zewnętrznej grupy konsumentów, która zna się na muzyce pobieżnie i kupuje płyty przed świętami Bożego Narodzenia. To właśnie ta zewnętrzna grupa konsumentów jest przyczyną, że w drzewie minimalnej rozpiętości dla artystów - które pokazałem w poprzednim rozdziale (Rys. 4.14) – zamiast sektora odpowiadającego estetycznie muzyce pop istnieje sektor gwiazdorski, łączący największe gwiazdy odmiennych od siebie gatunków.

## Rozdział 6. Oddziaływanie wielu rynków - rynek singli.

### 1.1 Ścieżki dyfuzji czyli cykle życia produktu między rynkami w Europie

Poprzednie 2 rozdziały zdominowało ujęcie rynku fonograficznego od strony czysto ekonomicznej i ilościowej – trajektorii sprzedaży albumów najpopularniejszych artystów. Rynek płytowy to jednak także single, czyli pojedyncze piosenki, które dla wytwórni i konsumentów mają czasem większe znaczenie, niż wykonujący je artysta. Pod względem ilościowym miara popularności przebojów zawsze była niejednoznaczna. Bowiem o ile brytyjskie, irlandzkie czy kanadyjskie listy powstawały w oparciu o czystą sprzedaż singli, to w USA gorąca setka *Hot 100* magazynu Billboard oprócz sprzedaży uwzględniała także obecność piosenek w stacjach radiowych i telewizyjnych, podobnie jak część europejskich list (np. Hiszpania). Wynika to z faktu, że do 1970 roku wiodące brytyjskie media państwowe jak BBC ze względów obyczajowych promowały prawie wyłącznie muzykę klasyczną i pop, unikając dominującego w stacjach prywatnych i pirackich rock and rolla, w przeciwieństwie do Stanów Zjednoczonych, gdzie panowała większa różnorodność. Ze względu na spadek sprzedaży fizycznych nośników, do sprzedaży singli w XXI wieku dolicza się też liczbę legalnych płatnych ściągnięć danego utworu w postaci mp3, a od 2005 roku także dzwonek na telefony komórkowe, przy czym ilość ta jest unormowana adekwatnie do ceny jaką za ściągnięcie utworu zapłacił klient.

Brak jednoznacznej ilościowej miary popularności utworu w różnych krajach sprawia, że nawet dziś najczęściej stosowanym odzwierciedleniem sukcesu jest pozycja piosenki na liście przebojów, odzwierciedlająca jej cykl życia jako produktu. Od końca lat 80-tych XX-wieku popularność piosenek praktycznie we wszystkich krajach mierzona jest w odstępach cotygodniowych. Dlatego w mojej pracy, mierząc cykle życia piosenek jako produktów, zdecydowałem się także użyć pozycji na liście przebojów jako miary ilościowej. Co ciekawe, w takim ujęciu ilościowym kształt trajektorii popularności singli różni się od omówionej poprzednio trajektorii dla albumów i bardziej przypomina klasyczny cykl życia produktu [90], na który składają się innowatorzy (*innovators*), wprowadzający (*early adopters*), wczesna większość (*early majority*), późna większość (*late majority*) i maruderzy (*laggards*).

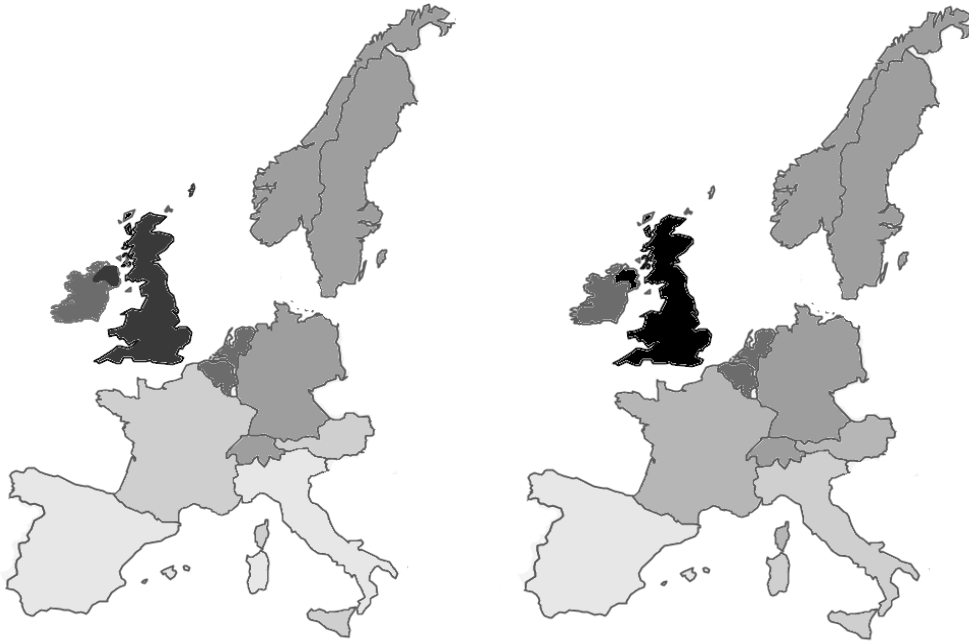
Ciekawy jest również sposób przekazywania popularności singla do innych krajów. Ze względu na dostępność danych i historyczne istnienie rynków sprzedaży singli, w moich badaniach analizowałem 1786 najpopularniejszych singli wydanych w latach 1966-2016 na 12 największych rynkach fonograficznych Europy, takich jak: Austria (A), Belgia (B), Francja (F), Niemcy (D), Irlandia (IRL), Włochy (I), Holandia (NL), Norwegia (N), Hiszpania (E), Szwecja (S), Szwajcaria (CH), Wielka Brytania (GB)

Celem było ustalenie, w jaki sposób popularność singla na jednym rynku oddziałuje na inne rynki, a w szczególności ustalenie najbardziej prawdopodobnych ścieżek, według których cykl życia produktu przebiega. Oczywiście na przestrzeni 50 lat mało singli osiągnęło popularność we wszystkich 12 krajach, jednak udało mi się zebrać 100 największych w skali europejskiej przebojów i przeanalizować ich cykle życia w Europie. Naturalnym kandydatem do opisu rozprzestrzeniania się ich popularności było zastosowanie podejścia epidemiologicznego. W tym celu wygodnie jest założyć, iż singiel zdobył (zaraził popularnością) dany rynek, jeśli osiągnął na nim najwyższe miejsce listy przebojów i już nie awansuje wyżej. A następnie obserwować w jakim odstępie czasu osiąga najwyższe pozycje także na innych rynkach. Przykładowo: singiel 'We Found Love' pochodzącej z Barbadosu Rihanny, wszedł na europejskie listy 8 października 2011, a jego popularność w ciągu pierwszych 8 tygodni przedstawia tabela 6.1.

**Tab. 6.1** Miejsca singla Rihanny 'We Found Love' na lokalnych listach w ciągu pierwszych 8 tygodni od premiery. Odkąd singiel zdobył na lokalnym rynku najwyższe miejsce, jego kolejne tygodnie zaznaczone są jasnym kolorem

Kraj	1 tydzień	2 tydzień	3 tydzień	4 tydzień	5 tydzień	6 tydzień	7 tydzień	8 tydzień
A	-	-	-	12	12	4		
B	3							
CH	3	3	2	2	2	1		
D	-	-	-	1				
E	15	14	14	9	9	9	5	3
F	1							
GB	-	1						
I	-	-	-	7	4	4		
IRL	-	3	1					
N	1							
NL	14	3						
S	2	2	2	2	1			

W tym przypadku źródłem infekcji okazała się Belgia, choć singiel w pierwszym tygodniu zaraził także Francję i Norwegię. Ostatnimi krajami, które przejęły jego popularność, okazały się Włochy i Hiszpania. W podobny sposób, analizując cykle życia 100 badanych piosenek, można zbudować mapę Europy, pokazującą zdobywanie popularności [Rys. 6.1]



**Rys. 6.1.** Średnia (po lewej) i mediana (po prawej) czasu osiągnięcia szczytu popularności na lokalnych listach singli

Co więcej, dla każdej piosenki możliwe jest określenie macierzy opóźnień (Tab. 6.2), o elementach określonych przez odległości:

$$d_{xy} = |t_x - t_y| \quad (6.1)$$

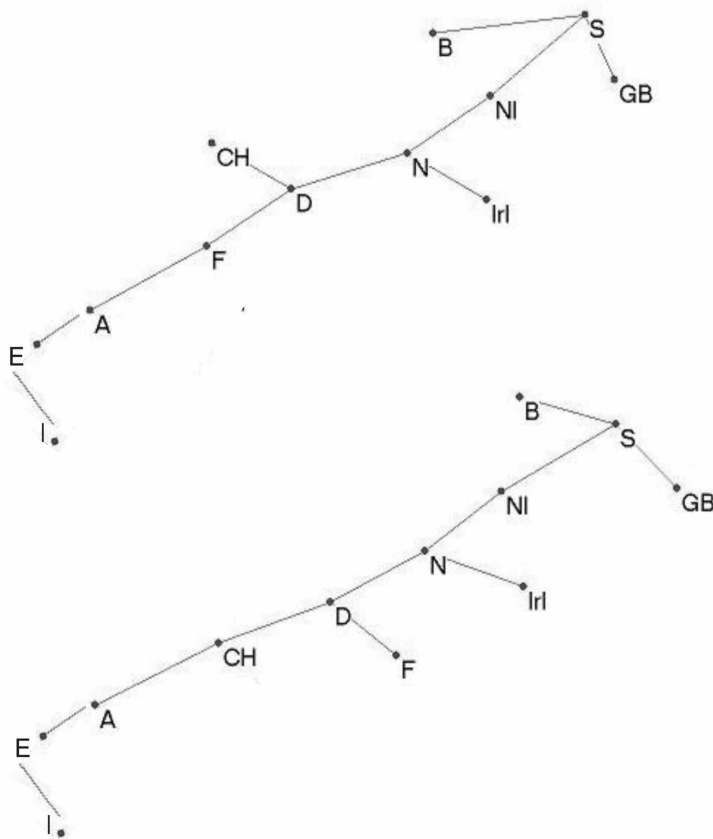
gdzie  $t_x$  oraz  $t_y$  reprezentują tygodnie, w których piosenka osiągnęła szczyt list przebojów w krajach  $X$  oraz  $Y$ . Tak zdefiniowana macierz jest symetryczna.

**Tab. 6. 2** Macierz opóźnień  $d_{xy}$  między krajami dla singla Rihanny 'We Found Love' liczonych w tygodniach

	A	B	CH	D	E	F	GB	I	IRL	N	NL	S
A	■	5	0	2	2	5	4	0	3	5	4	1
B	5	■	5	3	7	0	1	5	2	0	1	4
CH	0	5	■	2	2	5	4	0	3	5	4	1
D	2	3	2	■	4	3	2	2	1	3	2	1
E	2	7	2	4	■	7	6	2	5	7	6	3
F	5	0	5	3	7	■	1	5	2	0	1	4
GB	4	1	4	2	6	1	■	4	1	1	0	3
I	0	5	0	2	2	5	4	■	3	5	4	1
IRL	3	2	3	1	5	2	1	3	■	2	1	2
N	5	0	5	3	7	0	1	5	2	■	1	4
NL	4	1	4	2	6	1	0	4	1	1	■	3
S	1	4	1	1	3	4	3	1	2	4	3	■

Na podstawie macierzy opóźnień możemy skonstruować drzewo minimalnej rozpiętości (MST) i zastosować algorytm przyłączania według tabeli 6.1. Okazuje się jednak, że deterministyczna konstrukcja drzewa ma wiele wariantów, bowiem nie sposób ustalić kolejności czy np. w drugim tygodniu popularność singla Rihanny przejęła najpierw Wielka Brytania, czy też Holandia (istnieją co prawda znane wytwórniom płytowym dane częściowe, tzw. *midweek charts*, ale w ograniczonym zakresie). Nawet 25 lat wcześniej, gdy w erze przed Internetem cykl popularności piosenki przebiegał dłużej (nawet do 20 tygodni), ustalone drzewa minimalnej rozpiętości miały wiele wariantów i były zdegenerowane jak w przypadku przeboju 'Dancing Queen' szwedzkiej grupy ABBA [Rys. 6.2]. W tym przypadku źródłem infekcji jest kraj pochodzenia artysty – Szwecja.

Większość badanych artystów pochodzi jednak z Wielkiej Brytanii, stąd na mapie Europy [Rys. 6.1] Wielka Brytania jako częste źródło infekcji zaznaczona jest najciemniej, a Hiszpania i Włochy, jako podlegające opóźnieniu – najjaśniej. Dla artystów spoza Europy jak pochodzący z USA Michael Jackson jako źródło infekcji przyjmowałem pierwsze europejskie kraje, w których jego utwór wszedł na listę (np. w przypadku jego przeboju 'Black Or White' były to niezależnie od siebie Norwegia i Włochy).



**Rys. 6.2** Drzewo minimalnej rozpiętości uzyskane w dwóch wariantach na podstawie macierzy opóźnień między krajami dla singla ABBA 'Dancing Queen' (1976).

## 6.2 Algorytm przyłączania

Aby wykryć najbardziej prawdopodobne ścieżki zarażenia w oparciu o średni czas przyłączania, wykonano symulacje dla 50 najpopularniejszych singli z lat 1966-2016, które opanowały wszystkie 12 krajów Europy Zachodniej [92]. Najpierw zostało ustalone źródło infekcji (kraj pochodzenia, lub – w przypadku pozauropejskich artystów – kraj pierwszego wejścia na listę). Algorytm tworzy drzewa minimalnej rozpiętości dla każdego singla (1000 razy) poprzez przyłączanie kolejnych krajów (węzłów) w sposób następujący:

- (1) ustalamy źródło infekcji (np. ABBA 'Dancing Queen' pochodzi ze Szwecji)
- (2) przyłączamy kolejne węzły (np. do Szwecji dołączają w drugim kroku jednocześnie Belgia i Wielka Brytania)
- (3) węzły dołączane są w sposób losowy, ale z uwzględnieniem wielkości rynków i opóźnień (6.1). Współczynnik zdarzeń sprzyjających przyłączeniu kraju  $X$  do kraju



$Y$  (gdzie  $t_x \neq t_y$ , bo kraje  $X$  i  $Y$  z założenia są różne) zachodzi według wzoru:

$$P_{x-y} \sim \frac{1}{t_x - t_y} * \frac{N_Y}{N_{AVG}} \quad (6.2)$$

gdzie:

$P_{x-y}$  – współczynnik zdarzeń sprzyjających dołączeniu węzła  $X$  do węzła  $Y$

$N_Y$  – wielkość rynku  $Y$

$N_{AVG}$  – średnia wielkość rynku w Europie

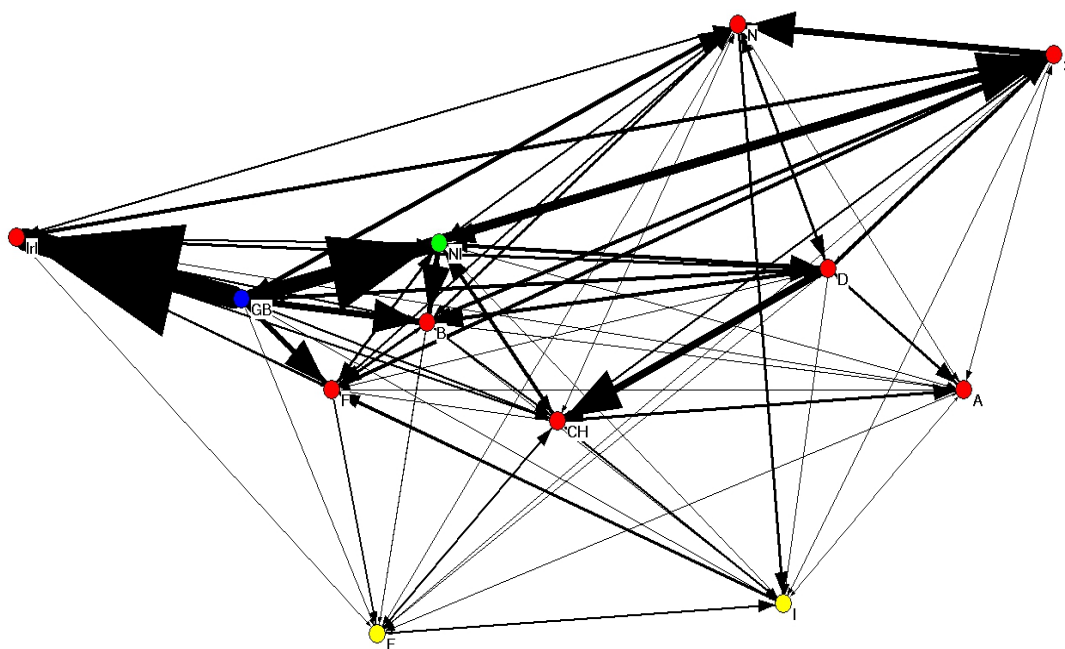
(4) zakładamy, że dla Europy Zachodniej wielkość rynku proporcjonalna

jest do liczby ludności krajów

Założenie (4) dla badanych 12 krajów Europy Zachodniej zgodne jest z raportami IFPI, które omawiałem w rozdziale czwartym. W efekcie otrzymujemy skierowaną sieć rozprzestrzeniania się popularności singli między krajami (Rys. 6.2), gdzie Wielka Brytania pełni rolę huba, Holandia – mostu szybko przenoszącego popularność, a Włochy i Hiszpania – krajów peryferyjnych Europy Zachodniej. Tabela 6.3, na podstawie wyników otrzymanych przez stosowanie algorytmu przyłączania, przedstawia ilościowo kraje pod względem importu (*indegrees*) lub eksportu (*outdegrees*) singli na zewnątrz.

**Tab. 6.3** Wielkość eksportu/importu na podstawie symulacji rozprzestrzeniania się popularności singli w Europie

kraj	Eksportują	Importują
<b>A</b>	<b>23</b>	<b>36</b>
<b>B</b>	<b>60</b>	<b>65</b>
<b>CH</b>	<b>45</b>	<b>63</b>
<b>D</b>	<b>75</b>	<b>53</b>
<b>E</b>	<b>16</b>	<b>27</b>
<b>F</b>	<b>43</b>	<b>50</b>
<b>GB</b>	<b>192</b>	<b>36</b>
<b>I</b>	<b>17</b>	<b>36</b>
<b>Irl</b>	<b>43</b>	<b>89</b>
<b>N</b>	<b>49</b>	<b>65</b>
<b>NI</b>	<b>74</b>	<b>94</b>
<b>S</b>	<b>55</b>	<b>78</b>



**Rys. 6.3** Skierowana sieć rozprzestrzeniania się popularności singli w Europie Zachodniej

Otrzymany wynik, że Wielka Brytania ma ogromną zdolność kreowania przebojów w Europie (przy jednocześnie dość przeciętnej zdolności importowania przebojów z zewnątrz), jest zgodny z danymi empirycznymi. Jednak dopiero dzięki przedstawieniu mapy Europy jako skierowanej sieci połączeń między krajami, można ją interpretować i wydzielić na niej 2 klastry:

- Skandynawia, Benelux, Wielka Brytania, Irlandia i Niemcy
- Hiszpania, Włochy, Szwajcaria i Austria

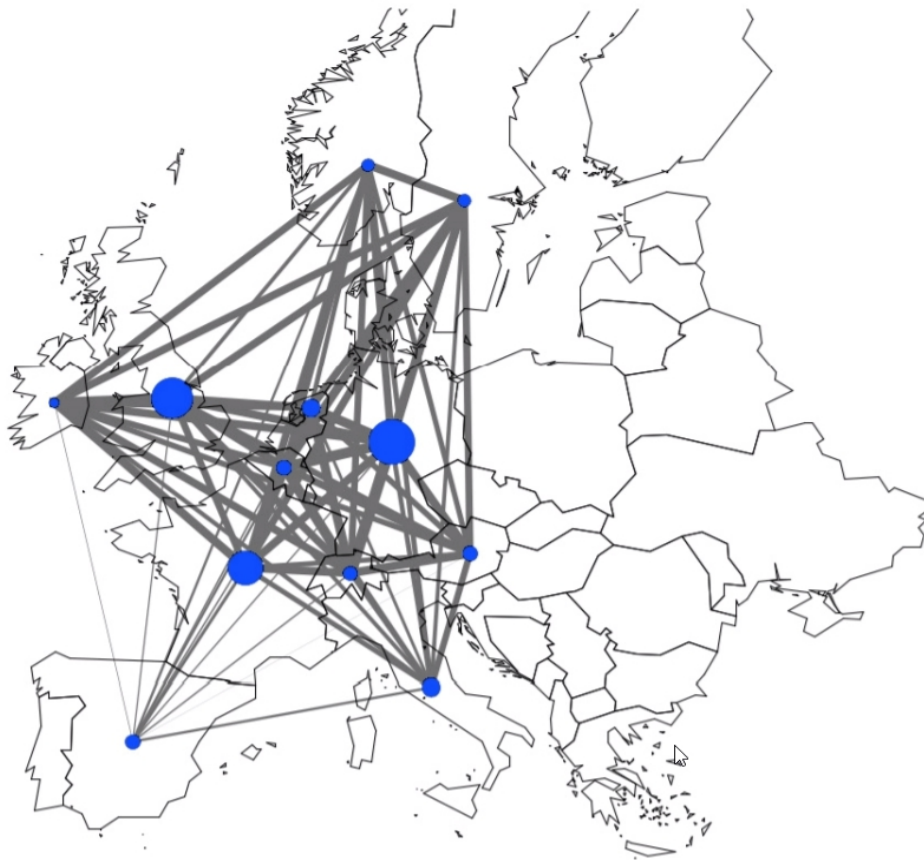
Pierwszy z nich reprezentuje kraje szybko przyswajające nowy przebój, a drugi kraje konserwatywne, absorbujące popularność singla jako ostatnie. Więcej wyników na ten temat przedstawia praca [92], w której ponadto jest omówiona szczególna rola Francji, która raz należy do klastra krajów innowacyjnych, a raz do krajów konserwatywnych (w zależności od stosowanych metod przyłączania węzłów). Na taką rolę Francji może mieć wpływ francuskie prawo autorskie, nakazujące stacjom radiowym grać w przynajmniej 50% francuskojęzyczny repertuar. W efekcie (jak wskazuje doświadczenie, a także

przeprowadzone symulacje) Francja ma ograniczoną przez prawo zdolność przyjmowania nowych przebojów z zagranicy, a więc jeśli nie przyjmie ich od razu, to w następnych krokach nie przyjmie ich w ogóle.

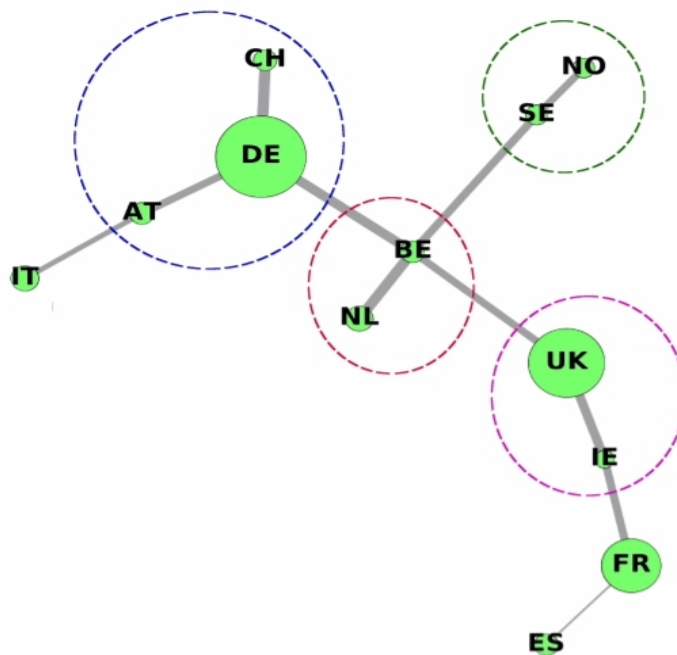
**Tab. 6.4** Macierz korelacji między seriami czasowymi reprezentującymi opóźnienia między najwyższymi miejscami na listach przebojów w 12 krajach Europy dla 72 najpopularniejszych singli z lat 1966-2015

	A	B	CH	D	E	F	IRL	I	NL	N	S	GB
A	1	0,66	0,73	0,73	0,18	0,49	0,55	0,59	0,55	0,51	0,63	0,58
B	0,66	1	0,83	0,86	0,32	0,65	0,59	0,41	0,9	0,55	0,75	0,72
CH	0,73	0,83	1	0,92	0,27	0,71	0,68	0,49	0,76	0,61	0,73	0,68
D	0,73	0,86	0,92	1	0,26	0,7	0,7	0,48	0,79	0,66	0,72	0,7
E	0,18	0,32	0,27	0,26	1	0,37	0,21	0,34	0,28	0,34	0,27	0,25
F	0,49	0,65	0,71	0,7	0,37	1	0,73	0,51	0,59	0,66	0,64	0,62
IRL	0,55	0,59	0,68	0,7	0,21	0,73	1	0,43	0,71	0,43	0,69	0,81
I	0,59	0,41	0,49	0,48	0,34	0,51	0,43	1	0,37	0,5	0,43	0,43
NL	0,55	0,9	0,76	0,79	0,28	0,59	0,71	0,37	1	0,48	0,67	0,71
N	0,51	0,55	0,61	0,66	0,34	0,66	0,43	0,5	0,48	1	0,71	0,43
S	0,63	0,75	0,73	0,72	0,27	0,64	0,69	0,43	0,67	0,71	1	0,69
GB	0,58	0,72	0,68	0,7	0,25	0,62	0,81	0,43	0,71	0,43	0,69	1

Omówione wyżej zjawiska rzutują na macierz korelacji Pearsona, której elementy dane są wzorem (1.3) dla opóźnień między trajektoriami popularności piosenek w badanych 12 krajach Europy (Tab. 6.4), gdzie podobieństwa geograficzne czy kulturowe również zostały odzwierciedlone w postaci klastrów (Rys. 6.4 – 6.5). Powyższe dane dotyczą 72 najpopularniejszych singli, które opanowały każdy z 12 lokalnych rynków i ich wybór wynika nie tylko z dostępności danych (niższe pozycje na listach krajowych zwykle nie są publikowane), ale przede wszystkim z ekonomicznej dominacji jeśli chodzi o popularność, przewyższającej pozostałe z 17 000 badanych singli.



Rys. 6.4 Sieć korelacji dla opóźnień rozprzestrzeniania się popularności singli w Europie



Rys. 6.5 Drzewo minimalnej rozpiętości odpowiadające korelacjom między opóźnieniami w zdobywaniu popularności z uwzględnieniem klastrów

Warto zaznaczyć, iż na rysunkach 6.4-6.5 grubość linii łączącej kraje jest proporcjonalna do siły korelacji między nimi, dlatego możliwe jest wykrycie klastrów wspólnotowych, zgodnie ze wzorem:

$$Q = \frac{1}{2m} \sum_{ij} \left[ w_{ij} - \frac{k_i k_j}{2m} \right] \delta_{c_i, c_j} \quad (6.3)$$

gdzie  $m$  oznacza ilość krawędzi sieci, a  $w_{ij}$  reprezentuje wagę połączenia między węzłem  $i$  o krotności  $k_i$ , a węzłem  $j$  o krotności  $k_j$ . Natomiast oznaczone w delcie Kroneckera  $c_i$  oznacza wspólnotę, do której należy węzeł  $i$ . Wówczas możliwe jest wykrycie istnienia wspólnot [91] w badanej sieci, jeśli  $Q > 0$ . Metoda wykrywania wspólnot wykorzystuje pojęcie międzywęzłowości krawędzi (ang. *edge betweenness*), które określa ile geodezyjnych połączeń między parami węzłów zawiera dana krawędź. W każdym kolejnym kroku obliczana jest międzywęzłowość dla każdej krawędzi w sieci wraz z jej całkowitą modularnością, a następnie usuwana jest krawędź o największej wartości współczynnika międzywęzłowości. Algorytm jest zatrzymywany po rozdzieleniu sieci na poszczególne węzły. Struktura wspólnot jest wyodrębniana optymalnie w kroku o największym  $Q$ . Dla sieci ważonej, aby poprawnie dokonać jej dekompozycji, należy przekształcić ją do postaci multigrafu (tzn, wagi zastępuje się proporcjonalnie wieloma krawędziami) i podzielić współczynnik międzywęzłowości przez krotność krawędzi (przed wybraniem krawędzi do usunięcia). W ten sposób na rysunku 6.5 można wyodrębnić klastry zawierające: 1) Niemcy, Szwajcarię i Austrię, 2) Norwegię i Szwecję, 3) Belgię i Holandię, 4) Wielką Brytanię i Irlandię. Pozostałe kraje jak Francja, Włochy i Hiszpania nie należą tu do żadnego klastra. Co ciekawe, wynik ten – uzyskany wyłącznie wskutek badania korelacji między sprzedażą singli – pokrywa się z omawianym w dalszej części rozdziału podobieństwem kulturowym czy językowym między krajami w Europie.

### 6.3 Wielka Brytania jako inicjator – przepływ popularności w ramach jednego rynku

Interesująca byłaby także analiza sieciowa, jak rozprzestrzenia się popularność w ramach jednego rynku. Niestety oficjalne listy przebojów nie odzwierciedlają tego zagadnienia, bowiem są publikowane co najwyżej dla regionów posiadających autonomię (jak Walonia należąca do królestwa Belgii, czy Szkocja, należąca do Wielkiej Brytanii), co przy podziale kraju na 2 regiony dawałoby mało efektywne statystyki. Jedyne znane mi systematyczne cotygodniowe pomiary popularności singli pochodzą z sondażu Instytutu Gallupa z 1978 roku, który podzielił obszar Wielkiej Brytanii na 7 regionów geograficznych. Przedstawia je mapa:



**Rys. 6.6** Podział Wielkiej Brytanii na historyczne regiony w pomiarach cotygodniowej lokalnej sprzedaży singli według Gallupa (1978)

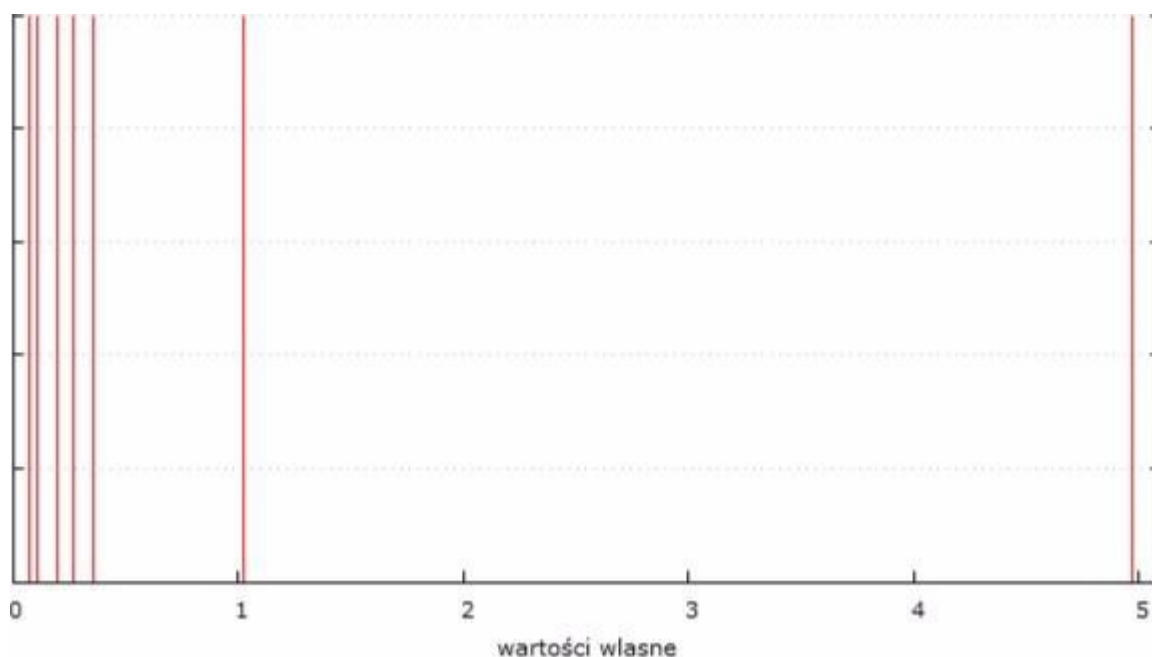
Instytut prowadził cotygodniowe pomiary sprzedaży singli w każdym z 7 regionów, od połowy marca 1978 do października 1978 (tworząc cotygodniowe listy 50 najpopularniejszych singli w każdym regionie z osobna), po czym uznał je za zbyt kosztowne i mało opłacalne. Jednak mimo krótkiego, półrocznego okresu udało się wyodrębnić 50 piosenek, które były notowane na listach we wszystkich 7 regionach, ponieważ rynek brytyjski jako największy w Europie i trzeci na świecie opiera się głównie na detalicznej sprzedaży sieciowej, a więc każdy wydany singiel staje się automatycznie dostępny we wszystkich regionach kraju, co pozwala mi badać przepływy popularności.

**Tab. 6.5** Macierz korelacji sprzedaży singli między różnymi regionami Wielkiej Brytanii (w dwóch ekstremalnych przypadkach podana jest także niedokładność oszacowania współczynnika korelacji)

	L	W&W	NE	YS	SC	MD	NW
London	1	0.79	0.69 $\pm$ 0.10	0.81	0.84	0.82	0.74
Wales & West	0.79	1	0.75	0.79	0.87	0.81	0.76
North East	0.69	0.75	1	0.78	0.77	0.82	0.72
Yorkshire	0.81	0.79	0.78	1	0.90 $\pm$ 0.06	0.82	0.75
Scotland	0.84	0.87	0.77	0.90	1	0.81	0.74
Midlands	0.82	0.81	0.82	0.82	0.81	1	0.87
North West	0.74	0.76	0.72	0.75	0.74	0.87	1

Dla półrocznej serii czasowej cotygodniowych obserwacji (marzec-październik 1978) obliczono opóźnienia zdefiniowane jak we wzorze (6.1) oraz macierz korelacji sprzedaży między regionami (Tabela 6.5). Regionem najszybciej adaptującym nowości okazał się Północny Zachód (North West) z miastami Manchester i Liverpool (średnie opóźnienie w stosunku do premiery: 4,18 tygodnia, mediana = 3 tygodnie). Region ten z najslabszymi korelacjami odstaje od reszty Wielkiej Brytanii i najbliżej mu do Midlands.

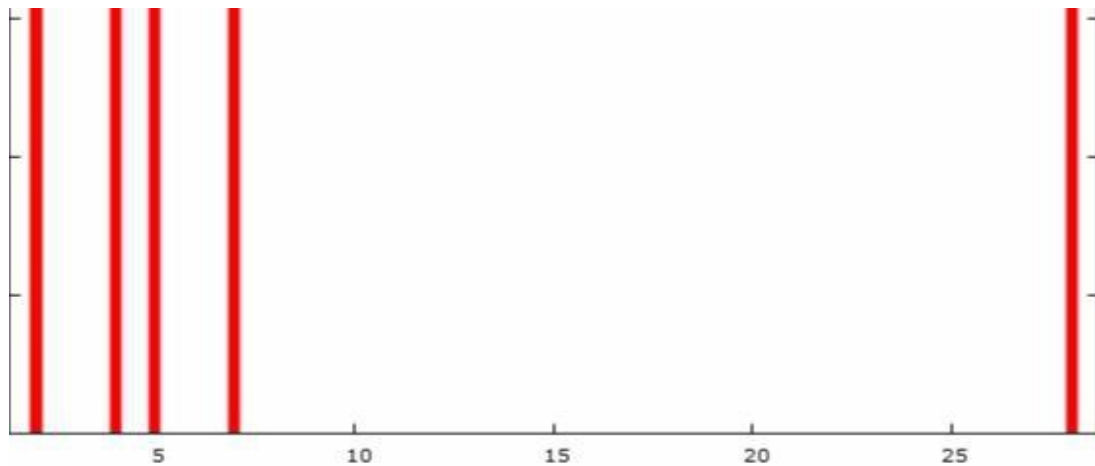
Macierz korelacji można zdiagonalizować i otrzymać jej 7 wartości własnych (Rys. 6.6). Podobnie jak w wielu sytuacjach, gdy obserwujemy niezerową średnią wartość współczynnika korelacji, spektrum wartości własnych posiada jedną silnie oddaloną od pozostałych wartość własną. Kolejne wartości własne odpowiadają za efekty drugorzędne, czyli niejednorodności wartości elementów macierzy.



**Rys. 6.8** Widmo wartości własnych, które odpowiadają macierzy korelacji między regionami

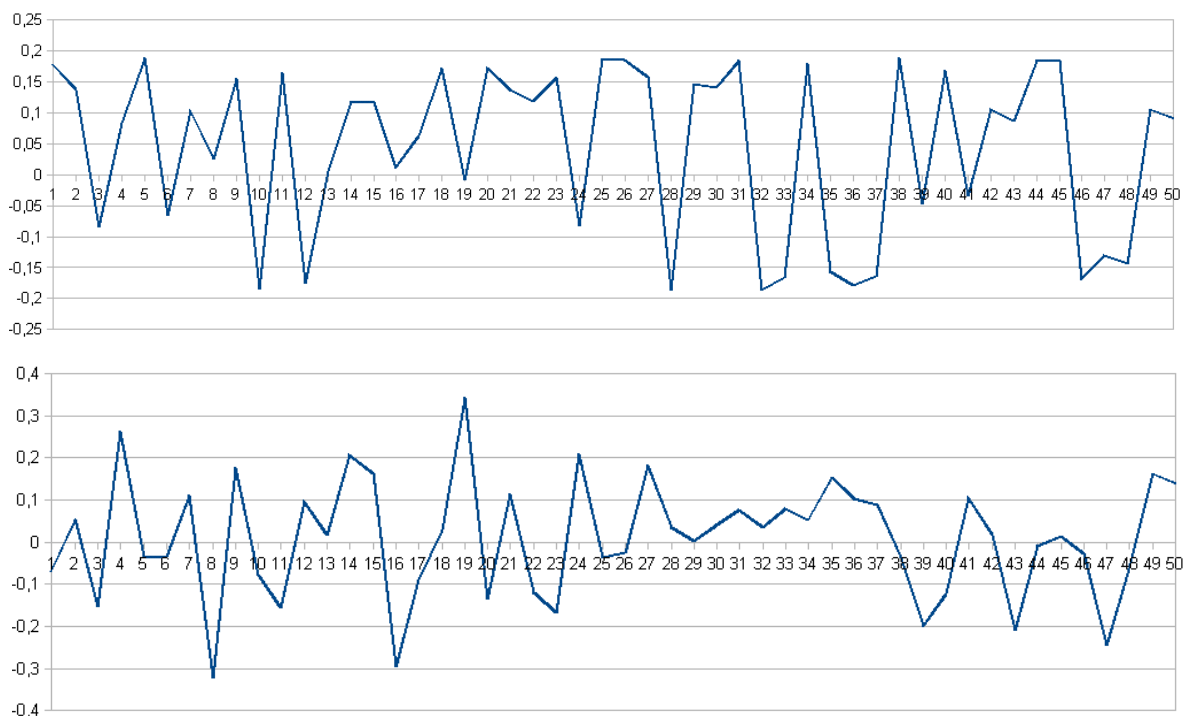
Podobną analizę wartości własnych można wykonać także dla macierzy korelacji sprzedażowych między badanymi utworami (50x50), ale wówczas uzyskana informacja będzie odzwierciedlać scenariusze zdobywania popularności przez przeboje. Rysunek 6.9 obrazuje (niezerowe) wartości własne macierzy korelacji sprzedażowych między utworami z których maksymalna, wyraźnie odstająca od pozostałych wartość wynosi 28,3.





**Rys. 6.9** Widmo niezerowych wartości własnych macierzy korelacji sprzedażowych między utworami

Wynik ten wskazuje na istnienie określonego sposobu zdobywania popularności, bowiem w badanym okresie utwory zwykle osiągały najwyższe pozycje między 3 a 7 tygodniem od premiery. Bardziej interesujący jest też przedstawiony na Rysunku 6.10 wykres sygnału własnego, odpowiadający wskazanej wartości własnej 28,3.



**Rys. 6.10** Wykres sygnału własnego, który odpowiada maksymalnej wartości własnej 28,3 macierzy korelacji między utworami (na górze), drugiej wartości własnej 7,2 (na dole) dla kolejnych 50 utworów.

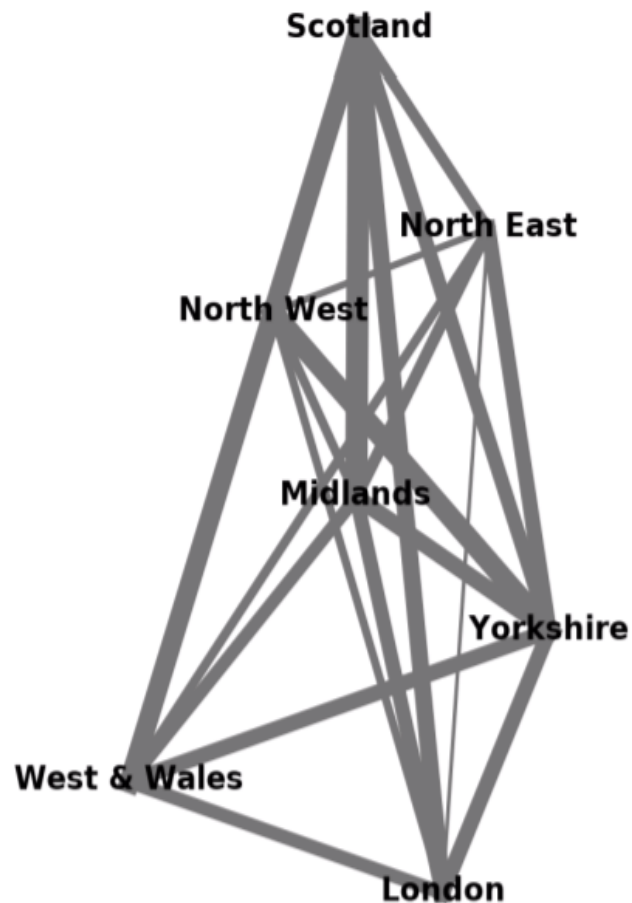
Sygnal własny  $Z_k(t_i)$  odpowiadający danej wartości własnej  $k$  z definicji [12] można obliczyć następująco:

$$Z_k(t_i) = \sum_{j=1}^{50} v_{jk} \cdot S(t_i) \quad (6.4)$$

gdzie  $v_{jk}$  oznacza składową  $j$  wektora własnego, który odpowiada  $k$ -tej wartości własnej, a  $S(t_i)$  oznacza popularność singla  $j$  w tygodniu  $t_i$ . Seria  $j$  od 1 do 50 na osi odciętych odpowiada kolejnym utworom wydanym od marca do października 1978, które trafiły na lokalne listy we wszystkich 7 brytyjskich regionach, jednak wahania sygnału własnego pokazują na powtarzające się zróżnicowanie w zdobywaniu popularności. Jedynie pod koniec serii sygnał własny ma mniejsze wahania, co odpowiada przebojom z października, a więc okresu w którym do rynku może się włączyć dodatkowa grupa konsumentów, opisana w poprzednich rozdziałach. Jednak brak danych o lokalnej sprzedaży za listopad i grudzień 1978 uniemożliwia głębsze badanie ich wpływu na sygnał własny.

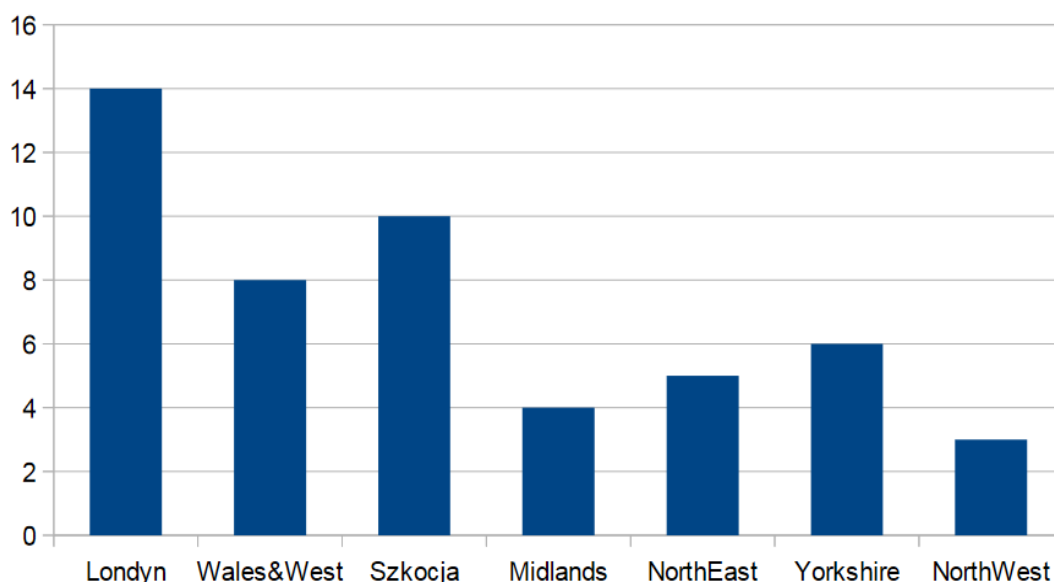
**Tab. 6.6** Średnie opóźnienia  $d_{xy}$  w tygodniach od premiery do szczytu popularności w regionach

London	$4.90 \pm 0.27$	Scotland	$4.78 \pm 0.28$
Wales & West	$4.75 \pm 0.25$	Midlands	$5.02 \pm 0.27$
North East	$5.08 \pm 0.25$	North West	$4.92 \pm 0.25$
Yorkshire	$4.92 \pm 0.28$		



**Rys. 6.11** Sieć odpowiadająca macierzy korelacji średnich opóźnień  $d_{xy}$  między regionami Wielkiej Brytanii

Co ciekawe, regionem najczęściej inicjującym przeboje jest Londyn (28%, ale prawie połowę tej liczby stanowią nowości od artystów spoza Wielkiej Brytanii). W dalszej kolejności wśród inicjatorów są stanowiące odrębne narody Szkocja i Walia. Natomiast Północny Zachód - mimo najszybszego w całym kraju przejmowania piosenek – miał w 1978 roku najmniej zainicjowanych przez siebie nowości. Warto jednak dodać, że dane Instytutu Gallupa dotyczą jedynie wycinka historii z 1978 roku (czyli nie obejmują lat 60-tych XX wieku, zdominowanych przez Liverpool, czy lat 80-tych gdzie nowości dostarczał Manchester).



**Rys. 6.12** Ilość inicjowanych przebojów przez regiony Wielkiej Brytanii (1978)

Warto dodać, że za transmisję zagranicznych utworów na rynek brytyjski odpowiada głównie Londyn (wyjątkiem jest Szkocja, która zainicjowała popularność nagrań australijskiej grupy AC/DC, której muzycy są szkockiego pochodzenia), jednak w kreowaniu popularności nowych utworów pochodzenia brytyjskiego stolica nie odbiega ilościowo od pozostałych 6 regionów kraju. Ponadto macierz średnich opóźnień (Tabela 6.6) pokazuje co najwyżej tygodniowe opóźnienia między regionami w przyjmowaniu utworów. Jednocześnie ta wynika ze zdominowania rynku brytyjskiego przez duże sieci handlowe (wyjątki od reguły jednoczesności zdarzały się rzadko, np. okazjonalna piosenka drużyny piłkarskiej Nottingham Forest była przebojem jedynie w Midlands i nigdzie indziej). Dla porównania: analogiczne odległości opóźnień  $d_{xy}$  między krajami Europy wynoszą do kilku tygodni (Rys. 6.1) i odzwierciedlają geograficzną strukturę naszego kontynentu.

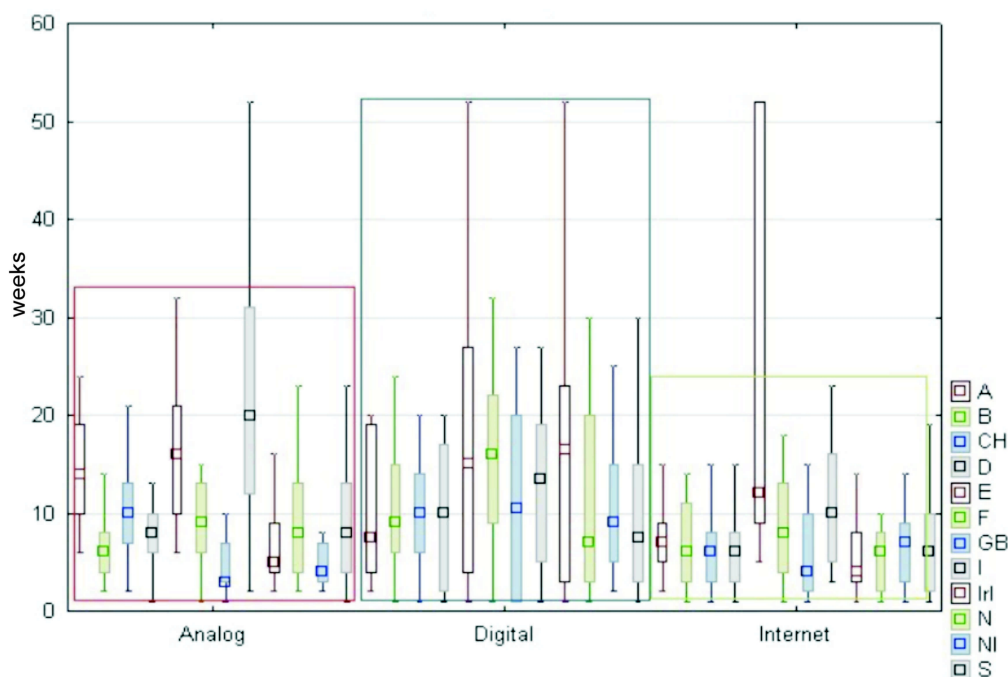
## 6.4 Sieć rynkowa a sieć kulturowa w różnych erach rozwoju nośników

Pokazane wyżej rezultaty, zarówno na mapie Europy (Rys. 6.1) jak i sieci skierowanej między lokalnymi rynkami (Rys. 6.3) wskazują po pierwsze na silną zależność geograficzną między graniczącymi ze sobą krajowymi rynkami (zwłaszcza, że dla zbadanych przeze mnie największych przebojów z lat 1966-2016 popularność w 97 przypadkach na 100 rozprzestrzeniała się właśnie według reguł geograficznych, czyli poprzez bezpośrednio sąsiadujące ze sobą kraje). Po drugie, połączenia między rynkami mogą wynikać także z języka i kultury krajów, oraz podlegać ewolucji w zależności od stopnia rozwoju rynku.

Historycznie, stopień rozwoju rynku definiuję jako:

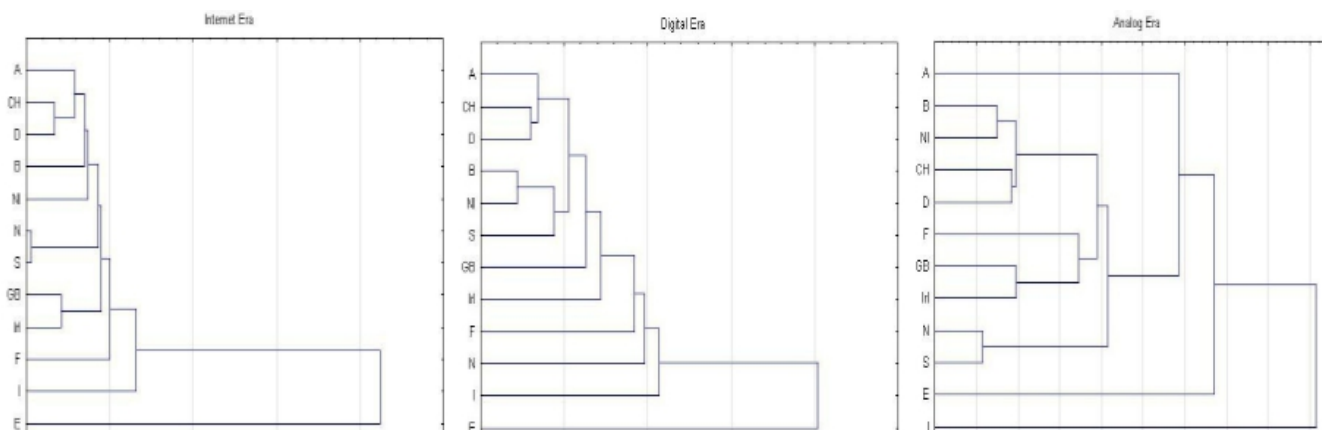
- 1) erę analogową (do 1987 roku)
- 2) erę nośników cyfrowych CD (1988-2003)
- 3) erę internetową (od 2004 roku do dziś)

Powyższy podział historyczny wynika z rodzaju nośnika, jaki dominował na globalnym rynku we wskazanych latach. Miało to oczywiście wpływ na szybkość rozprzestrzeniania się popularności utworu, bowiem w erze analogowej i cyfrowej zapotrzebowanie na singla skutkowało wytłoczeniem jego kolejnych partii (co zajmowało wiele tygodni i kolejek w tłoczniach płyt). Z kolei w erze internetowej, gdzie piosenka jako plik mp3 lub dzwonek na komórkę jest wirtualna, sprośanie zapotrzebowaniu rynkowemu następuje natychmiast (w ciągu kilku sekund).



**Rys. 6.12** Mierzone w tygodniach (*weeks*) opóźnienia w przepływie popularności przebojów między krajami w Europie Zachodniej w różnych erach: analogowej (*analog*), cyfrowej (*digital*), internetowej (*Internet*) w oparciu o pracę [93]

Na podstawie wzoru (6.1) można określić średnie opóźnienia rozprzestrzeniania się popularności singli między krajami Europy Zachodniej w różnych erach historycznych (Rys. 6.12), a także wyznaczyć diagramy hierarchiczne, stowarzyszone z drzewami minimalnej rozpiętości (Rys. 6.13). Wynika z nich, że w erze internetowej cykl życia przeboju (czyli osiągnięcie najwyższych pozycji na wszystkich 12 listach) istotnie zmalał nawet do kilku tygodni, co może być efektem braku fizycznego nośnika dla piosenki, ale także postępującej globalizacji. Co ciekawe, we wszystkich erach krajami szybko adaptującymi popularne przeboje są Skandynawia, Benelux, Wielka Brytania, Irlandia i Niemcy, a zmiany obserwujemy jedynie w przypadku klastra konserwatywnych krajów. Bowiem w erze analogowej i cyfrowej najbardziej odstającym krajem od reszty Europy były Włochy, a w erze internetowej najbardziej odstaje Hiszpania, co być może jest skutkiem globalizacji i jej bliskości kulturowej z krajami Ameryki Łacińskiej, a więc oddalania się od Europy.



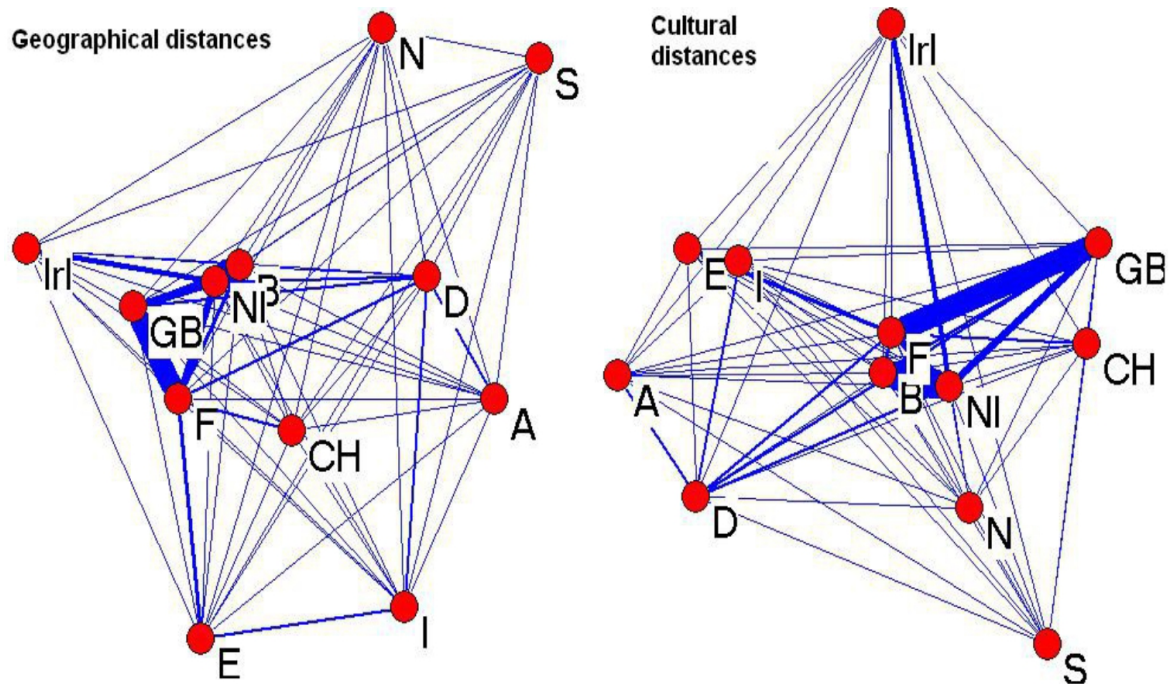
**Rys. 6.13** Diagramy hierarchiczne wyznaczone na podstawie opóźnień przepływu popularności między krajami w Europie w różnych erach: analogowej, cyfrowej, internetowej w oparciu o [93]

Ciekawe jest też porównanie powyższych wyników z odległościami kulturowymi między krajami (zdefiniowanymi w publikacji [94]), oraz ich odległościami geograficznymi (według współrzędnych geograficznych), przedstawione na Rys. 6.14, 6.15 oraz w Tab. 6.8

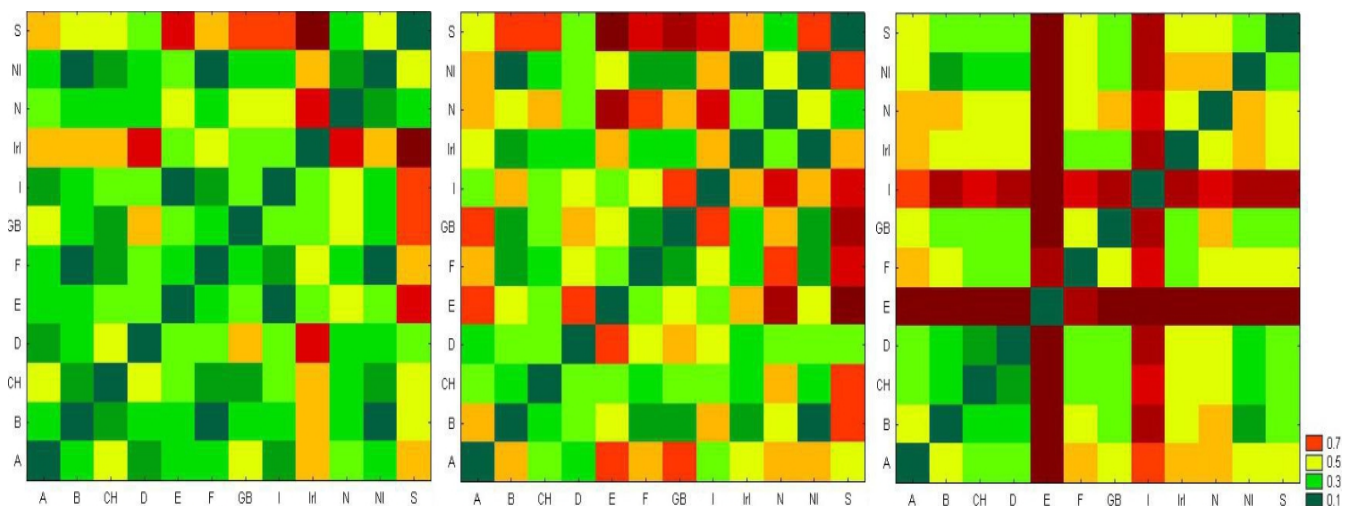
**Tab. 6.8** Korelacje między opóźnieniami w przekazywaniu popularności singli z Rys. 6.12 (średnia oraz mediana), a odległością geograficzną (zdefiniowaną według współrzędnych) i odległością kulturową, określoną według publikacji [94].

Era	Analogowa (średnia)	Analogowa (mediana)	Cyfrowa (średnia)	Cyfrowa (mediana)	Internetowa (średnia)	Internetowa (mediana)
Geograficzna	<b>0.56</b>	<b>0.63</b>	<b>0.12</b>	<b>-0.28</b>	<b>0.24</b>	<b>0.34</b>
Kulturowa	<b>-0.12</b>	<b>-0.32</b>	<b>0.26</b>	<b>0.50</b>	<b>-0.09</b>	<b>0.33</b>

Otóż w erze analogowej (1966-1987) zachodzi wyraźna korelacja między opóźnieniami w przekazywaniu popularności singla, a odległością geograficzną. W erze cyfrowej (1988-2003) bardziej niż geograficzne, liczą się odległości kulturowe. Dopiero w erze internetowej (po 2004 roku) wszystkie badane korelacje słabną. Przedstawione wyniki odzwierciedlają więc wpływ postępującej globalizacji na lokalne rynki i szybkość rozprzestrzeniania się popularności singli. Zagadnienie to można w ogólności rozszerzyć także na inne produkty i towary, zdobywające popularność na wielu sąsiadujących ze sobą rynkach, analogicznie do rozprzestrzeniania się mody lub epidemii.



**Rys. 6.14** Odległość geograficzna (po lewej) i kulturowa (po prawej) między badanymi rynkami na podstawie publikacji [94]. Wagi połączeń odpowiadają ilościowo wymianie utworów między krajami, omówionej w artykule [93]



**Rys. 6.15** Normalizowana odległość kulturowa (po lewej), geograficzna (na środku) w porównaniu z normalizowaną odległością pod względem opóźnień w przyjmowaniu utworów (po prawej) między badanymi rynkami na podstawie publikacji [94] oraz artykułu [93]



## Rozdział 7. Ścieżki zdobywania popularności na wielu rynkach - modelowanie agentowe

### 7.1 Ścieżki zdobywania popularności między rynkami

W poprzednim rozdziale pokazałem ilościowo oddziaływanie na siebie wielu rynków i wynikające z niego ścieżki zdobywania popularności. Badania przeprowadziłem dla singli w 12 graniczących ze sobą krajach Europy Zachodniej i dodatkowo wewnątrz jednego rynku (brytyjskiego). Podobne rozważania warto rozszerzyć na ogólny przypadek utworu, którego popularność rozprzestrzenia się w Internecie, najpierw na lokalnym rynku, a następnie globalnym. Jak wiadomo, Internet jako sieć połączeń między użytkownikami ma strukturę bezskalową [19-20], dlatego do jego opisu w modelu zostały użyte sieci Barabasięgo-Albert [18, 95], na których podstawie są tworzone lokalne rynki (podsieci, opisane macierzą sąsiedztwa), składające się na rynek globalny. Natomiast proces zdobywania popularności rynkowej (rozumiany jako przebieg epidemii lub cykl życia produktu) rozważany jest w epidemiologicznym modelu agentowym, w którym węzły sieci (agenci) przekazują popularność produktu swoim najbliższym sąsiadom.

Założenia modelu przewidują, że rynek globalny składa się z ustalonej liczby  $M = 100$  rynków krajowych  $M_i$  (numerowanych od  $i = 1$  do 100). Każdy rynek jest oddzielną siecią bezskalową o strukturze Barabasięgo-Albert i nieukierunkowanych, binarnych krawędziach. Ta konkretna architektura niekoniecznie odtwarza wszystkie szczegóły struktury rzeczywistych uczestników rynku, który jest znacznie bardziej złożony niż model BA, ale obejmuje jeden z kluczowych aspektów kontaktów międzyludzkich, czyli uwzględnia heterogeniczność węzłów o dużych wahaniami liczby połączeń.

Rynki różnią się wielkością (zdefiniowaną przez liczbę  $N$  węzłów) w zakresie obejmującym dwa rzędy wielkości:  $100 \leq N \leq 10\ 000$ . Wartości  $N$  są przypisane w taki sposób, aby wartości  $\log N$  były równomiernie próbkowane na przedziale (2,4). Badane są różne wartości parametru liczby połączeń węzła sieci BA  $m$  ( $m$  określa liczbę krawędzi binarnych, którymi każdy nowy węzeł jest połączony z już istniejącą częścią sieci):  $1 \leq m \leq 10$ . Nie rozpatrujemy większych wartości parametru  $m$ , ponieważ sieci uczestników rynku nie powinny być zbyt gęste.

W każdej sieci istnieje uprzywilejowana grupa, odpowiadająca za przyjęcie utworu z zewnątrz jako pierwsza (np. profesjonalni influencerzy czy wpływowi, kosmopolityczni trendsetterzy), która stanowi także węzeł o największej liczby połączeń (*hub*). Można go utożsamiać z centrum i połączyć z innymi centrami we wszystkich pozostałych sieciach. Założenie to wynika z faktu, że lokalne narodowe rynki fonograficzne są scentralizowane jeśli chodzi o dystrybucje muzyki i jej promocję. Taki węzeł centralny może odpowiadać za przejmowanie z zewnątrz popularności produktu i przekazywanie grupom najbardziej połączonym z tym centrum. Z kolei za samo wprowadzenie utworu na rynek może odpowiadać dowolny, losowo wybrany węzeł tak skonstruowanych sieci – nawet węzeł peryferyjny (o małej krotności).

Dzięki temu możemy stworzyć hierarchiczną sieć globalną o dwupoziomowej strukturze. Na jej wyższym poziomie mamy w pełni połączoną "supersieć" złożoną z "sieci krajowych", podczas gdy na niższym poziomie mamy szereg znacznie bardziej rozproszonych, niezależnych od skali sieci pojedynczych węzłów. Połączeniami supersieci są unikalne ścieżki, dzięki którym infekcja może rozprzestrzeniać się z jednego rynku na drugi. W celu odtworzenia głównych cech rynku rzeczywistego, krawędziom supersieci przypisane są wagi  $w_{ij}$ , które wyrażają kulturowe lub geograficzne odległości między krajami  $i$  oraz  $j$ . Im większa waga, tym silniejsza bliskość kulturowa lub geograficzna. Wagi są symetryczne ( $w_{ij} = w_{ji}$ ) i wylosowane z pewnego rozkładu na przedziale  $(0,1)$ .

Oprócz bliskości, każdy rynek  $i$  posiada specyficzny potencjał oddziaływania związany z jego rozmiarem  $N_i$ . W celu jego uwzględnienia stosujemy wielkość:  $z_{ji} = N_i / (N_i + N_j)$ , której wartości są ograniczone do  $0 < z_{ji} < 1$ . Jeżeli dwa rynki  $i$  oraz  $j$  znajdują się tak blisko siebie, że  $w_{ij} \approx 1$  i jeśli  $N_i \gg N_j$ , infekcja może łatwo rozprzestrzenić się w kierunku od  $i$  do  $j$ . Rozprzestrzenianie infekcji w kierunku przeciwnym:  $j \rightarrow i$ , staje się jednak mało prawdopodobne ponieważ wartość  $z_{ij}$  jest zbyt mała, gdyż oprócz bliskości kulturowej lub geograficznej uwzględnia także rozmiary obu rynków. Cecha ta przekształca globalną sieć z nieukierunkowanej na ukierunkowaną.

Bliskość i rozmiar mogą być postrzegane jako właściwości rynkowe, które są niezależne od wirusa. Z drugiej strony, potencjał komercyjny utworu jest wyrażony przez atrakcyjność (agresywność) wirusa  $a$ : im większa  $a$ , tym wirus jako produkt jest bardziej agresywny (czyli atrakcyjny). Badania w niniejszej pracy zostały przeprowadzone dla

zakresu wartości:  $0.01 \leq a \leq 90.0$ .

Co więcej, nawet bardzo agresywny wirus ma większe szanse na rozprzestrzenienie się na inne rynki, jeśli zainfekował już znaczną część swojego macierzystego rynku. Dlatego warto wprowadzić też współczynnik sukcesu rynkowego, zdefiniowany jako  $r_i = N_i^*/N_i$ , gdzie  $N_i^*$  oznacza liczbę zainfekowanych węzłów rynku  $i$ . Niech  $S_{ph}(i)$  oznacza stan stołecznego węzła (*hubu*) rynku  $M_i$ . Biorąc to wszystko pod uwagę, można zdefiniować czynnik  $I_{ij}(a, t)$ , który opisuje zdolność wirusa o agresywności  $a$  obecnego na rynku  $i$  do rozprzestrzeniania się na rynku  $j$  w czasie  $t$ :

$$I_{ij}(a, t) = w_{ij} z_{ij} a r_i(t) S_{ph}^{(i)} = w_{ij} \frac{N_i}{N_i + N_j} a \frac{N_i^*(t)}{N_i} S_{ph}^{(i)} = w_{ij} \frac{N_i^*(t)}{N_i + N_j} a S_{ph}^{(i)} \quad (7.1)$$

Ponieważ liczba zarażonych węzłów  $N_i^*(t)$  zależy od czasu, współczynnik  $I_{ij}(a, t)$  jest zależny od czasu również. Warto zauważyć, że wirus może się rozprzestrzeniać między rynkami wtedy i tylko wtedy, gdy zarażony jest węzeł centralny jednego z nich. Proces zainfekowania węzła może być utożsamiany z czynnością nabycia utworu. Dlatego każdy z węzłów w rozumieniu modeli opisanych skrótem SIR może być w stanie: podatnym (*Susceptible*), zarażonym (*Infected*) lub usuniętym (*Removed*).

Węzły te mogą być postrzegane jako reprezentujące uczestników rynku, którzy mogą być identyfikowani z grupami lub osobami, które zarówno promują, jak i konsumują (kupują) muzykę w tym samym czasie. Na rzeczywistym rynku influencerzy (mogą to być nawet media w ogóle) którzy promują muzykę niekoniecznie muszą być zaliczani do typowych konsumentów, ponieważ otrzymują często darmowe egzemplarze promocyjne utworów i ich decyzje o zakupie muzyki dotyczą jedynie utworów trudno dostępnych, które nie są dla nich darmowe. Dla uproszczenia jednak warto założyć, że zachowują się oni jak konsumenci z większą liczbą powiązań niż typowi konsumenci. Dlatego też każdy agent w modelu zarówno kupuje, jak i promuje utwory muzyczne jednocześnie wśród swoich znajomych i grupy docelowej. Krawędzie sieci wewnątrzrynkowej odpowiadają więziom międzyludzkim wśród agentów oznaczonych parami węzłów (nie definiujemy tych więzi, wystarczy, że dzięki powiązaniom grupa  $i$  może promować muzykę dla grupy  $j$ ).

Węzeł pozostaje w stanie zarażenia przez  $d$  kolejnych kroków czasowych, co może odpowiadać czasowi transakcyjnemu, w którym agent jest zdolny do promowania utworu muzycznego w swoim otoczeniu powiązań. Z perspektywy promocji muzyki można wyobrazić sobie inny paradygmat epidemii, tj. podatny-zakażony (SI). W tym podejściu, gdy węzeł zostanie zainfekowany, pozostaje on w tym stanie na zawsze i nigdy nie przechodzi do stanu usuniętego (można go traktować jako węzeł zarażony - szczególny przypadek naszego modelu z bardzo dużym  $d$ ). Odpowiada to pozornie realistycznej sytuacji, w której agent, który lubi i kupuje utwór muzyczny, będzie go promował w dowolnym momencie w przyszłości (rodzaj propagacji informacji [96]). Jednakże, gdyby ten paradygmat został włączony do modelu, prowadziłoby to do ekonomicznie nierealistycznego przypadku, w którym w końcu wszystkie węzły sieci są zainfekowane (tzn. wszyscy agenci kupują każdy utwór). Dlatego też podejście SIR wydaje się lepsze.

Każda infekcja rozpoczyna się od jednego lub więcej węzłów początkowych na wybranym rynku  $M_{i_0}$ . W każdej sieci BA reprezentującej dany rynek  $M_i$ , wirus może być przenoszony pomiędzy węzłami na tej samej zasadzie, ale z różnymi prawdopodobieństwami. Zakładamy, że spośród czterech czynników rządzących infekcją pomiędzy rynkami ( $w_{ij}$ ,  $z_{ij}$ ,  $r_i$ ,  $a$ ) tylko dwa poniższe mają znaczenie na poziomie wewnątrzrynkowym:

- (1) agresywność wirusa  $a$
- (2) bliskość kulturowa  $w_{i_0j}$  danego rynku  $M_j$  wobec rynku  $M_{i_0}$ .

Model zakłada też, że jeśli wirus zdołał już dotrzeć na jakiś rynek, jego dalsze rozprzestrzenianie się wśród węzłów nie zależy już ani od wielkości rynku  $M_{i_0}$ , ani od współczynnika sukcesu  $r_{i_0}$ . Wewnątrz pojedynczego rynku  $M_i$ , który ma strukturę sieci BA, wirus rozprzestrzenia się z węzła na węzeł według tych samych reguł, jednak z inaczej określonym prawdopodobieństwem.

Dla wirusa (utworu) o określonej agresywności (atrakcyjności) prawdopodobieństwo, że węzeł  $p$  zostanie zarażony zależy zarówno od stanu  $S_q(j)$  sąsiedniego węzła  $q$  jak i od współczynnika krotności (*connectivity*)  $k_p(j)$  węzła  $p$ :

$$P_{pq}^{(ij)} = \frac{w_{ij} S_q^{(j)}}{k_p^{(j)}} a \quad (7.2)$$

gdzie parametr  $S_q(j) = 1$  oznacza, że węzeł jest zarażony (*infected*), z kolei  $S_q(j) = 0$  oznacza, że zostaje on po przebyciu infekcji usunięty (*removed*). Oczywiście w ramach macierzystego rynku  $M_{i_0}$  we wzorze (7.2) należy założyć, że waga bliskości kulturowej oznacza tożsamość  $w_{ii} = 1$ .

Założenie, że prawdopodobieństwo zarażenia węzła jest odwrotnie proporcjonalne do jego krotności  $k_p(j)$  może być rozumiane jako indywidualna odporność węzła na zarażenie i odzwierciedla fakt, że rynek jest niejednorodny. Z ekonomicznego punktu widzenia założenie to uwzględnia ważne zasady wolnorynkowej konkurencji, gdyż agenci o szerszych kontaktach mają więcej sąsiadów i większe możliwości wyboru. Oznacza to, że wśród sąsiadów takiego węzła o dużej krotności zajdzie większa konkurencja. Z epidemiologicznego punktu widzenia węzeł taki jest bardziej odporny na zarażenie, co jest odzwierciedleniem zjawiska, że specjaliści o szerokim dostępie do źródeł muzyki wybierają bardziej wyrafinowane, atrakcyjne utwory. Z kolei węzły pozbawione dużej liczby sąsiadów lub peryferyjne mają mniejszy wybór, bo między ich sąsiadami konkurencja już nie zachodzi. Dlatego są mniej odporne na oferowane im produkty.

Powyższe założenie konkurencyjności pozwala uprościć model do sytuacji, w której jeden wirus rozpoczyna infekcję z wybranego losowo węzła. Oczywiście w tak zdefiniowanej sieci możliwe jest także obserwowanie wielu jednoczesnych infekcji, jednak w przypadku cyklu życia produktu jakim jest utwór muzyczny warto założyć, że poszczególne piosenki podobają się odbiorcom niezależnie od siebie, tj. fakt, że jeden utwór zdobył nasze polubienie nie będzie umniejszał naszego polubienia dla innych piosenek, które – co jest jak najbardziej zgodne z rzeczywistością - mogą podobać nam się przecież jednocześnie, bez wzajemnej szkody dla siebie. Dlatego w parametrach modelu uwzględniamy:

- agresywność wirusa  $a$ ,
- początkową liczbę zarażonych węzłów  $n_{seed}$ ,
- czas zarażenia węzła  $d$  (czas transakcyjny) po którym węzeł jako „wyleczony” zostaje usunięty (*removed*)

Zakładam ponadto, że wszystkie pozostałe węzły sieci są podatne (*susceptible*) na infekcję. Wówczas wybieramy rynek źródłowy  $M_{i_0}$ , w którym węzły już na początku są

zarażone, czyli  $p_{0k}$  ( $k = 0, 1, 2, \dots, n_{\text{seed}}$ ) są w stanie  $S_{p_{0k}}(i_0) = 1$ . Odtąd stany wszystkich węzłów w kolejnych krokach czasowych podlegają ewolucji opisanej równaniami (7.2) oraz (7.1). Może zdarzyć się, że wśród początkowych zarażonych węzłów znajdzie się węzeł centralny  $p_h(i_0)$  - czyli hub rynku  $M_{i_0}$  i wtedy infekcja przedostanie się na inne rynki, od samego początku zarażając jednocześnie rynek macierzysty. W przeciwnym razie, gdy węzeł centralny jest jedynie podatny na infekcję (*susceptible*), to przekaże ją innym rynkom dopiero wtedy, kiedy zostanie zarażony od węzłów rynku macierzystego. Oczywiście w ten sposób, przy odpowiednio dużej agresywności  $a$ , wirus jest w stanie opanować wszystkie rynki.

Iteracja zatrzymuje się w momencie, gdy poprzednio zakażone węzły zostają usunięte i źródło wirusa w sieci przestaje istnieć. Ze względów statystycznych, dla raz ustalonych parametrów, każda iteracja powtarzana jest 100 razy. Można sobie wyobrazić, że węzły zarażone w chwili początkowej mogą być wybrane w dowolnym miejscu takiej sieci, gdyż muzyka może być tworzona i promowana w każdym miejscu na świecie. W dalszej części pracy przedstawię najbardziej interesujące przypadki.

Tak skonstruowany model przypomina nieco ekologiczne modele metapopulacji [37, 38], w których gatunki charakterystyczne dla danego regionu mogą okupować nowe, przestrzennie odległe habitaty przez określony czas, a następnie być tam uznane za wymarłe. Model agentowy rynku fonograficznego prędzej czy później prowadzi do ustania infekcji niezależnie od tego, ile rynków zostało zarażonych po drodze, gdyż odzwierciedla on klasyczny cykl życia przeboju jako produktu – bez cykliczności, która może zdarzać się w modelach metapopulacji (i niektórych produktach o cyklicznej przydatności do użycia).

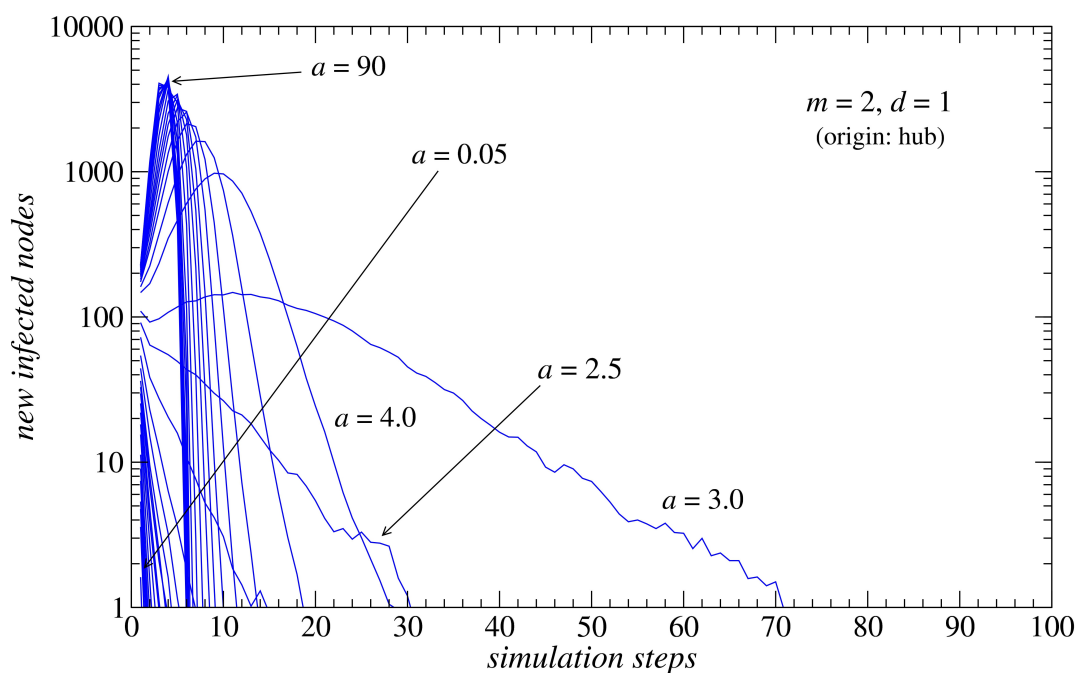
## 7.2 Model zdobywania popularności dla rynku lokalnego

Na początek warto rozważyć infekcję w ramach pojedynczego, lokalnego rynku, czyli założyć że rynek  $M_0$  jest izolowany od pozostałych i posiada maksymalną liczbę węzłów  $N_0 = 10\,000$ , przy czym struktura  $M_0$  jest skonstruowana według algorytmu Barabasiego-Albert dla  $m = 2$  (struktury dla wartości innych, niż  $m = 2$  zostaną omówione w dalszej części). Najbardziej interesujące jest sprawdzenie, jaki wpływ na liczbę zarażonych węzłów w kolejnych krokach ma agresywność wirusa  $a$ , w tym celu rozważmy 2 przeciwne do siebie przypadki:

- 1) początek infekcji wychodzi z centralnego węzła (*hub*)
- 2) początek infekcji wychodzi z peryferyjnego węzła

Pierwszy z nich odzwierciedla sytuację, gdy utwór już na starcie promowany jest przez media głównego nurtu, drugi – sytuację przeciwną, gdy piosenka należy do niszy gatunkowej i na starcie opanowała jedynie peryferyjne węzły rynku. Oczywiście przypadek pierwszy jest uprzywilejowany, co potwierdzają wyniki (Rys. 7.1) dla różnych  $a$  i  $d$ .

W tradycyjnym modelu rynkowym najbardziej odpowiadająca rzeczywistości jest wartość  $d = 2$ , oznaczająca sytuację konsumenta, który po nabyciu towaru jest w stanie zarażenia jeszcze drugi tydzień, podczas którego dzieli się z najbliższymi sąsiadami atrakcyjnością kupionego produktu. Wówczas atrakcyjność  $a$  decyduje o nastawieniu sąsiadów (zmianie stanu) do tego produktu. Jednocześnie akt sprzedaży towaru determinuje u sprzedawcy zamówienie kolejnej partii towaru na kolejny okres. Zjawisko to jest typowe dla rynków towarowych, a źródłem zapotrzebowania na produkt jest tzw. „marketing szeptany” (*whisper marketing*), charakteryzujący się na ogół dużą wiarygodnością [96]. Na rynku fonograficznym czas zazwyczaj mierzony jest w tygodniach, jednak aż do połowy lat 80. XX wieku we Francji listy popularności publikowano co miesiąc, a w Niemczech i Austrii co 2 tygodnie. Dlatego w modelu warto traktować kolejne kroki czasowe w sposób ogólny i uniwersalny, zwłaszcza że w czasach Internetu, podobnie jak na rynkach finansowych, czas transakcyjny i towarzyszące mu ilościowe informacje (np. wyświetlenia piosenek i streaming) mogą być dostępne w czasie rzeczywistym.



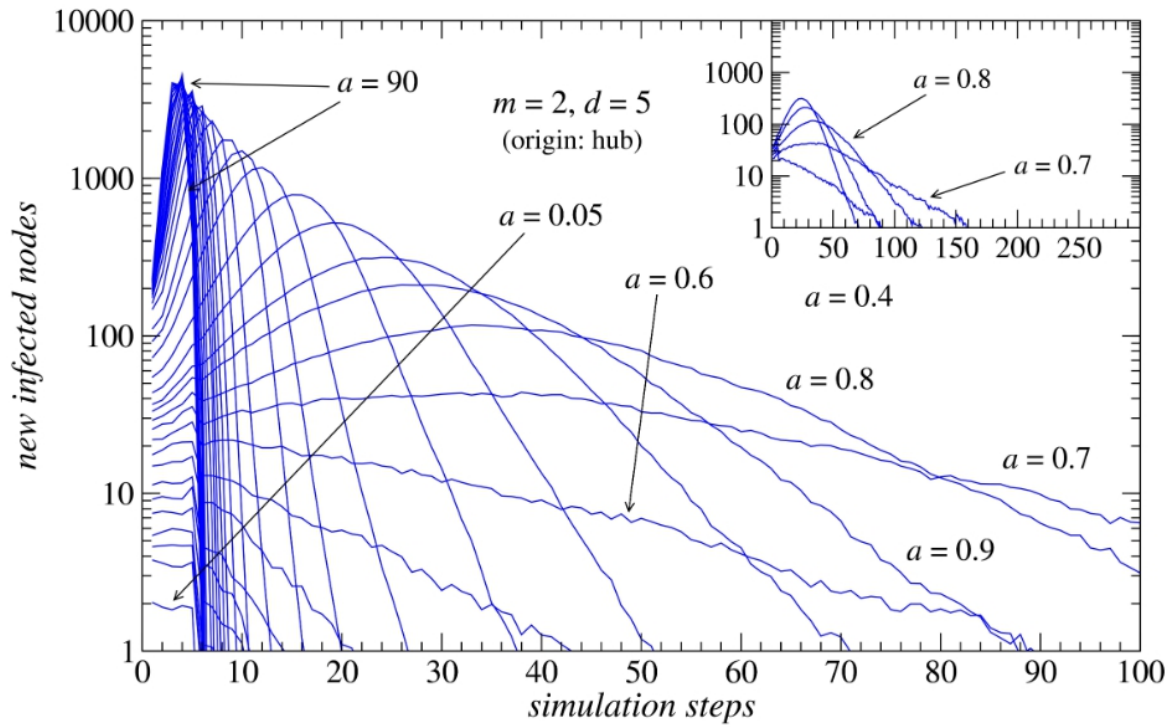
**Rys. 7.1** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach symulacji (*simulation steps*) dla różnych  $a$  i  $d=1$  (u góry) oraz  $d = 2$  (na dole) gdy źródłem infekcji jest węzeł centralny (*hub*).

W skrajnym przypadku można sobie wyobrazić sprzedaż produktu dla czasu transakcyjnego  $d = 1$ . Oznacza to, iż:

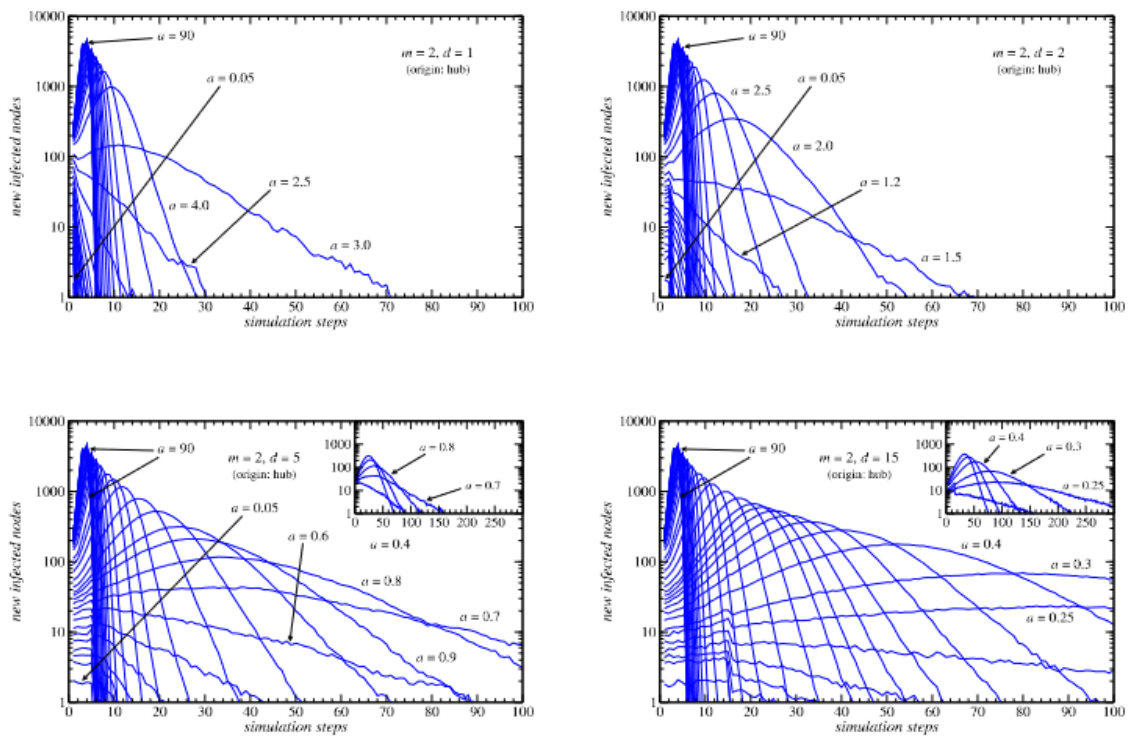
- konsument po nabyciu towaru, w drugim tygodniu nie podzielił się informacją o produkcie z sąsiadami,
- sprzedawca po dokonaniu sprzedaży zamówił kolejną partię towaru do sklepu

Zjawisko to w praktyce występuje rzadko i najczęściej oznacza, że wytwórca produktu anonimowo kupuje od sprzedawcy partię własnego towaru, aby sztucznie zwiększyć zamówienia na własny produkt. Na rynku fonograficznym znany jest przypadek z 1962 roku, kiedy manager Brian Epstein kupił 10 000 egzemplarzy pierwszego singla 'Love Me Do' kierowanego przez siebie zespołu THE BEATLES tylko po to, aby umieścić go na 17 miejscu brytyjskiej listy sprzedaży singli i sprowokować kolejne zamówienia także poza Liverpooliem (gdzie zespół na żywo występował). Dla potrzeb modelu warto zbadać całą różnorodność sytuacji i warianty rynkowej rzeczywistości, w których atrakcyjność utworu  $a$  rozumiana jako agresywność wirusa mieści się w szerokim zakresie wartości, podobnie jak czas transakcyjny  $d$ , rozumiany jako czas przebywania węzła w stanie zainfekowania.

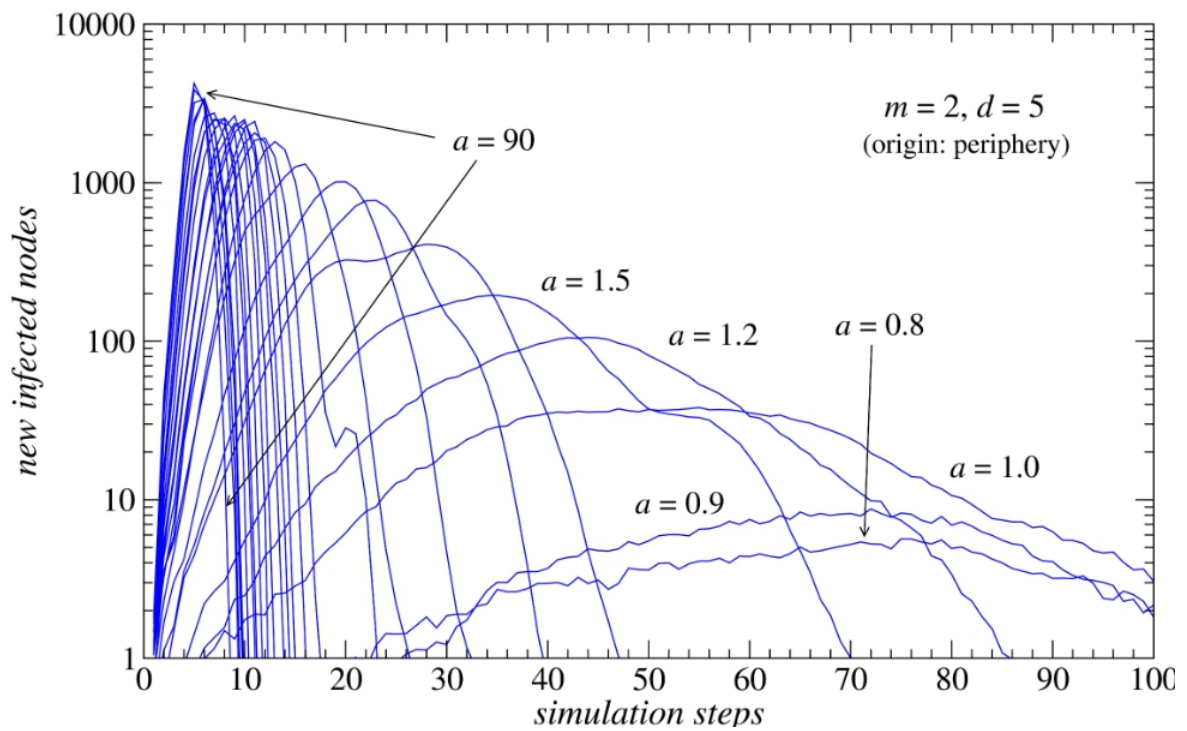




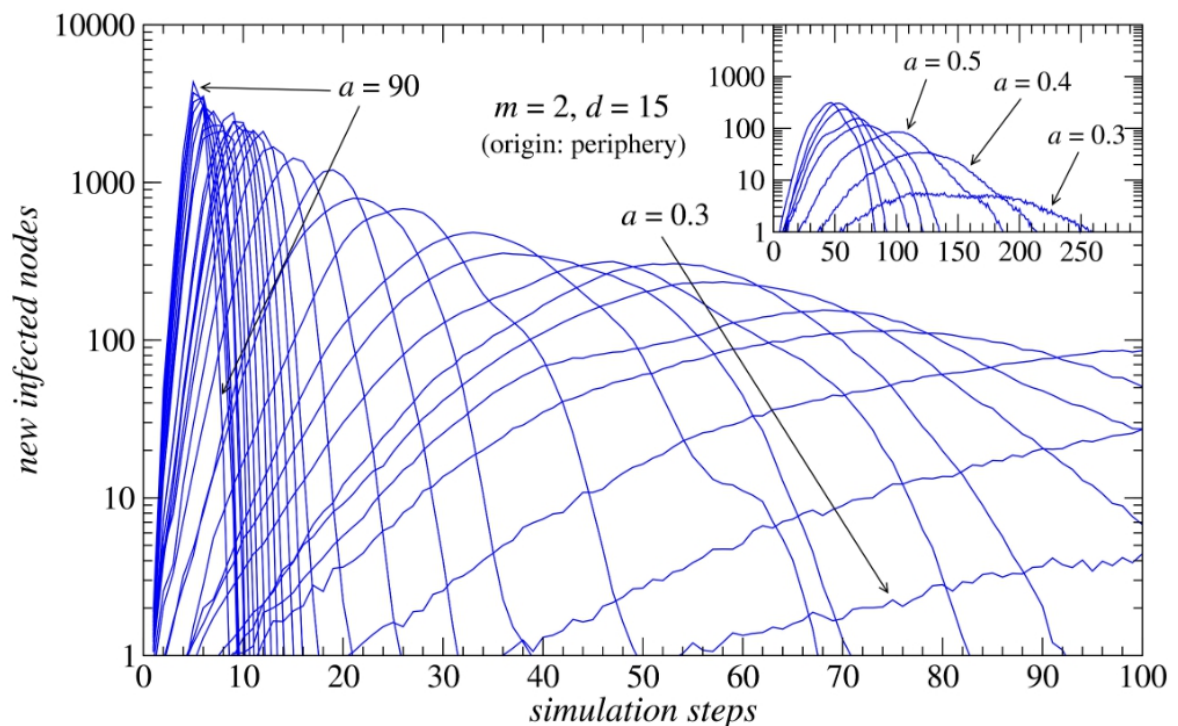
**Rys. 7.2** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach symulacji (*simulation steps*) dla różnych  $a$  i  $d=5$  gdy źródłem infekcji jest węzeł stołeczny (*hub*).



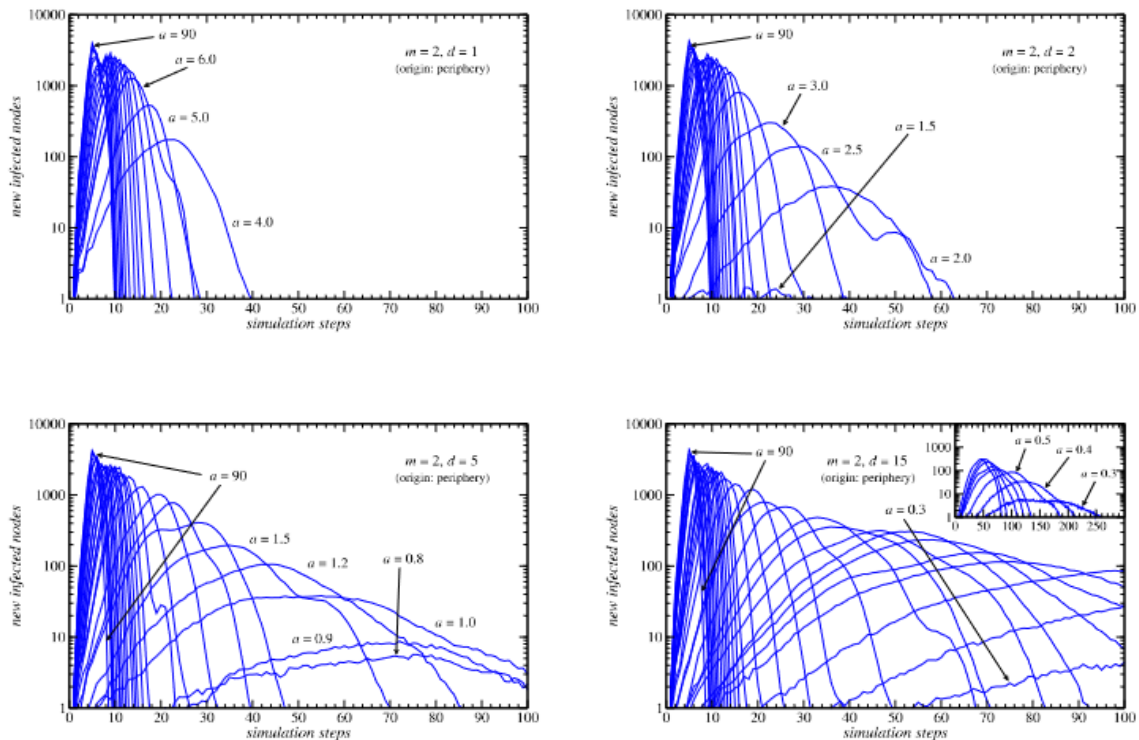
**Rys. 7.3** Podsumowanie przebiegu ścieżek z poprzednich rysunków w kolejnych krokach symulacji (*simulation steps*) dla różnych  $a$  i  $d$  gdy źródłem infekcji jest węzeł stołeczny (*hub*).



**Rys. 7.4** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach symulacji (*simulation steps*) dla różnych  $a$  i  $d=2$  (u góry) oraz  $d = 5$  (na dole) gdy źródłem infekcji jest węzeł peryferyjny (*origin: periphery*).



**Rys. 7.5** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach symulacji (*simulation steps*) dla różnych  $a$  i skrajnego  $d=15$  gdy źródłem infekcji jest węzeł peryferyjny (*origin: periphery*).



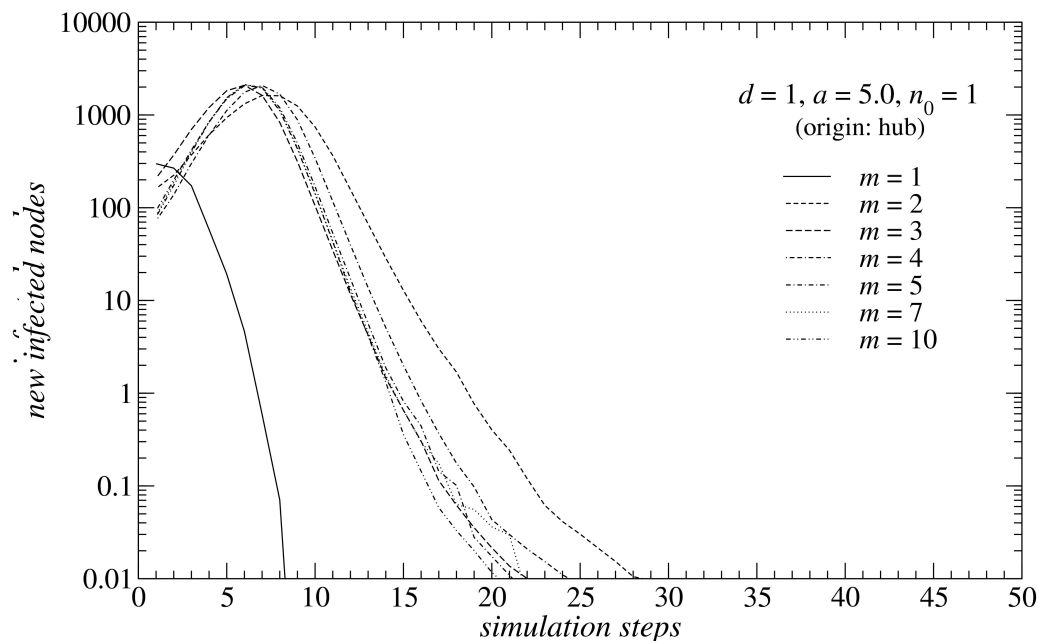
**Rys. 7.6** Podsumowanie przebiegu ścieżek z poprzednich rysunków w kolejnych krokach symulacji (*simulation steps*) dla różnych  $a$  i  $d$  gdy źródłem infekcji jest węzeł peryferyjny (*origin: periphery*).

Przedstawione wykresy pokazują, że dla danego  $d$  zwiększanie  $a$  powoduje wzrost liczby zarażonych węzłów, co odpowiada wykresom cyklu życia produktu. Jeśli wirus początkowo pojawi się w hubie, a jego agresywność jest powyżej pewnego progu  $a_{\min}$  (potrzebnego, by zarazić sąsiedni węzeł), to dla  $d < 5$  maksimum liczby nowo zarażonych węzłów w kolejnych krokach przypada zwykle na krok pierwszy, lub kilka kolejnych, a następnie maleje do zera. Odzwierciedla to z grubsza przebieg epidemii wyrażonej jako miejsce piosenki na liście najlepiej sprzedawanych singli w kolejnych tygodniach od premiery, omówione w poprzednim rozdziale.

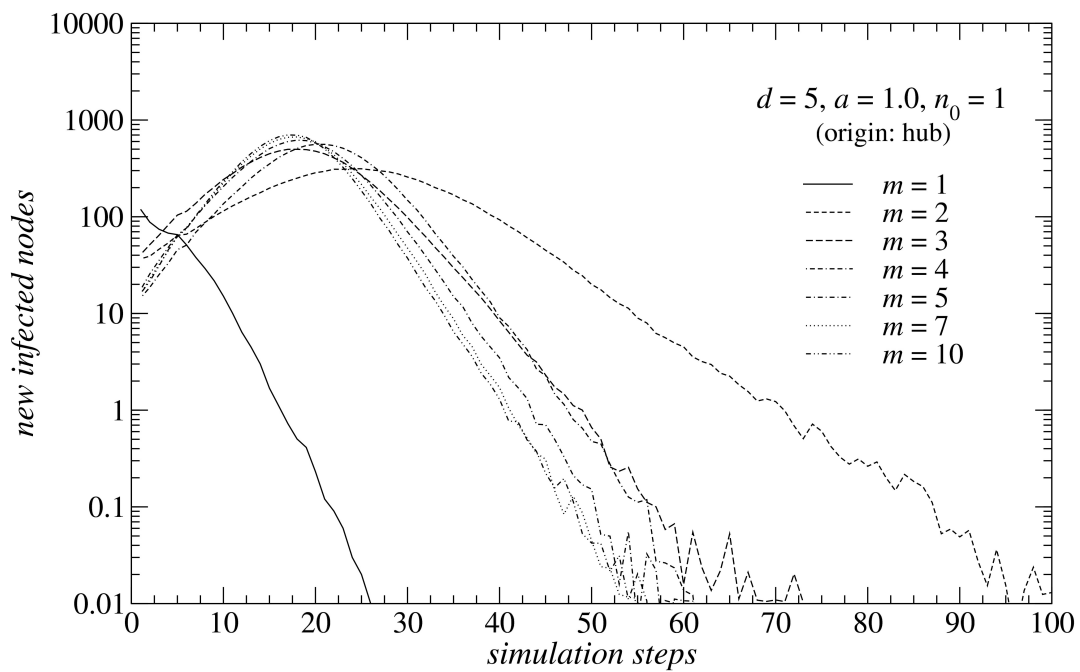
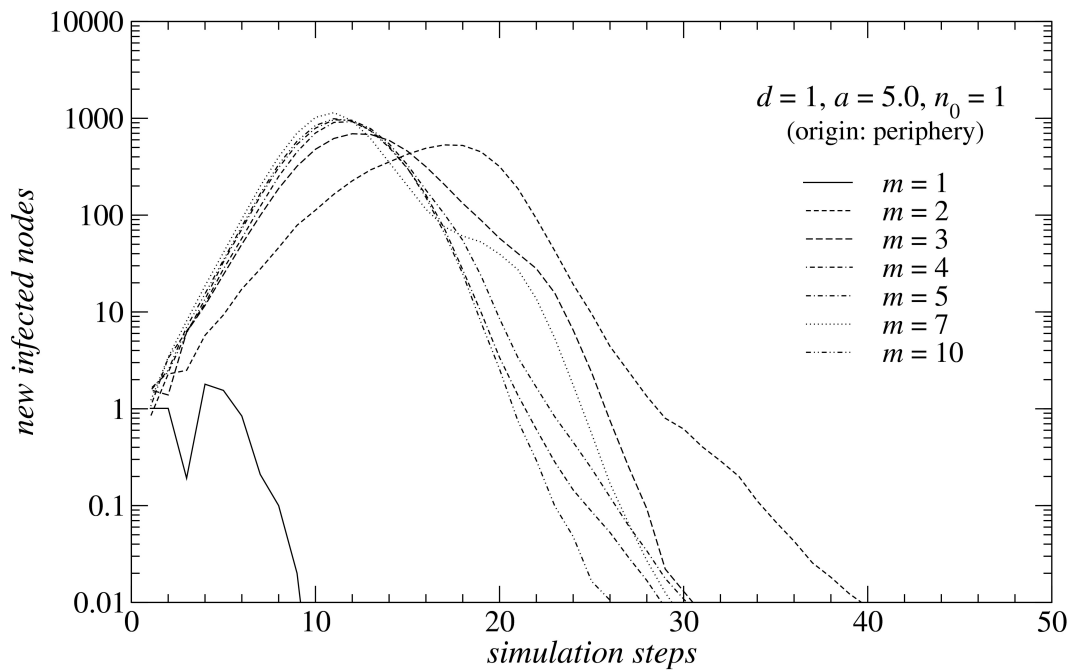
Z punktu widzenia epidemiologii, dla danego  $d$ , ciekawy jest czas trwania infekcji w zależności od agresywności wirusa  $a$ , ponieważ z wykresów wynika, że w niektórych przypadkach istnieje  $a$  dla którego czas trwania epidemii jest maksymalny i zbyt duże czy też zbyt małe  $a$  powoduje jego skracanie. Taka sytuacja ma miejsce przy odpowiedniej kombinacji parametrów  $d$  oraz  $a$ . Przykładowo: dla małego  $d = 1$  infekcja rozpoczynająca się w hubie będzie trwała najdłużej jeśli  $a = 3$ , a w przypadku startu z peryferii jej czas trwania będzie najdłuższy przy  $a = 4$  (Rys. 7.6). Z kolei dla dużych  $d = 15$  epidemia zaczynająca się od huba będzie trwała najdłużej (300 kroków czasowych) dla stosunkowo małej agresywności wirusa  $a = 0,25$  (Rys. 7.3) a epidemia wychodząca z peryferii (260

kroków) dla  $a = 0,3$  (Rys. 7.5).

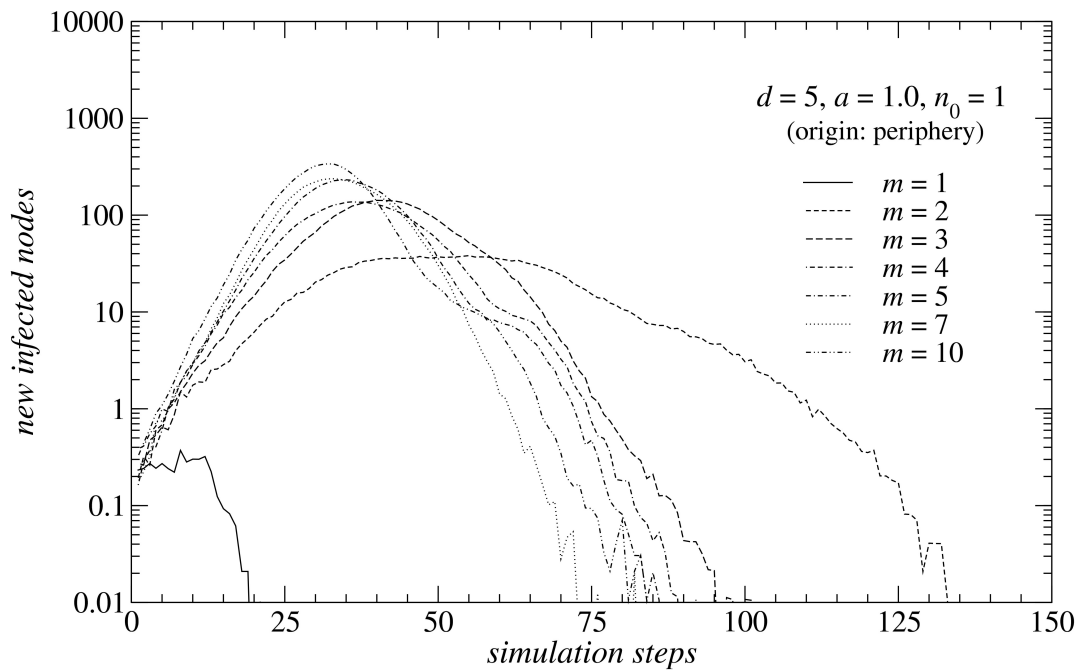
Granice stosowalności modelu agentowego warto zbadać także z punktu widzenia modelu Barabasięgo-Albert, czyli rozważenia także innych wartości parametru preferencyjnego przyłączania  $m$  (Rys. 7.7-7.9) niż stosowane dotychczas  $m = 2$ . Przykłady (Rys. 7.7-7.10) pokazują przebieg epidemii na nowo zakażonych węzłach, co może być odpowiednikiem sprzedaży singla w kolejnych tygodniach od premiery. Prezentowane krzywe dla poszczęólnych wartości  $m$  powstały poprzez uśrednienie cyklu życia dla 100 niezależnych wirusów. Podobnie jak w przypadku  $m = 2$ , dla innych parametrów  $m$  zasięgi epidemii jest szerszy w przypadku jej startu z węzła centralnego (Rys. 7.7) niż peryferyjnego (Rys. 7,8). Zjawisko to zachodzi także dla większych agresywności wirusa  $a$  (Rys. 7.9), gdyż liczba zarażonych węzłów w przypadku startu z huba jest przynajmniej o rząd wielkości większa, niż w przypadku startu epidemii z peryferii.



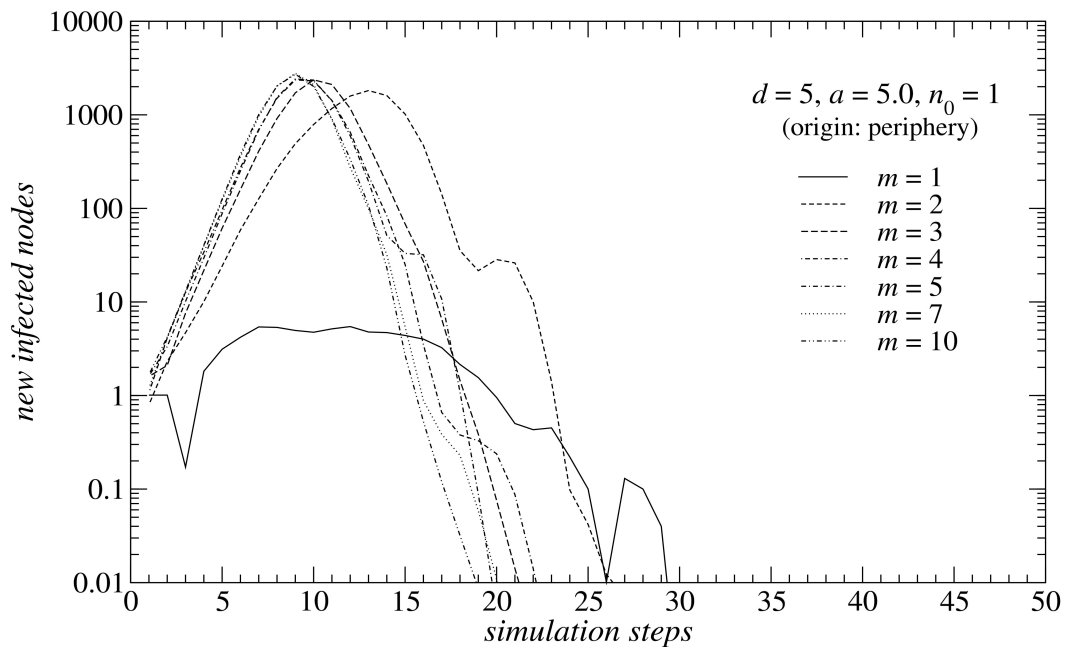
**Rys. 7.7** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach (*simulation steps*) epidemii wychodzącej z huba dla  $d = 1$  przy agresywności  $a = 5$  w sieciach Barabasięgo-Albert o różnych parametrach  $m$ .



**Rys. 7.8** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach (*simulation steps*) epidemii wychodzącej z periferii dla  $d = 5$  przy agresywności  $a = 1$  (u góry) oraz wychodzącej z węzła stołecznego (na dole) dla  $d = 1, a = 5$  w sieciach Barabasięgo-Albert o różnych parametrach  $m$ .

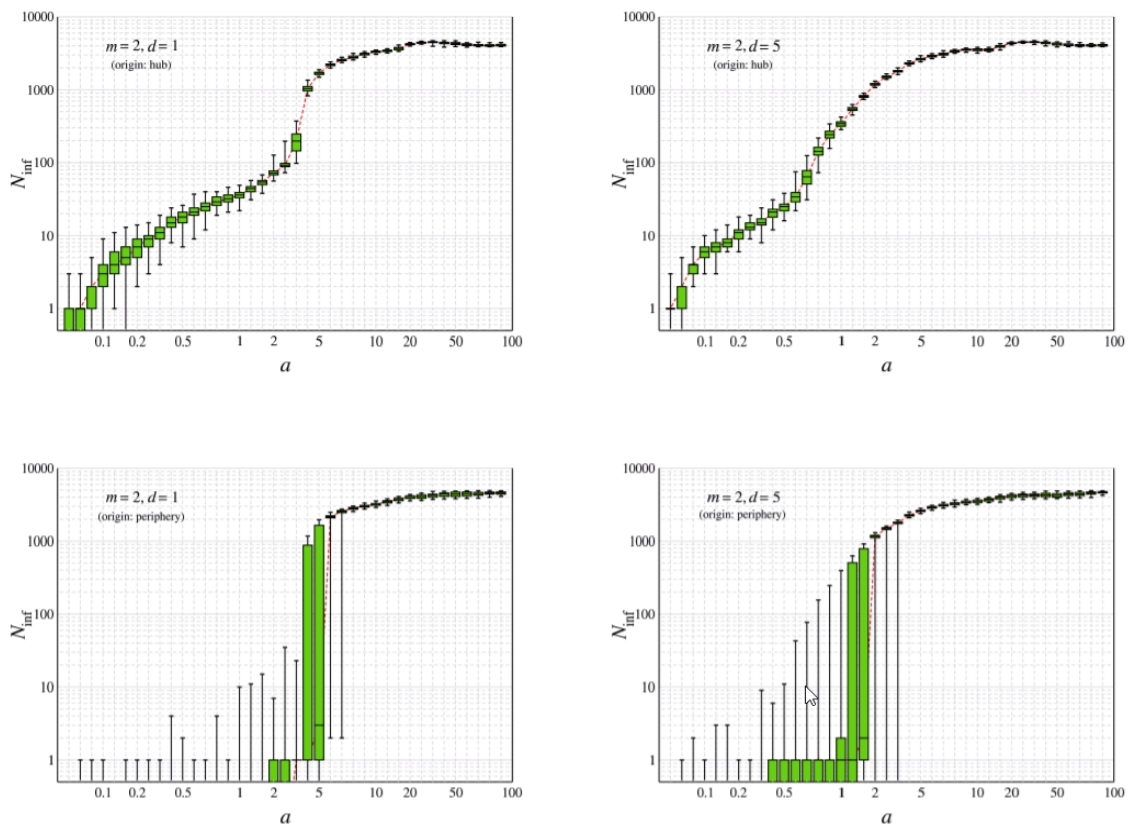


**.Rys. 7.9** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach (*simulation steps*) epidemii wychodzącej z periferii dla  $d = 5$  przy agresywności  $a = 1$  w sieciach Barabasięgo-Albert o różnych parametrach  $m$ .



**.Rys. 7.10** Nowo zarażone węzły (*newly infected nodes*) w kolejnych krokach (*simulation steps*) epidemii wychodzącej z periferii dla  $d = 5$  przy agresywności  $a = 5$  (na dole) w sieciach Barabasięgo-Albert o różnych parametrach  $m$

Warto odnotować także zjawisko charakterystyczne dla perkolacji, czyli fakt, że epidemia będzie trwała najdłużej jeśli sieć będzie miała parametr przyłączenia  $m = 2$ , gdyż dla  $m > 2$  i ściśle połączonej sieci mimo podobieństw, węzły zarażą się za szybko i cykl życia wirusa równie szybko zgaśnie. Oznacza to, że wartość  $m = 2$  wydaje się optymalna, co prowadzi do wniosku, że jeżeli przyjmiemy że odzwierciedla ona rzeczywistą sieć, to każda ingerencja w strukturę rynku polegająca np. na stworzeniu dodatkowych połączeń między węzłami jako uczestnikami rynku, mogłaby spowodować szybkie nasycenie zainteresowania muzyką, które nie prowadziłyby do realistycznych krzywych sprzedaży (podobnie jak wcześniej omówione zwiększenie agresywności wirusa  $a$  czy też czasu transakcyjnego  $d$  przebywania węzła w stanie zainfekowania). Z drugiej strony redukcja sieci do  $m = 1$  mogłaby spowodować brak epidemii na rynku.



**Rys. 7.11** Porównanie wariacji maksymalnej liczby nowo zarażonych węzłów  $N_{inf}(\max)$  dla czasu transakcyjnego  $d$  równego 1 oraz 5 jako funkcja agresywności wirusa wychodzącego ze centrum (u góry) oraz z peryferii (na dole). Zaznaczone na wykresie boxploty oznaczają różnicę między kwartylami Q1 oraz Q3, a ich środek oznacza średnią Q2 dla  $N_{inf}(\max)$ , w skali logarytmicznej.

Jak wspominałem wcześniej, sukces epidemii w pierwszej kolejności zależy od tego, czy pierwszy zarażony węzeł zainfekuje swojego sąsiada – czyli od agresywności wirusa  $a$ . Ale czy agresywność  $a$  ma duży wpływ na zarażanie dalszych węzłów, niż sąsiedni? Aby

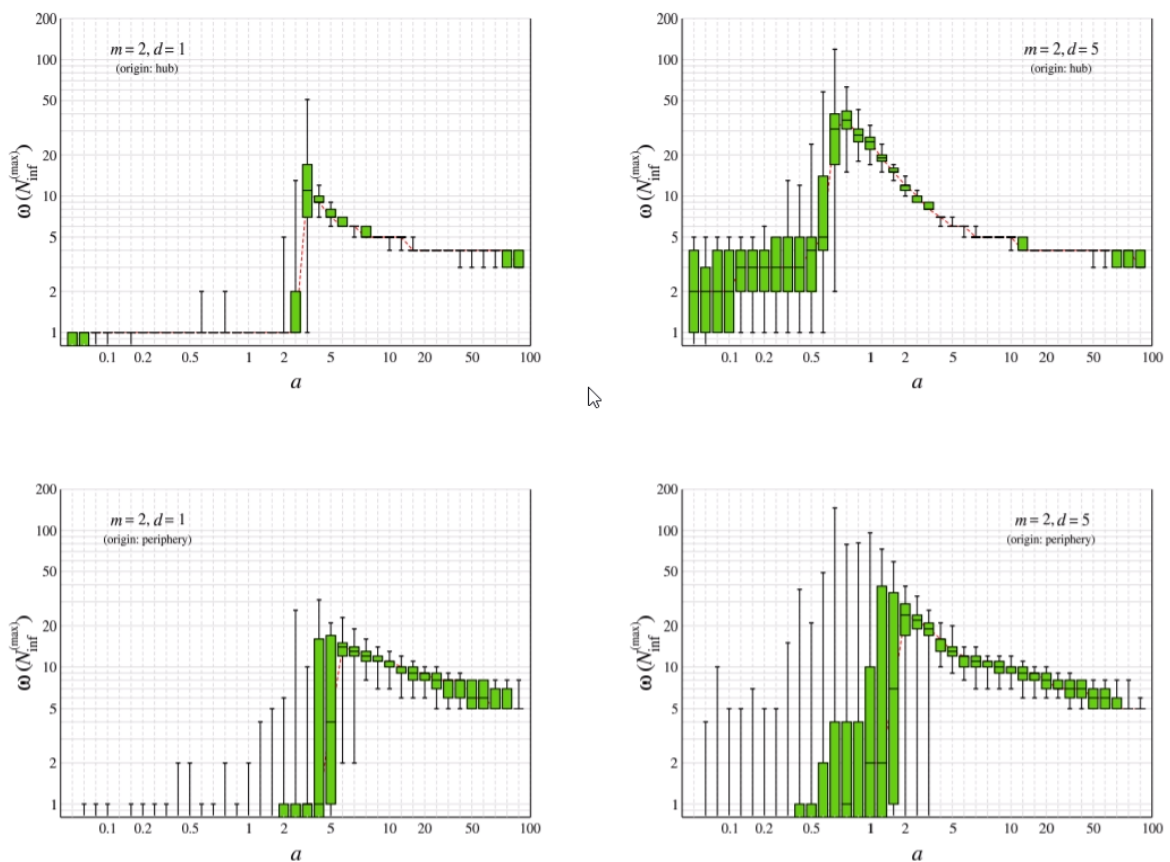
odpowiedzieć na to pytanie, warto zbadać cykle życia produktu pod kątem zależności maksimum liczby nowo zarażonych węzłów  $N_{inf}$  od agresywności wirusa  $a$ . Innymi słowy, chodzi o zbadanie jaki wpływ agresywność wirusa  $a$  miałyby na najwyższe miejsce singla na listach sprzedaży, rozumianej tu jako maksimum krzywej cyklu życia produktu. Ma to znaczenie, ponieważ w poprzednim rozdziale opisując najbardziej prawdopodobne ścieżki infekcji rozważaliśmy czas opóźnień między maksimami cykli w różnych krajach.

We wszystkich czterech prezentowanych wariantach (Rys. 7.11) mniej lub bardziej wyraźnie występuje krytyczny próg agresywności  $a$  powyżej którego zachodzi przełamanie, co wynika z faktu, że rozmiar sieci jest skończony. Oznacza to, że jeśli epidemia ma początek w hubie, to obserwujemy stopniowy, monotoniczny wzrost maksymalnej liczby nowo zarażonych węzłów aż osiągnięta zostanie saturacja, gdy dalsze zwiększanie agresywności wirusa nie powoduje istotnych zmian liczby zainfekowanych węzłów (co wynika oczywiście ze struktury połączeń międzywęzłowych w sieci i jej skończonego rozmiaru). Dla małych czasów  $d$  można zauważyć pewien skokowy wzrost maksymalnej liczby zainfekowanych węzłów przy pewnej wartości  $a$  (Rys. 7.11, lewy górny wykres), jednak nie wykazuje on cech perkolacji, a im większe  $a$ , tym mniejszy skok aż do jego zaniku i wygładzenia krzywej (Rys. 7.11, prawy górny wykres). Inaczej wygląda sytuacja dla epidemii zaczynających się w węzle peryferyjnym, gdy niezależnie od  $d$  nie obserwujemy stopniowego wzrostu, ale gwałtowny skok maksymalnej liczby nowo zainfekowanych węzłów przy pewnej progowej wartości agresywności  $a$ , a więc jest to przejście z sytuacji braku pandemii do sytuacji, gdy ona obejmuje znaczną część sieci (dolne wykresy na Rys. 7.11). Co ciekawe, w tym zakresie wartości parametru  $a$  obserwujemy najbogatszą dynamikę, gdy przy tym samym  $a$  wirus może wkrótce całkiem zaniknąć, stworzyć powoli rozwijającą się epidemię, jak i zarazić w kilku krokach całą sieć. Można ten obszar wartości  $a$  określić jako “krytyczny”.

Interesujące jest również porównanie wariancji czasu dojścia  $\omega$  do najwyższego miejsca na liście sprzedaży (rozumianego tu jako maksimum aktualnie nowo zarażonych węzłów, czyli maksimum na krzywej sprzedażowej). Przede wszystkim dlatego, że krytyczność w zależności od  $a$  jest tu o wiele bardziej widoczna i charakterystyczna dla pandemii, która szybko osiąga szczyt w sytuacji mało agresywnych wirusów, lub – z drugiej strony – dla ekstremalnie agresywnych. Natomiast w każdej w przedstawionych sytuacji istnieje próg agresywności  $a$ , dla którego wariancja czasu dojścia do szczytu  $\omega$  jest maksymalna.



Przekłada się na duże zróżnicowanie kształtów krzywej cyklu życia produktu i scenariuszy zdobywania popularności. Wydają się one bogate i najmniej przewidywalne w sytuacji startu z peryferyjnego węzła i przy długim czasie transakcyjnym  $d$ .



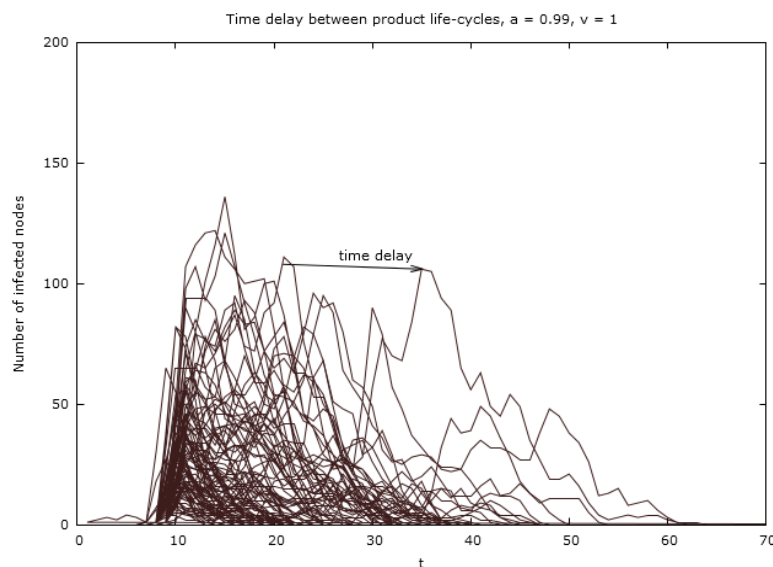
**Rys. 7.12** Porównanie wariancji czasu dojścia  $\omega$  do maksimum krzywej cyklu życia produktu dla czasu transakcyjnego równego 1 oraz 5 jako funkcja agresywności wirusa wychodzącego z centrum (na górze) i peryferii (na dole). Boxploty zdefiniowane są analogicznie jak na Rys. 7.11.

Wynik ten wskazuje, że zachowanie krytyczne zależy od pewnego parametru kontrolnego, gdzie poniżej pewnego progu dynamika subkrytyczna jest lokalna, a zdarzenia destrukcyjne o zasięgu globalnym nie są możliwe. Powyżej tego progu dynamika nadkrytyczna pozwala tylko na zdarzenia globalne, podczas gdy lokalne nie są obserwowane. W pobliżu progu krytycznego możliwe są jednak oba rodzaje zdarzeń. W modelu rolę takiego parametru kontrolnego pełni agresywność wirusa. Pozornie, wynik ten wydaje się być sprzeczny ze znaną właściwością sieci o architekturze bezskalowej, gdzie nie istnieje próg epidemii i wirus może rozpocząć epidemię nawet wtedy, gdy jego agresywność jest niska. Jednakże, odnosi się to do sieci, w których wagi krawędzi są tym, co determinuje prawdopodobieństwo infekcji dla sąsiednich węzłów. Zamiast tego, w modelu, w którym każdy węzeł ma swoją wewnętrzną odporność na infekcję odwrotnie

proporcjonalną do jego krotności, węzeł centralny ma najmniejsze prawdopodobieństwo bycia zarażonym i pełni rolę ekranu dla transmisji wirusa przynajmniej we wczesnym stadium epidemii, kiedy mało węzłów jest zarażonych. Ta własność sprawia, że - w przeciwieństwie do typowych modeli bezskalowych sieci z wagami – próg agresywności  $a$ , powyżej którego zachodzi epidemia, jest jak najbardziej możliwy.

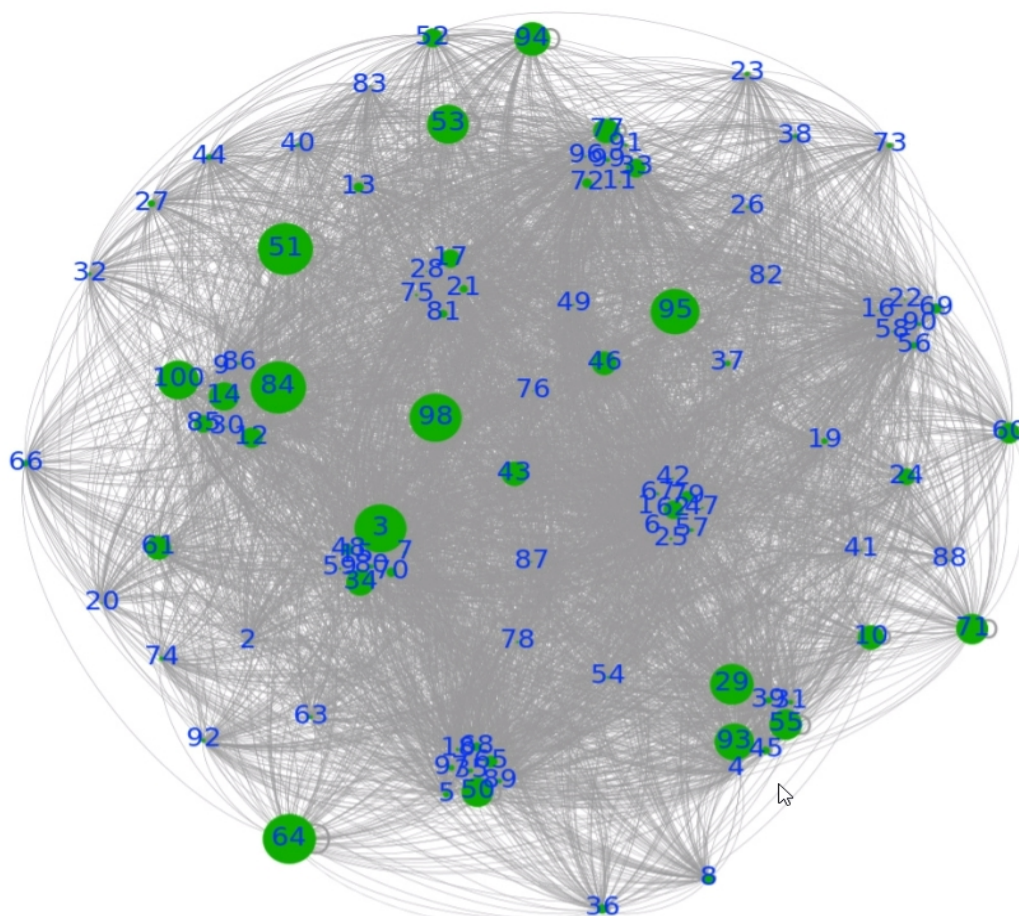
### 7.3 Przypadek wielu rynków

Po omówieniu w poprzednim podrozdziale ewolucji epidemii na pojedynczym rynku, warto wrócić do kompletnego modelu świata składającego się z  $M = 100$  rynków połączonych pojedynczymi krawędziami. Konstrukcyjnie, wszystkie rynki z wyjątkiem rynku pochodzenia wirusa mogą być zainfekowane wyłącznie przez centralny węzeł. Oznacza to, że wzorzec rozprzestrzeniania się infekcji na pierwotnym rynku  $M_{i0}$  może się różnić od odpowiadającego mu wzorca na każdym kolejnym rynku. Dotyczy to zwłaszcza peryferyjnych węzłów początkowych. Oznacza to, że dynamika infekcji na macierzystym rynku może być zazwyczaj bogatsza niż jej odpowiedniki na innych rynkach. Jest to wynik realistyczny, ponieważ wielkie koncerty rozważając eksport muzyki z rynku pochodzenia artysty do innych krajów, w pierwszej kolejności zadają pytanie: Jaka była sprzedaż jego płyt na rynku krajowym? Dlatego model przewiduje, że rozpowszechnianie muzyki za granicą jest bardziej prawdopodobne, jeśli jej komercyjny potencjał został już zweryfikowany na rynku macierzystym, niezależnie od kraju pochodzenia utworu.



**Rys.7.13** Przykładowy przebieg epidemii jako zbiór krzywych sprzedaży na wielu rynkach zdefiniowany jako zależność liczby zarażonych węzłów (*number of infected nodes*) od czasu  $t$ , dla modelu agentowego z zaznaczeniem opóźnienia między szczytami krzywych dla  $a = 0,99$  i  $d = 1$ .

W przeciwieństwie do wielu modeli sieci wzajemnie połączonych, skonstruowany model nie pozwala na żadne sprzężenia zwrotne między rynkami, które mogłyby potencjalnie zmienić rozprzestrzenianie się epidemii na rynku źródłowym poprzez wzajemne transmisje wirusów z innych rynków [97-98]. Dlatego dynamika międzyrynkowego przenoszenia wirusa (za pośrednictwem węzłów centralnych) może być uważana za niezależną od jego rozprzestrzeniania się na kolejnych rynkach lokalnych, co pozwala na oddzielne analizowanie tych dwóch poziomów sieci.

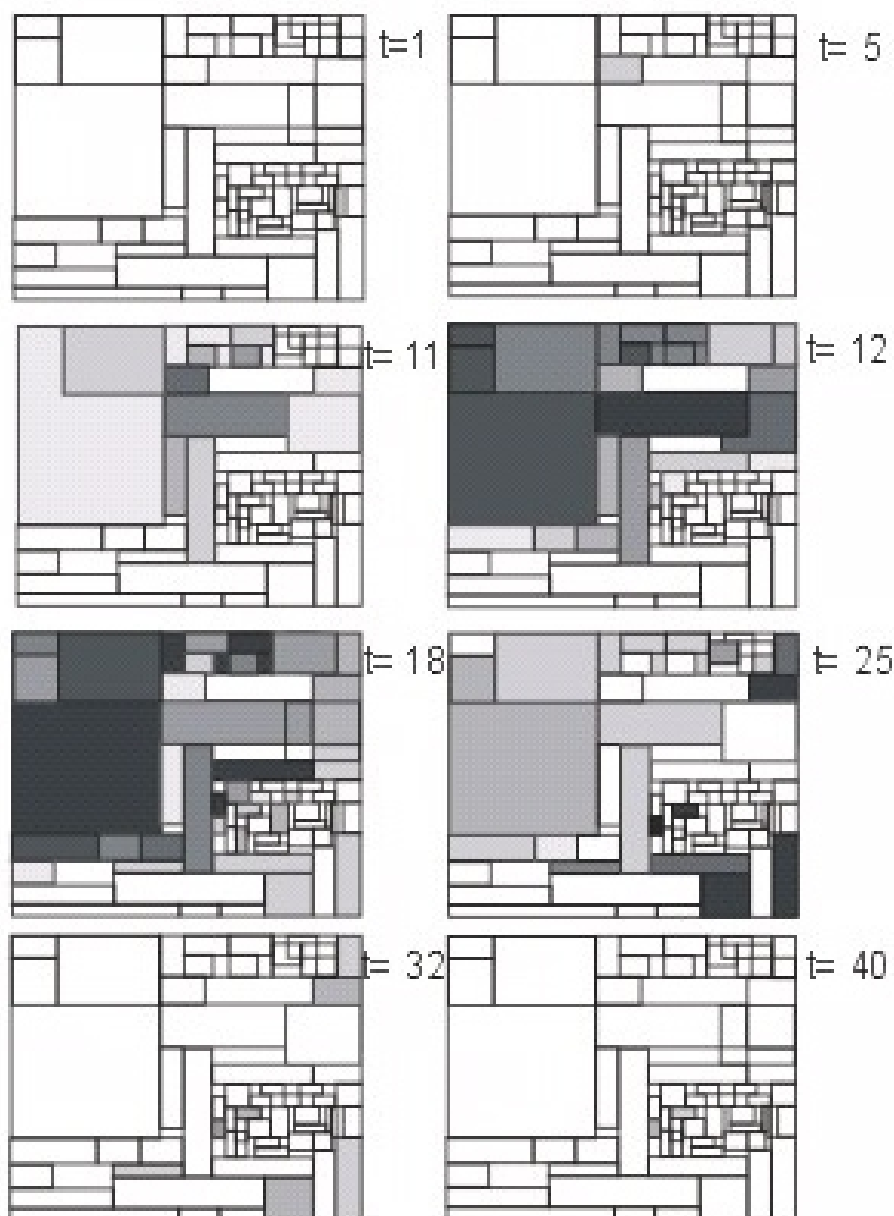


**Rys. 7.14** Globalny rynek złożony ze 100 rynków lokalnych zawierający kraje należące do klastrów kulturowych jak i będące poza klastrami. Wielkość symbolu reprezentującego rynek odpowiada liczbie węzłów tworzących podsieć odpowiadającą temu rynkowi, a węzły położone blisko siebie oznaczają rynki tworzące klastry kulturowe.

Pierwszym krokiem jest wygenerowanie globalnego rynku jako "świata" zgodnie z procedurą opisaną w podrozdziale 7.1. Choć prezentuję tu jedynie jego przykładową realizację, to jednak oddaje ona wszystkie interesujące cechy dynamiczne, jakie można zaobserwować w modelu. W takiej realizacji każdy węzeł  $M_i$  reprezentuje podsieć rynku lokalnego o liczbie węzłów  $N_i$ . Na początku przeprowadziłem symulacje modelu zdefiniowanego jak w podrozdziale 7.1. Wagi  $w_{ij}$  zostały wylosowane z rozkładu jednorodnego na przedziale  $[0,1]$ . Tak określone założenia pozwalają jedynie na przebieg

epidemii albo zlokalizowany jedynie na początkowym rynku  $M_{i0}$  albo na wszystkich rynkach (Rys. 7.13), lub brak jakichkolwiek zarażeń, przy czym relacje między liczbą tych pierwszych i tych drugich przypadków silnie zależy od agresywności  $a$ . Tak zdefiniowany model, mimo zastosowania go na szerokiej statystycznej próbie realizacji, nie generuje stanów pośrednich np. sytuacji zainfekowania tylko kilku wybranych rynków. Jest to ważne, bo nawet jeśli zjawisko zdobywania popularności w czasach Internetu po 2003 roku rzeczywiście ma już charakter globalny (czyli utwór – jako pokazuje rozdział poprzedni - może zdobyć popularność we wszystkich krajach, lub na rynku macierzystym, lub nie zdobyć jej w ogóle), to w czasach dominacji nośników fizycznych w XX wieku w pełni globalną popularność zdobywało jedynie kilkanaście utworów na dekadę, a dominującym scenariuszem było zdobywanie popularności jedynie w kilku krajach, najczęściej bliskich kulturowo. Dlatego aby odzwierciedlić to rynkowe zjawisko i nadać modelowi więcej, niż tylko 3 możliwe scenariusze zdobywania popularności [99], należy wzmocnić wagi połączeń między rynkami bliskimi kulturowo. Klastry kulturowe w modelu zostają utworzone w ten sposób, że wartości wag międzyrynkowych  $w_{ij}$  nie mogą należeć do przedziału  $[w_{thresh} - \Delta w, w_{thresh} + \Delta w]$ , co wyraźnie separuje zakres dopuszczalnych wartości dla wag między podsieciami nienależącymi do tego samego klastra kulturowego ( $w_{ij} < w_{thresh} - \Delta w$ ) od wartości wag między podsieciami ( $w_{ij} > w_{thresh} + \Delta w$ ) wewnątrz tego samego klastra. Przyjęto, że  $w_{thresh} = 0,5$  oraz  $\Delta w = 0,4$ , co daje  $w_{ij} < 0,1$  dla połączeń poza klastrami i  $w_{ij} > 0,9$  dla połączeń wewnątrz klastrów. (Wartości tych parametrów można oczywiście zmieniać w modelu, manipulując w ten sposób siłą wpływu podobieństwa kulturowego na dynamikę epidemii.) W ten sposób otrzymano heterogeniczną strukturę składającą się z silnie i słabo sprzężonych podsieci [101] oraz dwa typy procesów epidemicznych: szybkie (dyfuzja wirusa wewnątrz rynku i pomiędzy rynkami w obrębie klastra) i wolne (dyfuzja wirusa pomiędzy rynkami odległymi kulturowo) [102].

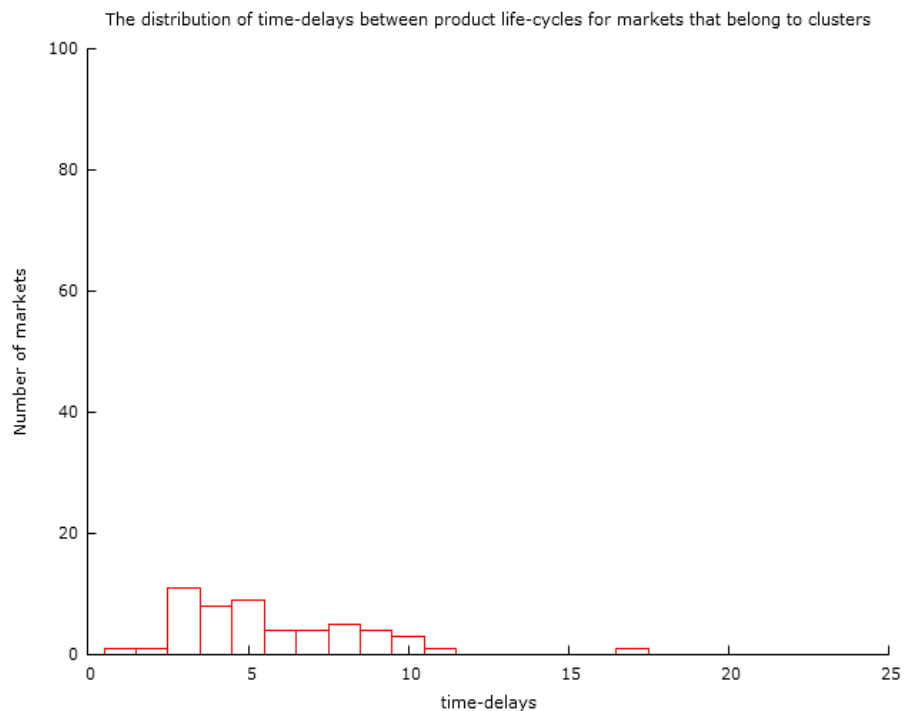
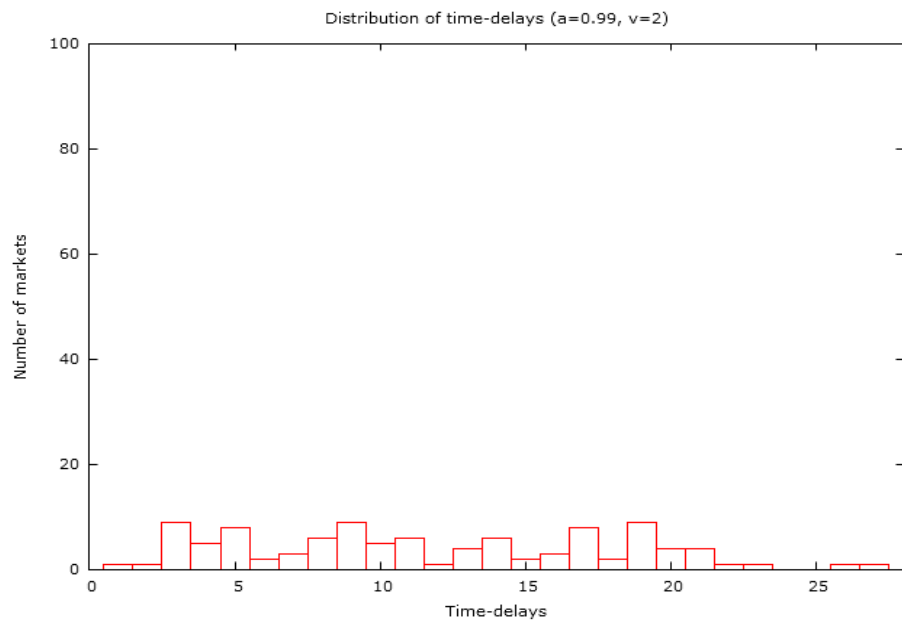
Efektym podziału na klastry jest odpowiadająca rzeczywistości XX wieku sytuacja, że utwór, którego popularność wydostanie się za granicę, zaraża już niekoniecznie wszystkie rynki (jak ma to często miejsce w XXI wieku), ale jedynie ich część. Ma to oczywiście wpływ na zdefiniowane w szóstym rozdziale wzorem (6.1) opóźnienia między szczytami popularności piosenki na różnych rynkach (pokazane także na Rys. 7.13). Przybierają one wówczas inny rozkład, bowiem podział na klastry ogranicza możliwość zarażania innych rynków nawet drogą okrężną [103].



**Rys. 7.15** Odpowiadająca Rys. 4.3 przykładowa mapa rozprzestrzeniania się popularności utworu w kolejnych krokach. Wielkość prostokąta odpowiada liczbie węzłów podsięci. Zaciemnienie odpowiada procentowi zarażonych węzłów.

Przykładowo, dla parametrów  $a = 0,99$  i czasu transakcyjnego  $d = 2$  porównując przykładowe realizacje ze 100 przeprowadzonych na rynkach bez klastrów i na rynkach z klastrami (Rys. 7.16), łatwo zauważyć, że gdy nie ma klastrów to rozkład między szczytami popularności (czyli rozkład liczby kroków po których liczba nowo zarażonych węzłów jest maksymalna) jest bardziej równomierny, natomiast wprowadzenie klastrów powoduje ograniczenie większości opóźnień do kilkunastu kroków czasowych. Intuicyjna część tego wyniku dotyczy rynków wewnątrz klastrów, gdzie epidemia ma szansę na szybkie rozprzestrzenianie się dzięki dużym wagom połączeń. W przypadku rynków

pozostających poza klastrami wynik ten może się jednak wydawać mniej intuicyjny, gdyż wagi połączeń takich rynków pozostają niewielkie, jednak staje się w pełni zrozumiałe, gdy weźmie się pod uwagę, że każdy z zainfekowanych rynków tworzących klastery ma swoje połączenie z każdym z rynków poza klastrami. Efektywnie więc rynki w ramach klastra tworzą razem jeden większy rynek, którego siła oddziaływania na inne rynki jest większa niż każdego z rynków składowych. Wprowadzenie klastrów jest więc równoważne z utworzeniem kilku ponadnormatywnie dużych centrów zakażenia.



Rys. 7.16 Rozkład opóźnień (*time delays*) między szczytami popularności na lokalnych rynkach bez grupowania ich w

klastry (u góry) oraz przy uwzględnieniu klastrów (na dole). Przykładowe realizacje są wykonane w obu wariantach dla  $a = 0,99$  i  $d = 2$

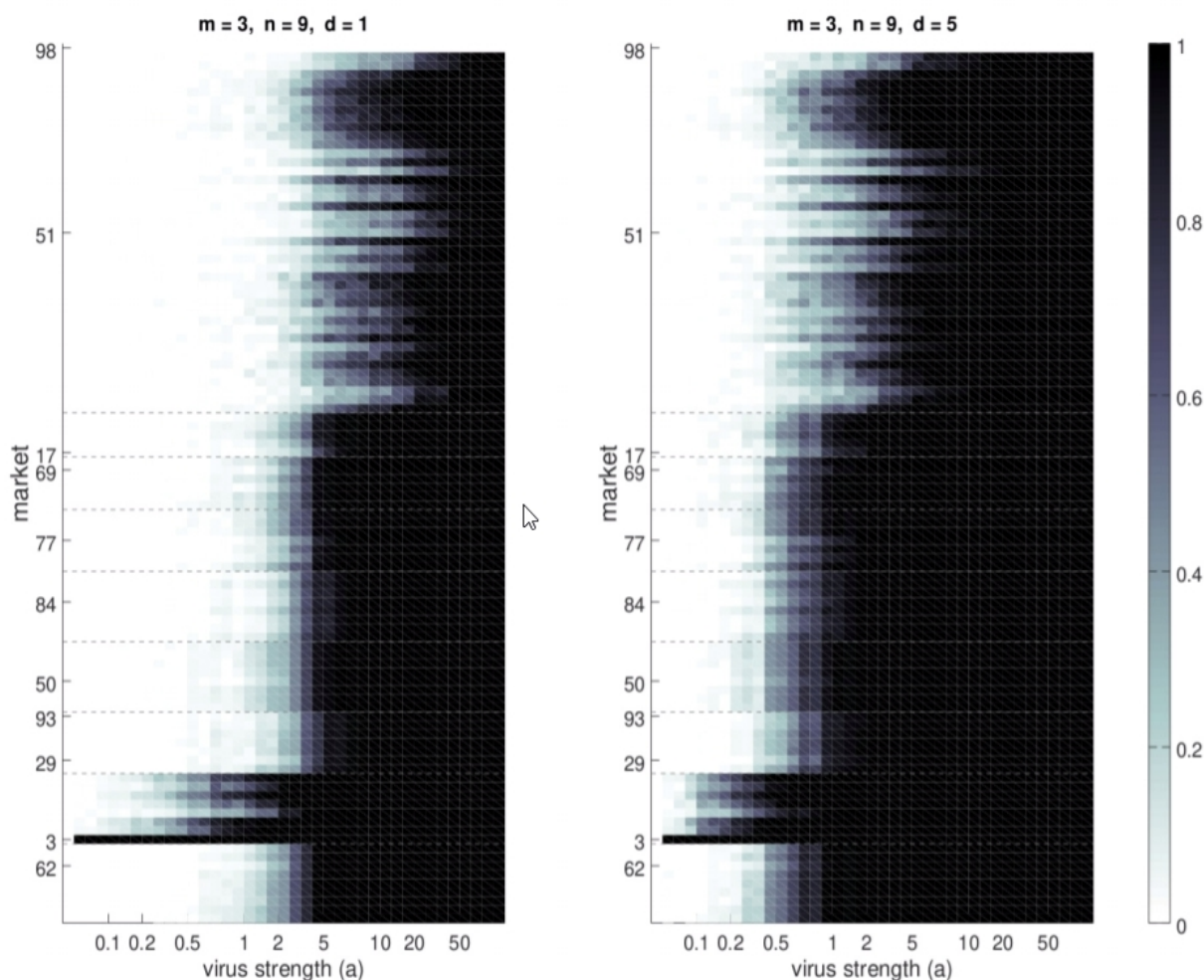
Zgodnie ze wzorami (7.1) i (7.2), oraz założeniem, że rynek globalny powinien uwzględniać bliskość kulturową krajów w postaci klastrów, przedstawioną wcześniej przykładową realizację można zastosować dla sieci 100 rynków lokalnych, w której największy rynek  $M_{51}$  ma rozmiar 9617 węzłów, a najmniejszy  $M_{20}$  jedynie 103 (Rys. 7.14). Globalna sieć zawiera 8 klastrów (złożonych z 5 do 9 bliskich kulturowo rynków), a także 42 kraje leżące poza nimi.

Po przeprowadzeniu symulacji dla tak zdefiniowanej sieci i różnych wyborach rynku początkowego, najciekawsze okazały się rezultaty dla trzech charakterystycznych przykładów wyboru węzła początkowego ścieżki zakażeń:

- 1) węzeł należy do dużego rynku  $M_{51}$  spoza klastrów
- 2) węzeł należy do dużego rynku  $M_3$  wewnątrz klastra
- 3) węzeł należy do małego rynku  $M_{76}$  spoza klastrów

Każdy z przypadków został zrealizowany 100 razy, celem zbadania jego własności statystycznych. Pozostałe scenariusze zdobywania popularności na rynku globalnym mieszczą się pomiędzy rezultatami uzyskanymi dla tych trzech charakterystycznych przypadków. Zdobywanie popularności jest najłatwiejsze, jeśli węzeł początkowy ścieżki pochodzi z dużego rynku należącego do jednego z klastrów (Rys. 7.15). Wówczas nawet niewielka agresywność pozwala wirusowi wydostać się na rynki zewnętrzne, przy czym najczęściej są to rynki z tego samego klastra.

Jeśli start epidemii następuje z rynku nienależącego do żadnych klastrów (Rys. 7.16) to jedynym sposobem na zarażenie pozostałych rynków musi być duża atrakcyjność utworu, gdyż zdobywanie popularności poza rynkiem macierzystym następuje według podobnych krzywych sprzedażowych. W obu przypadkach (Rys. 7.15 – 7.16) wydłużenie czasu transakcyjnego pozwala uzyskać popularność utworom o odpowiednio niższej atrakcyjności, bowiem dla obu wariantów istnieje próg krytyczny parametru  $a$ , po którego przekroczeniu singiel może opanować rynki inne, niż macierzysty. Warto zauważyć, że efekt zarażenia wszystkich rynków zachodzi dopiero gdy  $a$  jest bardzo duże.



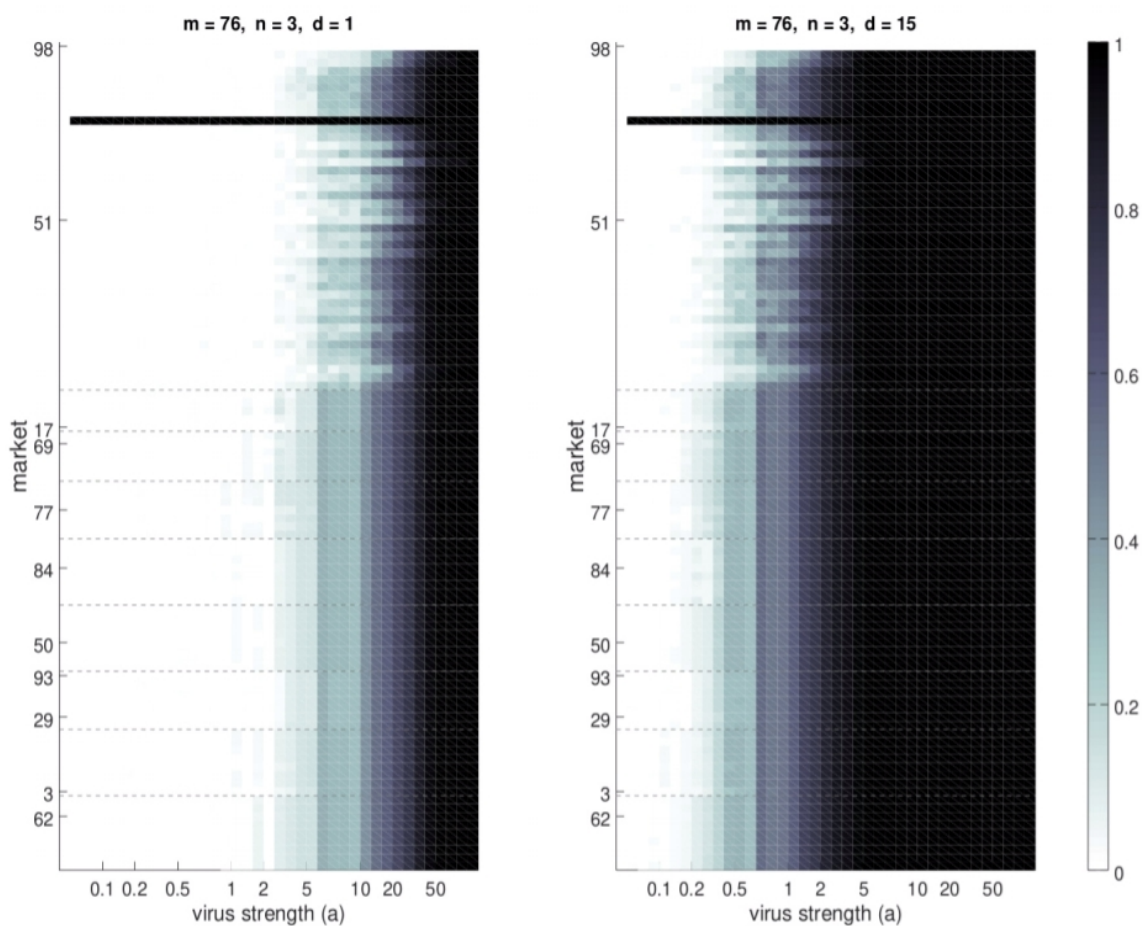
**Rys. 7.17** Przedstawiona w postaci zacięniowania (od 0 do 1) liczba realizacji zarażenia innych rynków przez rynek M3 (należący do jednego z klastrów) w zależności od agresywności wirusa  $a$ , dla  $d = 1$  (po lewej stronie) i  $d = 5$  (po prawej stronie). Parametr  $m$  określa numer rynku, parametr  $n$  wskazuje węzeł inicjujący. Przerywane linie poziome na diagramie oznaczają granice klastrów. Powyżej najwyżej położonej linii przerywanej zostały umieszczone 42 rynki pozaklastrowe.

Ścieżki zdobywania popularności dla piosenek o agresywności przecietnej (0,60), dużej (2,5 i 10) oraz ekstremalnej (90) zostały przedstawione na Rys. 7.21. Dla niewielkiej agresywności  $a = 0,6$  efekt popularności skuteczny jest jedynie na macierzystym rynku i to w śladowym zakresie - podobnie jak dla modelu agentowego bez klastrów (na Rys. 7.3).

Podobnie jak na Rys. 7.3, dopiero dla większej agresywności  $a$  popularność przenosi się na rynki zewnętrzne, ale wprowadzenie do modelu klastrów sprawia, że zamiast zarażać wszystkie 100 rynków, popularność piosenki jest widoczna jedynie w obrębie klastra, do którego należy pierwotny rynek. Warto przy tym zauważyć, że według modelu sukces na

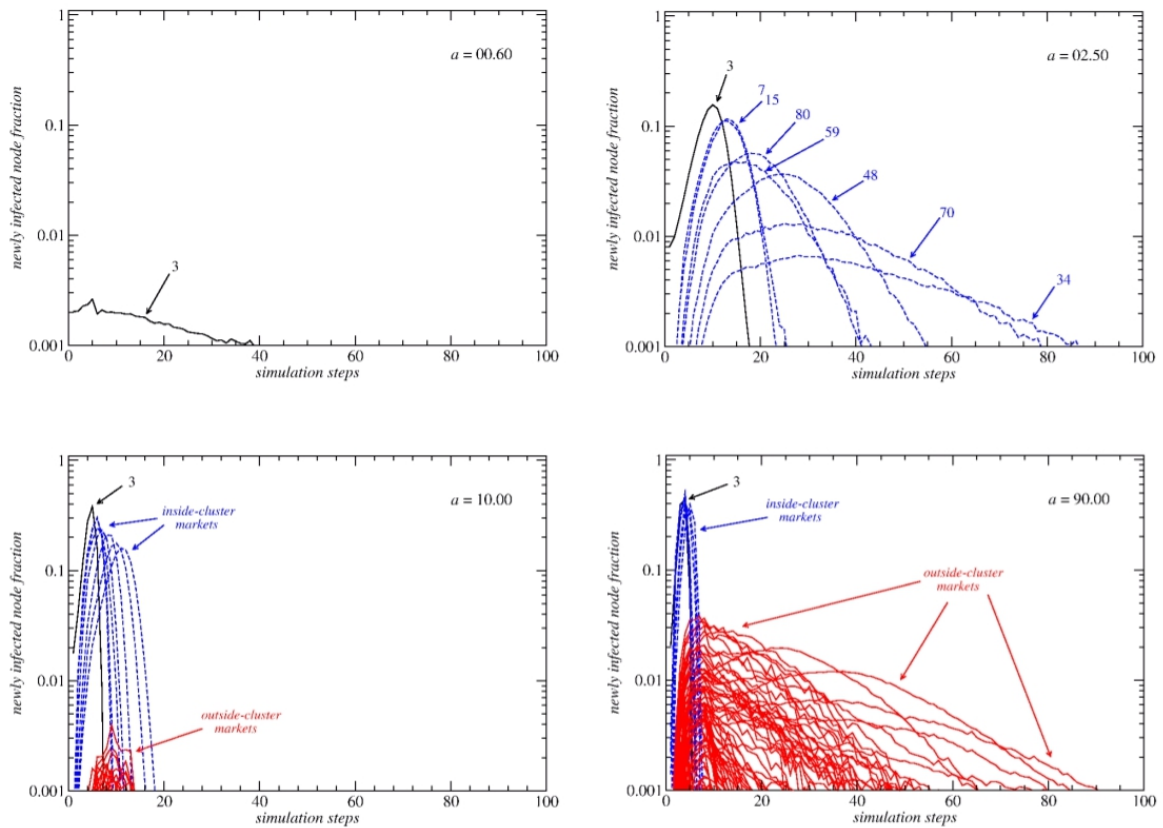


rynkach w klastrze jest powtórzeniem sukcesu na rynku macierzystym, co jest zgodne z rzeczywistymi decyzjami słuchaczy (i dostosowaną do nich polityką wielkich koncernów płytowych, które zanim zainwestują w eksport piosenek artysty na inne rynki, najpierw zadają kluczowe pytanie „Jaka jest popularność artysty na jego rynku macierzystym?”).



**Rys. 7.18** Identyczna zależność jak na Rys. 7.17, ale przedstawiona dla epidemii wychodzącej z małego rynku *M76* (leżący poza klastrami). Parametr  $m$  określa numer rynku, parametr  $n$  wskazuje węzeł inicjujący. Przerwane linie poziome na diagramie oznaczają granice klastrów. Powyżej najwyższej położonej linii przerywanej umieszczono 42 rynki pozaklastrowe

Co ciekawe, dopiero dla agresywności  $a = 10$  i większej, ścieżki cyklu życia produktu są coraz bardziej widoczne na rynkach spoza klastrów, a przy ekstremalnej wartości  $a$  piosenka może zarazić wszystkie 100 rynków (Rys. 7.19). Powyższe symulacje zostały wykonane dla czasu transakcyjnego  $d$  równego 2.

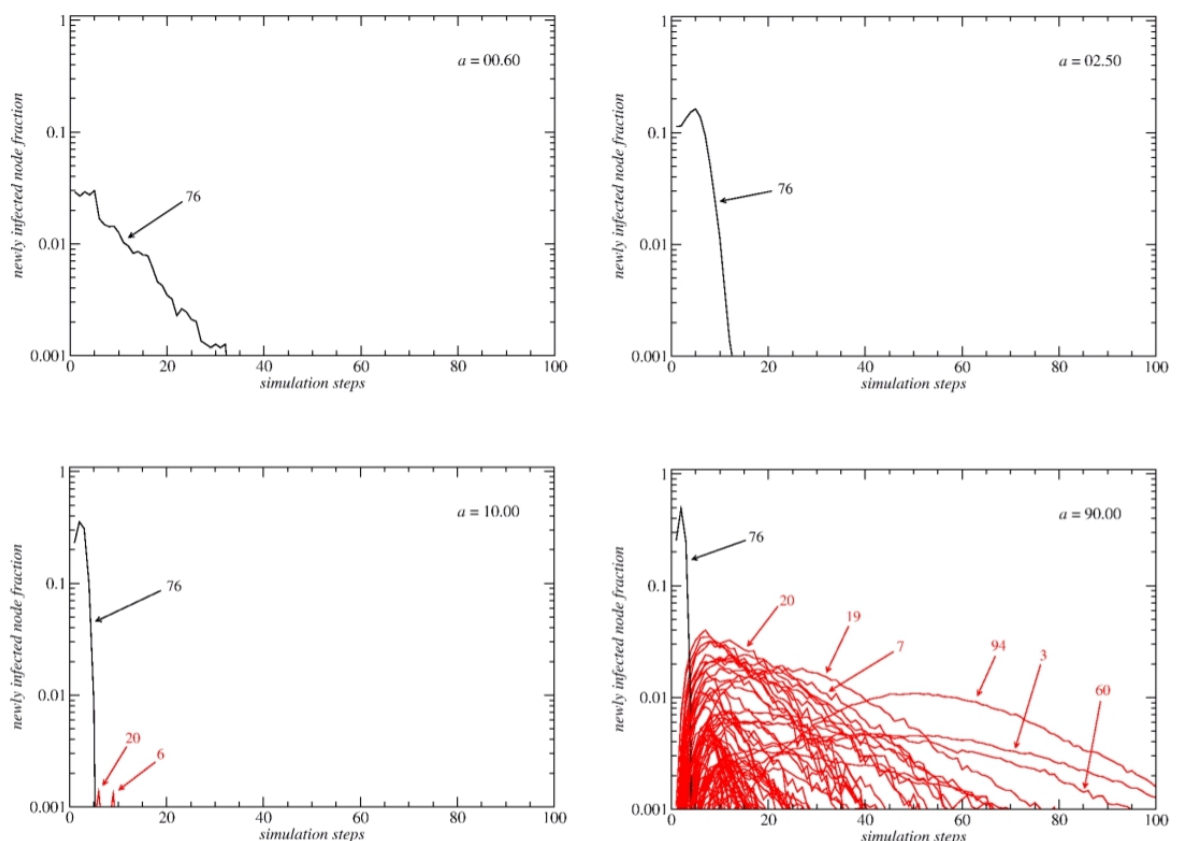


**Rys. 7.19** Przebiegi epidemii wychodzącej z centrum dużego rynku  $M3$  należącego do klastra, dla różnych agresywności  $a$  i czasu  $d=2$ . Na osi odciętych umieszczona jest liczba nowo zarażonych węzłów w stosunku do całości węzłów na danym rynku, w skali logarytmicznej. Linie ciągłe oznaczają rynek macierzysty. Ścieżki na rynkach pozaklastrowych zostały zaznaczone jaśniejszym kolorem.

Tymczasem jeśli piosenka pochodzić będzie z małego, niezależnego rynku spoza klastrów (Rys. 7.20), zdobywać popularność będzie jej o wiele trudniej i zarażenie innych rynków stanie się możliwe dopiero, gdy jej atrakcyjność będzie ekstremalna. Jednak wówczas przebiegi epidemii na rynkach zewnętrznych będą znacznie dłuższe, niż jej przebieg na rynku macierzystym. Rzeczywistym odzwierciedleniem tego jest choćby piosenka szwedzkiej grupy Europe 'The Final Countdown', która w maju 1986 na macierzystym rynku natychmiast awansowała z 8 miejsca na 1 (gdzie utrzymywała się przez 6 tygodni), po czym spadła od razu na miejsce 5, a w sierpniu opuściła szwedzką listę, a więc jej czas życia na liście w Szwecji, mimo ekstremalnej popularności, wynosił ledwie 14 tygodni. Dla porównania: na rynku brytyjskim ten sam singiel utrzymywał się przez prawie pół roku od października 1986 do lutego 1987, osiągając szczyt w grudniu 1986 i do wiosny 1987 zaraził wszystkie rynki światowe – w tym Stany Zjednoczone. Co prawda Szwecja wraz z Danią i Norwegią należą do klastra skandynawskiego, to jest on ekonomicznie

mało istotny jeśli chodzi o udział w wielkości globalnego rynku albumów (Rys. 4.3) i singli (Rys. 7.27) i mimo wylansowania kilku popularnych artystów, należy go traktować w skali globalnej jako rynek peryferyjny.

Dla zgromadzonych ścisłych danych z Europy 17000 singli w ostatnich 60 latach zdarzały się jedynie pojedyncze przypadki odpowiadające sytuacji jak na Rys. 7.20 – czyli opanowaniu wszystkich rynków przez utwór pochodzący pierwotnie spoza klastra. Przykładem jest tu singiel 'The Rhythm Of The Night' pochodzącej z Włoch grupy Corona, który w lutym 1994 roku osiągnął szczyt macierzystej listy przebojów na 8 tygodni, ale dopiero po 20 tygodniach za pośrednictwem Hiszpanii opanował większość rynków Europy, w tym - na jesieni - jako ostatnią Wielką Brytanię i Irlandię, docierając w tych krajach na miejsce 2. Do Stanów Zjednoczonych dotarł dopiero na początku 1995 roku, najwyżej na miejsce 11, a jego cykl życia trwał ponad rok.



**Rys. 7.20** Analogiczne jak na poprzednim rysunku przebiegi popularności utworu wychodzącego z centrum rynku  $M_{76}$  leżącego poza klastrem, dla różnych agresywności  $a$  i czasu transakcyjnego  $d=2$

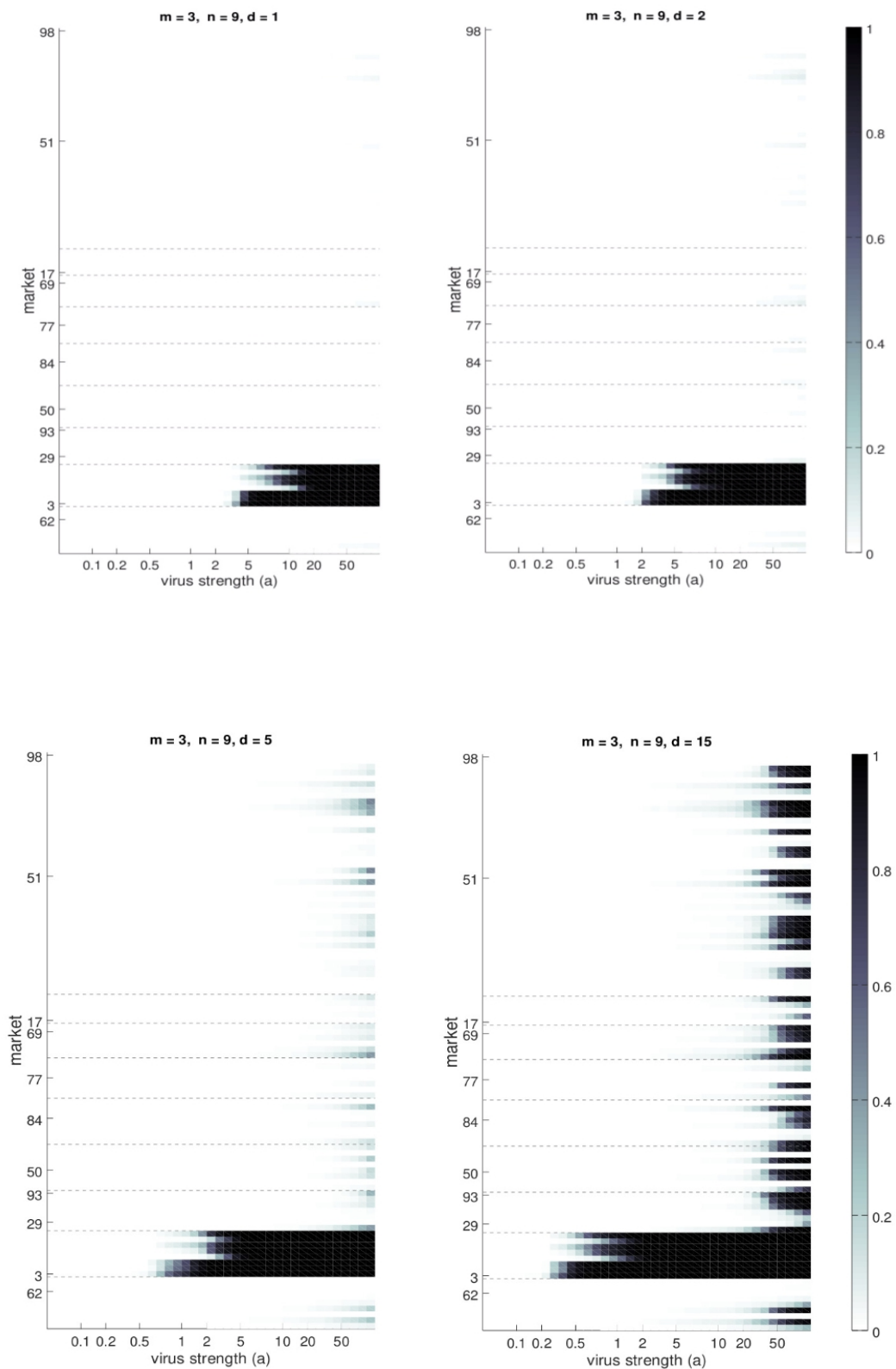
Ciekawym z punktu widzenia fizyki i entropii jest statystyczna możliwość istnienia

procesu, w którym popularność piosenki zaczyna się na rynku, który jest zwykle infekowany późno. Wówczas ostatnimi zarażonymi rynkami stają się te, które na ogół zarażają się jako pierwsze. Takie zjawisko, choć już w znacznie skróconej (do 7 tygodni) skali, nastąpiło w 2017 roku dla przeboju 'Despacito' Luisa Fonsi, który w erze Internetu opanował Europę zaczynając od Hiszpanii, a kończąc na Wielkiej Brytanii i Irlandii.

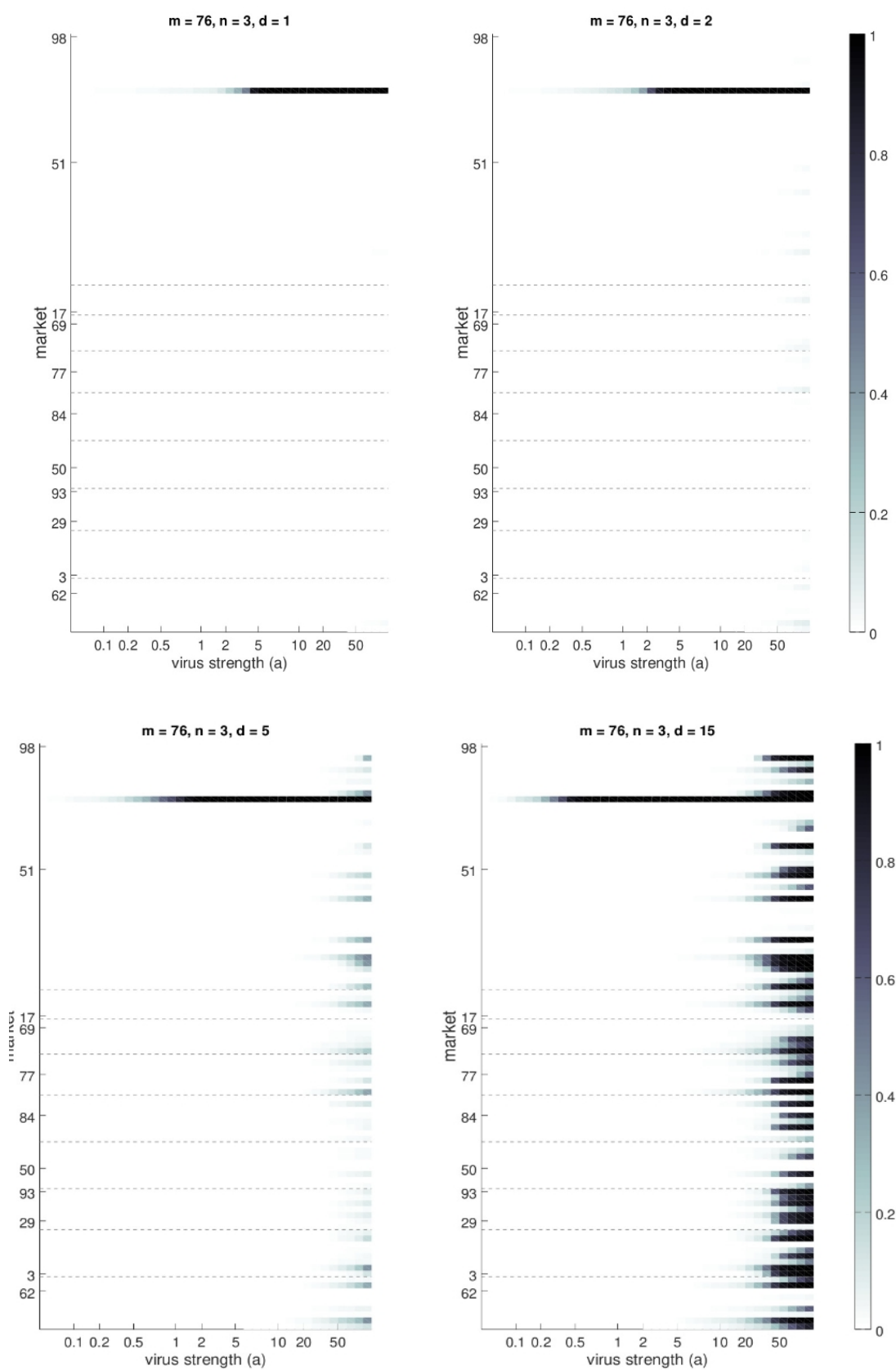
Aby wyprowadzić piosenkę na inne rynki oprócz jej atrakcyjności  $a$ , oraz miejsca startu epidemii, warto zbadać także zależność popularności od czasu pozostawania węzła w stanie zakażenia  $d$ . Okazuje się, że infekcje rozpoczęte z rynku należącego do klastrów jak i niezależnego rynku spoza klastrów przebiegają podobnie. Jednak różnicą jest fakt szybkiego zarażenia krajów należących do klastra, jeśli należy do niego kraj z którego pochodzi wirus. Jest to zgodne z rzeczywistością, bowiem praktycznie wszystkie przeboje z Wielkiej Brytanii zarażają Irlandię, piosenki z Niemiec są popularne w Austrii i Szwajcarii (klastry na które jest podzielona Europa przedstawia Rys. 6.5.). Natomiast dla zbyt krótkich czasów  $d$  globalna pandemia w ogóle nie występuje.

Na rys. 7.23 i 7.24 przedstawione zostały ścieżki transmisji wirusów między rynkami dla różnych przykładowych wartości parametru  $a$ , otrzymane każdorazowo na podstawie 100 niezależnych symulacji epidemii. Wagi połączeń między węzłami odpowiadającymi poszczególnym rynkom są proporcjonalne do ułamka udanych transmisji wirusa między tymi rynkami spośród wszystkich 100 symulacji. W przypadku infekcji zaczynających się na dużym rynku w ramach klastra dla względnie małej wartości  $a$  ścieżek jest niewiele i tworzą strukturę gwiazdzistą lub podwójnie gwiazdzistą (górne wykresy na Rys. 7.23). W tym drugim przypadku wewnętrzna gwiazda jest związana z infekcjami wewnątrz klastra, a zewnętrzna - poza nim. Dla dużych wartości  $a$  struktura ścieżek staje się już bardziej skomplikowana i obrazuje większe bogactwo możliwości (dolne wykresy na Rys. 7.23).

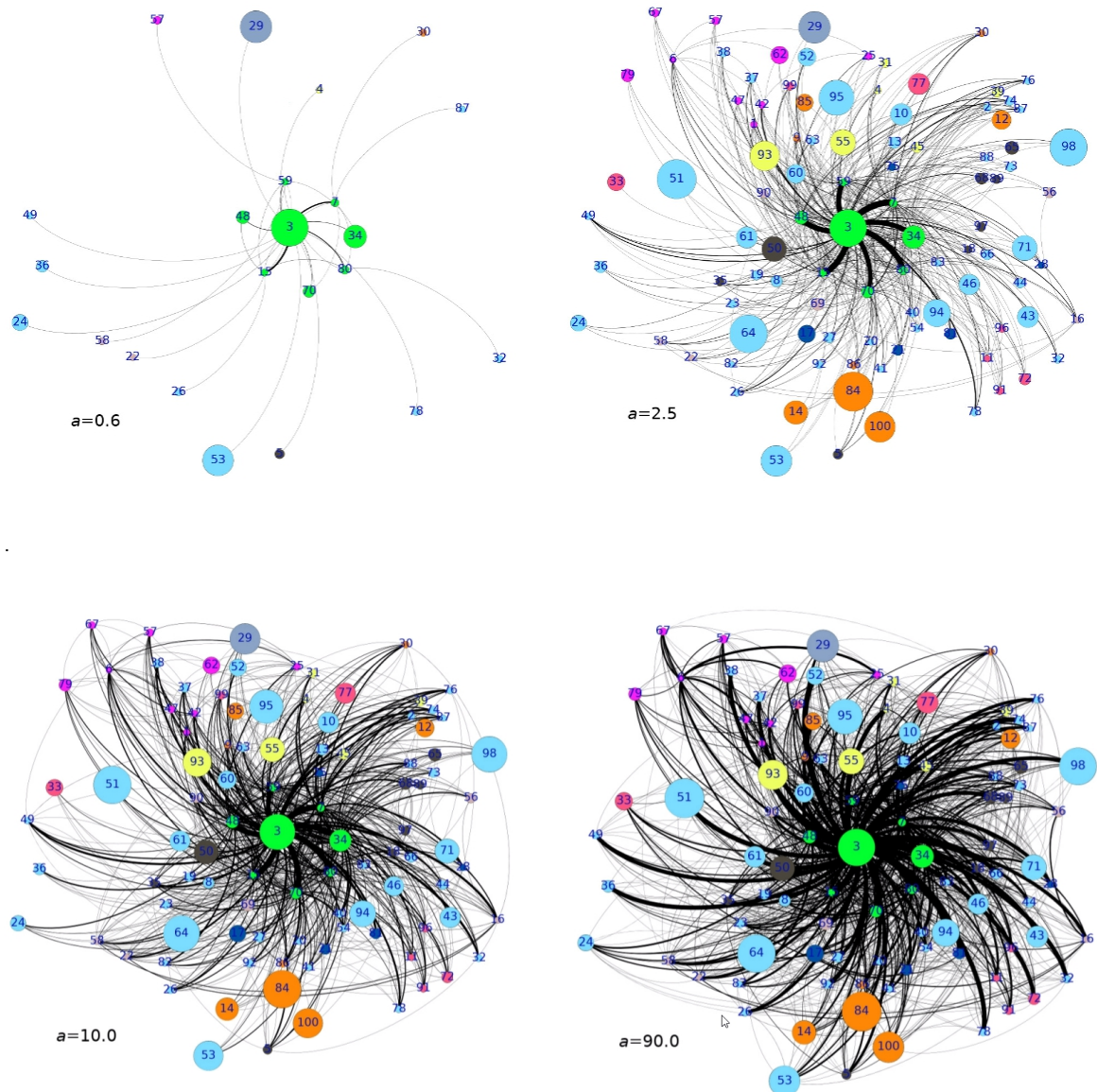
W przypadku infekcji rozpoczynającej się w peryferyjnym, niezależnym rynku sytuacja wygląda trochę inaczej, gdyż jedynie struktura pojedynczej gwiazdy jest wtedy widoczna (Rys. 7.24). Wraz ze wzrostem  $a$ , choć dla dużo większych wartości tego parametru niż w przypadku dużego rynku wewnątrz klastra, struktura ścieżek zaczyna obejmować całą sieć, podobnie jak na Rys. 7.23. Jest to naturalne, ponieważ przy odpowiednio dużej agresywności wirusa epidemia rozprzestrzenia się często na cały świat bez względu na konkretną strukturę globalnego rynku.



**Rys. 7.21** Zależność średniej liczby zainfekowanych węzłów w stosunku do całkowitej ilości zarażonych węzłów (pokazana w skali szarości) od agresywności wirusa dla różnych czasów transakcyjnych  $d$ : 1, 2, 5, 15 dla rynku M3 należącego do klastra 3. Przerwane linie poziome oznaczają granice klastrów.

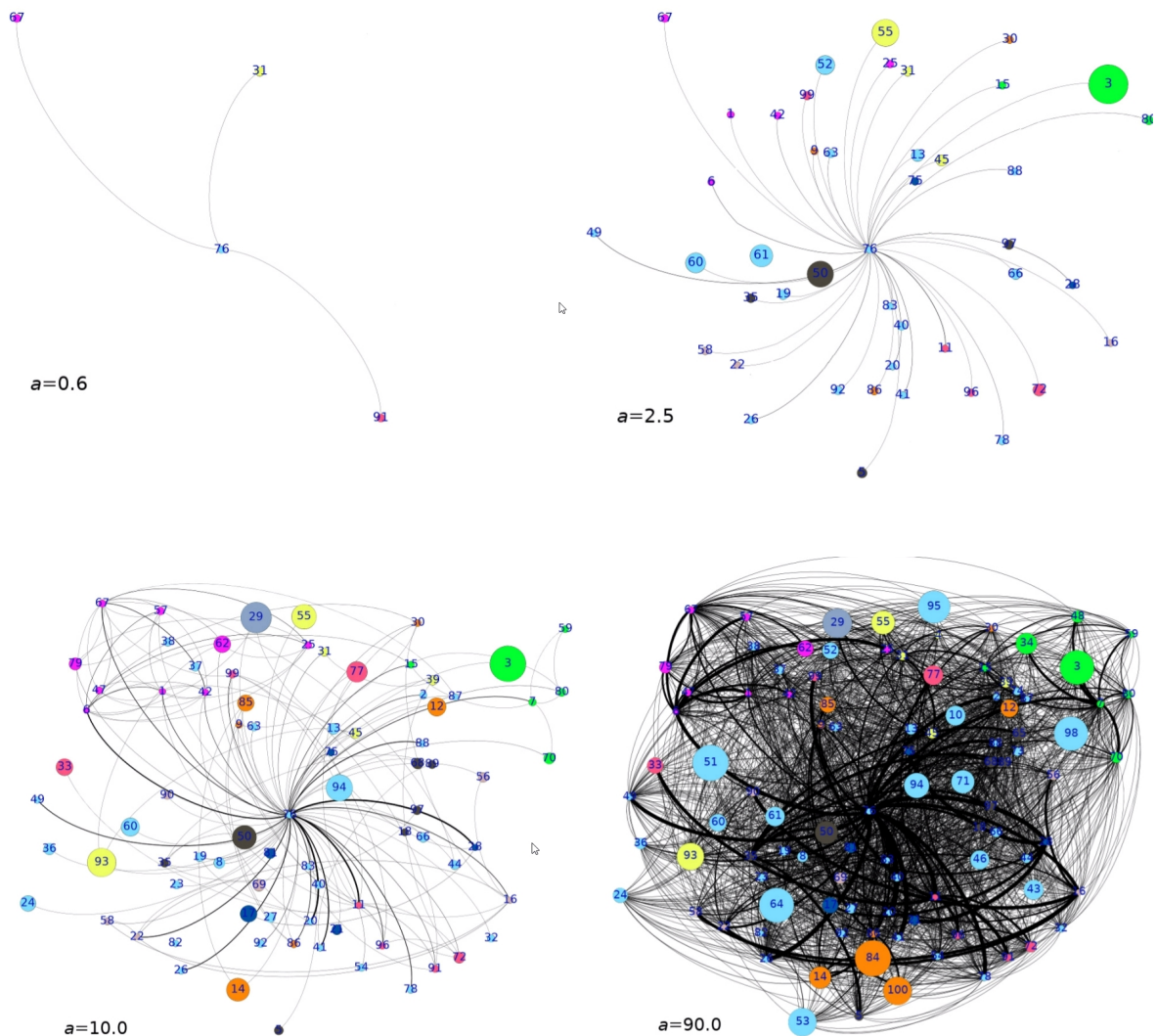


Rys. 7.22 Identyczna zależność jak na Rys. 7.21 w przypadku gdy epidemia wychodzi z węzła stołecznego rynku  $M76$



**Rys. 7.23** Ścieżki zarażania popularnością innych rynków wirusem wychodzącym z rynku  $M3$  należącego do klastra 3 (przy czasie transakcyjnym równym 2) dla różnych agresywności wirusa  $a$ .

Na dolnych wykresach Rys. 7.24 obserwujemy, że struktura jest mniej scentralizowana niż na dolnych wykresach Rys. 7.21. Faktycznie, na rzeczywistym rynku jeśli muzyka pochodzi z rynku dominującego, jej rozprzestrzenianie się może być bezpośrednie, ponieważ wpływ tego rynku na wszystkie inne jest duży. Przeciwnie dzieje się, gdy źródłem muzyki jest mały, peryferyjny kulturowo rynek. Wówczas muzyka może osiągnąć popularność przede wszystkim zdobywając jakieś większe rynki pośrednie i dopiero stamtąd powędrować na cały świat.



**Rys. 7.24** Ścieżki zarażania popularnością innych rynków wirusem wychodzącym z rynku *M76* leżącym poza klastrami (przy czasie transakcyjnym  $d = 2$ ) dla różnych agresywności wirusa  $a$ .

Przykładem zdobywania popularności przez „agresywną” piosenkę z peryferyjnego rynku poza klastrami jest rok 2012 i przebój 'Gangnam Style' koreańskiego rapera Psy (który w Europie jako pierwsza – na miesiąc przed innymi krajami – przyjęła Dania). Jednak moda na 'Gangnam Style' dotarła do swoich późniejszych fanów na całym świecie nie bezpośrednio z Korei Południowej, która w tamtym czasie miała jeszcze dość słabo rozwinięte powiązania kulturowe z resztą świata, ale poprzez serwis YouTube i media społecznościowe pochodzące z innych krajów Azji i z USA [100].



## 7.4 Porównanie z danymi empirycznymi

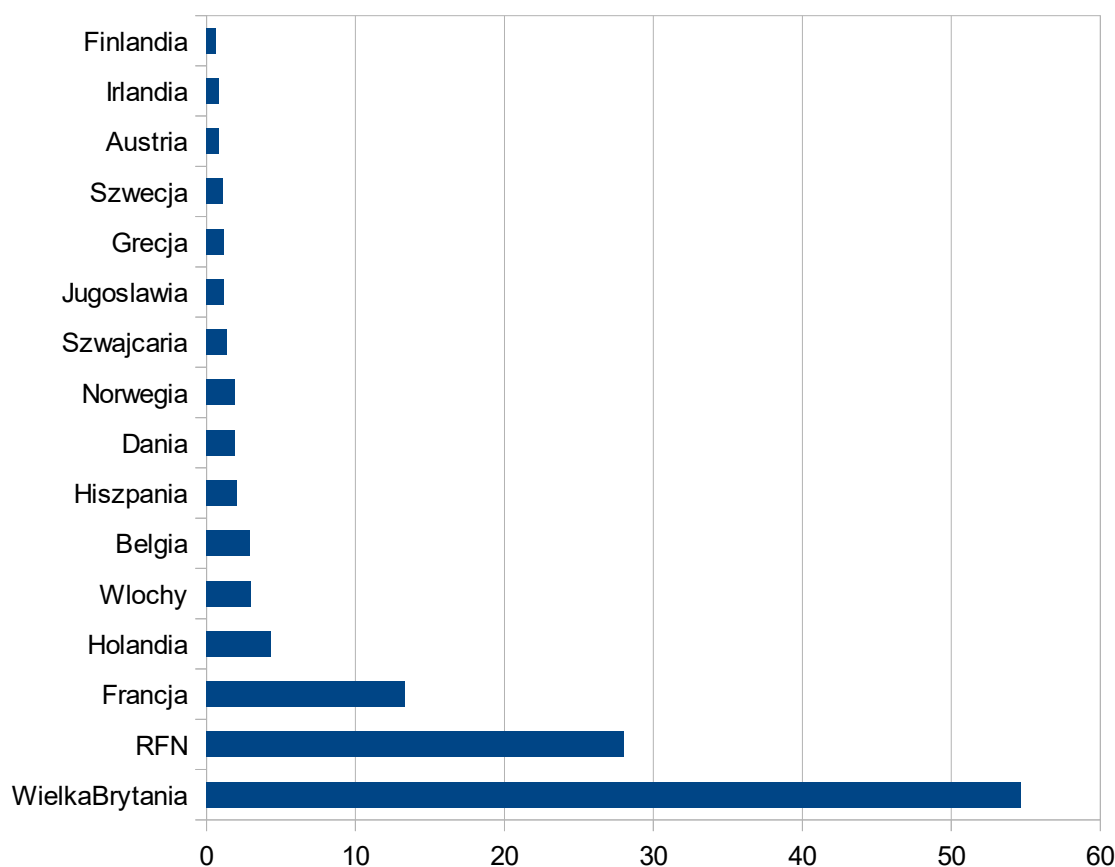
Podstawową trudnością do pokonania, aby model agentowy wyjaśniał procesy zachodzące na rzeczywistych rynkach, był fakt, że na przestrzeni 60 lat (1966-2016) jedynie 72 piosenki z 17 tysięcy opanowały swoją popularnością wszystkie 12 badanych krajów Europy Zachodniej, które notują popularność singli poprzez cotygodniowe listy przebojów. Co prawda dominacja ekonomiczna największych 72 przebojów przewyższa swoją popularnością wszystkie pozostałe z 17 tysięcy badanych piosenek (podobnie jak obroty spółek należących do głównych indeksów giełdowych przewyższają obroty pozostałych notowanych na giełdzie spółek), jednak fakt, że znikomy procent wszystkich realizacji przenosi swój sukces na wymiar międzynarodowy – mimo że jest zgodny z rzeczywistością – wymagał uzupełnienia.

Jak wykazałem w poprzednich rozdziałach, rynek fonograficzny w porównaniu z rynkami finansowymi jest przewidywalny. Predykcyjne zalety jego modelowania agentowego mogą być jednak ograniczone nie tylko technologią nośnika, ale też wydarzeniami historycznymi i kształtowanymi przez nie geografiami samych rynków. Dlatego też jazz, który apogeum przeżywał w latach 40-tych XX wieku nigdy nie przełożył swojej popularności na wielomilionowe nakłady płytowych przebojów, bowiem rozwój fonografii przerwała wówczas II wojna światowa. W efekcie upowszechnienie singla w formacie 7” nastąpiło dopiero pod koniec lat 50-tych XX wieku, a oparte o ich sprzedaż listy przebojów przyjęły się powszechnie w Europie Zachodniej dopiero w latach 60-tych.

Na podstawie danych oficjalnych (Niemcy Zachodnie, Belgia, Dania, Szwajcaria, Austria, Szwecja), danych cząstkowych (Wielka Brytania, Francja, Holandia, Hiszpania, Norwegia, Finlandia, Grecja, Irlandia) oraz własnych oszacowań (Włochy, Jugosławia) udało mi się odtworzyć rozkład udziału w europejskim rynku dla 16 krajów o największej sprzedaży singli w 1961 roku (Rys. 7.25).

Przykładowo: całkowity brytyjski rynek fonograficzny był w 1961 roku oficjalnie wart 82

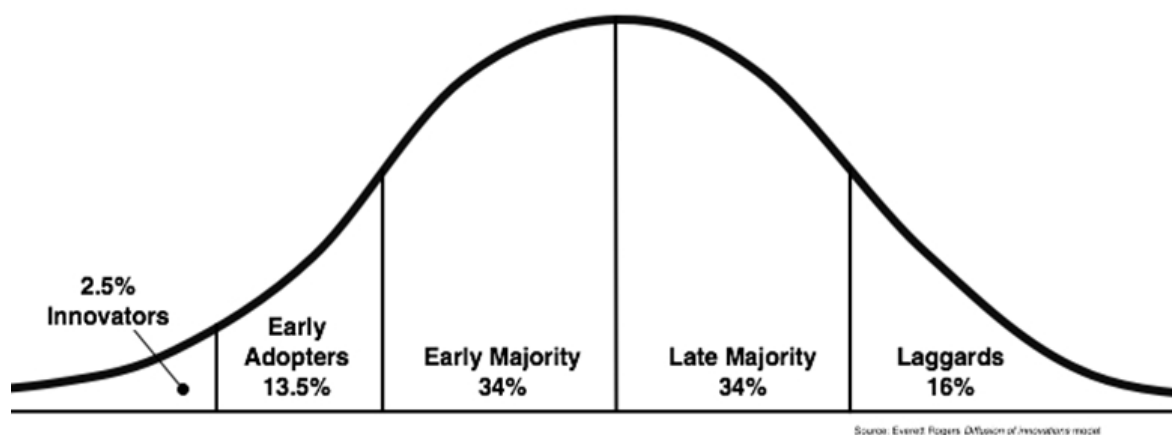
miliony dolarów. Wiedząc, że liczba wydanych singli dominowała nad liczbą płyt długogrających (albumów), oszacowałem wartość brytyjskiego rynku singli na 66%, całkowicie sprzedanych tam płyt, czyli 54,7 mln dolarów. Dla porównania: w 1961 roku oficjalne wartości rynku singli Niemiec Zachodnich to 28 mln dolarów, Belgii – 2,9 mln USD, Szwajcarii – 1,4 mln USD, Szwecji 1,1 mln USD i Austrii 826 tys. dolarów. Udział rynku Włoch i Jugosławii oszacowałem na podstawie liczby ludności tych krajów w proporcji do wielkości rynków sąsiednich (Hiszpanii i Grecji).



**Rys. 7.25** Rozkład udziału największych 16 krajów w rynku singli w Europie dla danych z 1961 roku (wartości poszczególnych rynków podaje w milionach dolarów).

Ze względu na uwarunkowania ekonomiczne Europy, można przyjąć, że do 1990 roku rozkład wskazany na Rys. 7.25 pozostawał bez zmian. Hierarchii tej nie zmieniło nawet powiększenie rynku zachodnioniemieckiego o obszar dawnego NRD. Znaczenie miał jedynie rozpad Jugosławii, dlatego w moich badaniach z poprzedniego i bieżącego rozdziału uwzględniam jedynie 12 największych i najbardziej stabilnych ekonomicznie europejskich rynków. Założenie to przyjąłem również dlatego, że popularność utworu w

kolejnych dekadach mierzona jest nie tylko sprzedażą (fizyczną lub cyfrową), ale (w niektórych krajach) także liczbą odtworzeń w mediach.



Rys. 7.26 Rozkład przyjmowania innowacji – od innowatorów (innovators), aż po maruderów (laggards) według źródła [90].

Warto zatem porównać rezultaty modelowania agentowego z danymi empirycznymi, które analizowałem w poprzednich rozdziałach. Zwłaszcza, że dla rynku singli (który analizuję według danych od lat 60. XX wieku), rozkład udziału krajów w rynku globalnym różni się od rozkładu udziału krajów w rynku albumów (pokazany na Rys. 4.3 dla 2003 roku).

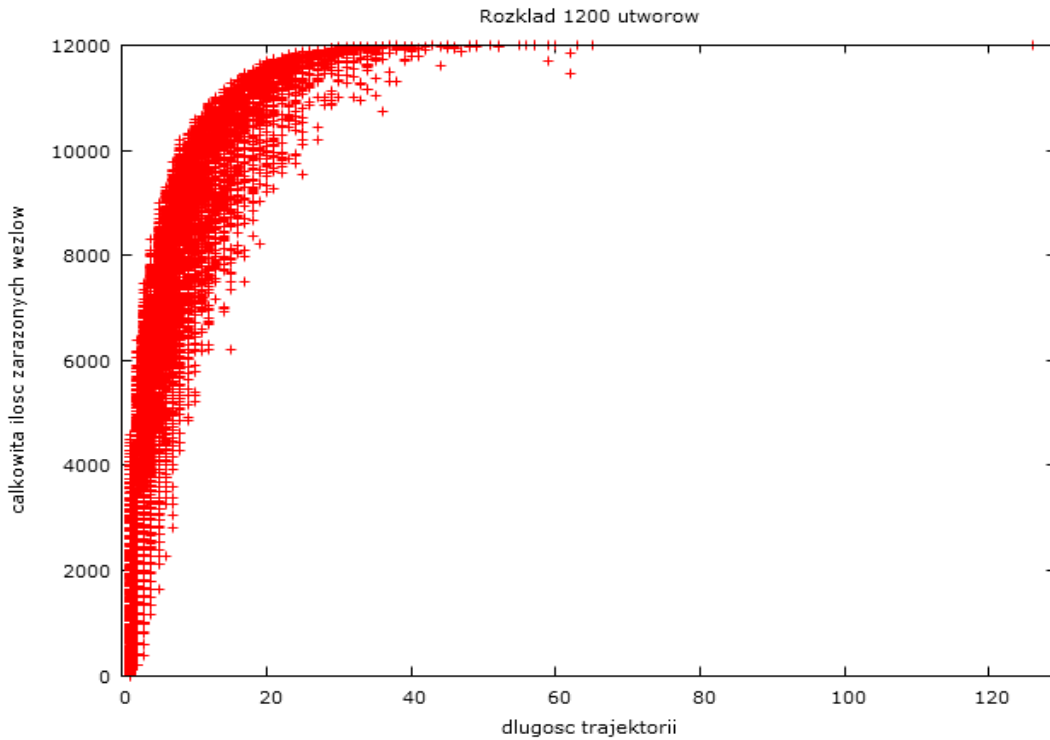
Ponadto – co pokazałem w rozdziale 6. – popularność przeboju na jednym lokalnym rynku trwa od kilku do kilkunastu tygodni. Biorąc pod uwagę wielkość rynków (Rys. 7.25), superpozycja krzywych sprzedaży (Rys. 7.13) może być wyrażona jako liczba zarażonych węzłów na rynku globalnym. W ten sposób przebieg epidemii może trwać nieco ponad rok. W praktyce jednak dla 12 badanych rynków typowy czas życia przeboju trwał od 26 tygodni do jednego roku, a jedynie w 2 przypadkach na 100 trwał dłużej. Miało to miejsce dla piosenek Take On Me (1984) norweskiej grupy a-ha (węzeł inicjujący: Norwegia, czas trwania epidemii: 2 lata), oraz Macarena (1992) hiszpańskiej grupy Los Del Rio, która do osiągnięcia światowej popularności w 1996 roku przebiegała się najpierw lokalnie z rynku austriackiego. Powyższe 2 wyjątkowe przypadki są zatem przykładem utworów o wysokiej chwytności, ale inicjowanych z peryferyjnych rynków (Rys. 7.18). Szczegółowy rozkład zależności całkowitej liczby zarażonych węzłów od rzeczywistej długości ścieżki sprzedaży dla 1200 najpopularniejszych utworów ostatnich 60 lat przedstawia Rys. 7.28.

week 16 / 95			<b>Border Breakers</b>		© BPI Communications BV	
Mainland European records breaking out of their country of signing						
TW	LW	WOC	Artist/Title	Original Label	Country Of Signing	TS
①	1	11	<b>SCATMAN JOHN/SCATMAN</b> (SKI-BA-BOP-BA-DOP-BOP)	(ICEBERG/RCA)	DENMARK	54
2	3	7	<b>Roxette/Vulnerable</b>	(EMI)	SWEDEN	48
③	>	NE	<b>Alex Party/Don't Give Me Your Life</b>	(UMM)	ITALY	44
4	2	10	<b>Snap/The First The Last Eternity (Till The End)</b>	(Ariola)	GERMANY	38
⑤	8	4	<b>Corona/Baby Baby</b>	(DWA)	ITALY	36
6	5	6	<b>2 Unlimited/Here I Go</b>	(Byte)	BELGIUM	24
7	4	17	<b>Rednex/Old Pop In An Oak</b>	(Jive)	HOLLAND	19
8	6	13	<b>Lavinia Jones/Sing It To You (Dee-Doob-Dee-Do)</b>	(Virgin)	GERMANY	21
9	9	7	<b>Sparks/When Do I Get To Sing My Way</b>	(Logic)	GERMANY	20
10	7	12	<b>Ten Sharp/After All The Love Has Gone</b>	(Columbia)	HOLLAND	19
⑪	>	NE	<b>Caught In The Act/Love Is Everywhere</b>	(Undercover Records)	HOLLAND	16
⑫	13	4	<b>La Bouche/Be My Lover</b>	(Hansa)	GERMANY	16
13	10	25	<b>MC Sar &amp; The Real McCoy/Another Night</b>	(Hansa)	GERMANY	13
⑭	>	NE	<b>Joey Tempest/A Place To Call Home</b>	(Polar)	SWEDEN	11
15	15	3	<b>Alliance Ethnik/Respect</b>	(Delabel)	FRANCE	13
⑯	>	NE	<b>MC Sar &amp; The Real McCoy/Love And Devotion</b>	(Hansa)	GERMANY	11
17	14	22	<b>Roxette/Run To You</b>	(EMI)	SWEDEN	9

**Rys. 7.27** Przykładowa lista piosenek, które zdobywały popularność przekraczając granice swoich macierzystych rynków według tygodnika Music & Media z 22 kwietnia 1995 nr 16/1995.

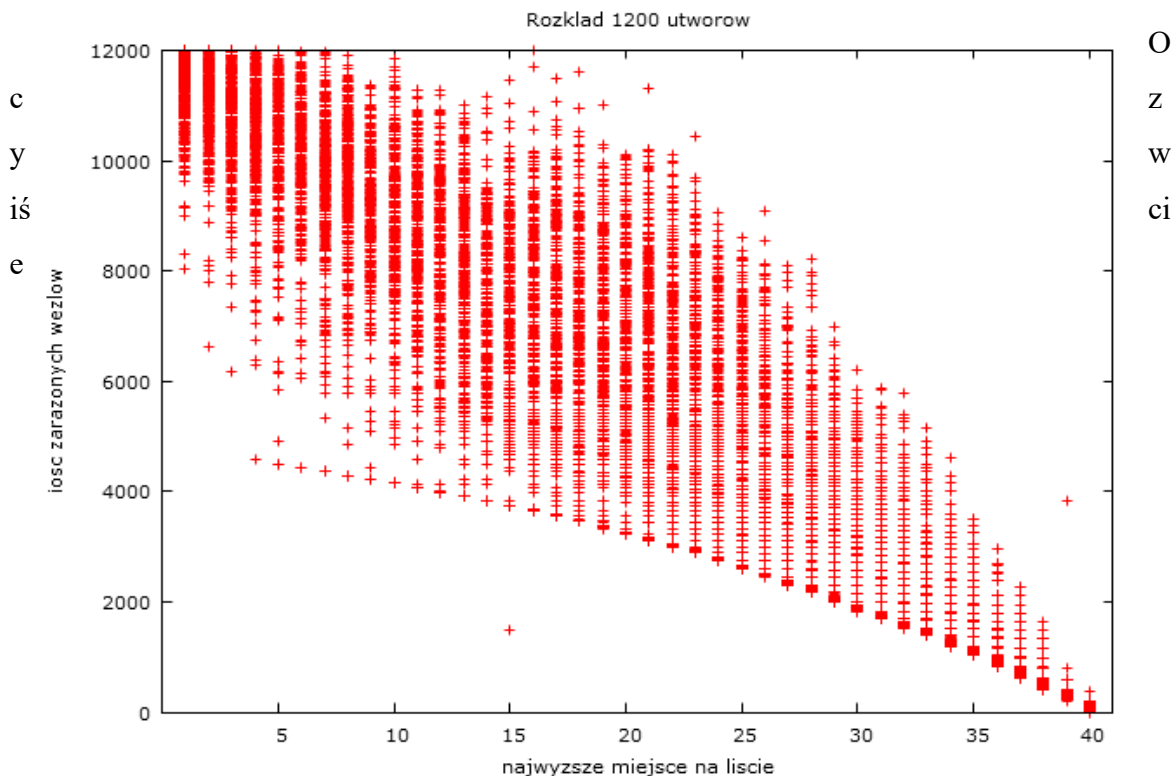
Analizowane 1200 utworów zostało wybrane z 17 tysięcy na podstawie kryterium ich popularności zdefiniowanej jako zarażenie macierzystego rynku (czyli osiągnięcie co najmniej miejsca w pierwszej piątce cotygodniowej listy przebojów) lub zarażenie innych rynków. Pozostałe utwory nie zdobyły bowiem takiej popularności przez co były mniej interesujące pod względem badawczym. Całkowita liczba zarażonych węzłów oznacza tu liczbę setek sprzedanych egzemplarzy na wszystkich 12 rynkach w Europie (bowiem oficjalne dane o sprzedaży zwykle podawane są z dokładnością do 100 egzemplarzy). Na tej podstawie możliwe jest układanie cotygodniowej listy najpopularniejszych singli w Europie, co zresztą robił tygodnik Music & Media. Analizując taką listę, możliwe jest zbadanie rzeczywistej sprzedaży całkowitej utworu od najwyższego miejsca na liście (Rys. 7.26) obejmującej pozycje od 1 do 40. Okazuje się wówczas, że przewidziany przez

model krytyczny próg atrakcyjności  $a$  powyżej którego utwór zdobywa globalną popularność, rzeczywiście może istnieć, ponieważ z reguły niższe miejsce przekłada się na niższą popularność (czyli mało zarażonych węzłów), a wyższe na więcej zarażeń. Jeśli jednak utwór osiągnie miejsce powyżej czwartego na macierzystym rynku, to jego popularność może rosnać już skokowo, dzięki przeniesieniu epidemii na inne rynki.



**Rys. 7.28** Rozkład zależności całkowitej liczby zarażonych węzłów (wyrażonej liczbą setek sprzedanych egzemplarzy singla) od długości krzywej sprzedażowej (liczonej w tygodniach)

**Rys. 7.29** Rozkład zależności całkowitej liczby zarażonych węzłów (wyrażonej liczbą setek sprzedanych egzemplarzy singla) od najwyższego miejsca na liście



na wyjaśnione w tym rozdziale scenariusze zdobywania popularności ma wpływ rozwój technologiczny zarówno nośnika utworu, jak i jego dystrybucji. Dlatego wprowadzone w modelu agentowym parametry kontrolne jak agresywność wirusa  $a$  czy czas transakcyjny  $d$  należy rozumieć w kontekście technologii przekazywania popularności i np. sytuacja gdzie atrakcyjność  $a$  staje się coraz większa historycznie mogłaby odpowiadać przyszłości, w której utwór jest bardziej atrakcyjnie wyprodukowany, niż utwory z obecnych czasów, ponieważ w technologii produkcji muzyki domyślnie zakładamy, że nowsze utwory mają przewagę nad starszymi. Z tego powodu – przy całym szacunku dla kompozytorów – zdobywanie popularności w czasach przed powstania rynku fonograficznego w XX wieku, można sobie wyobrażać jako ekspansję utworów o stosunkowo mniejszej atrakcyjności, ale za to z dłuższym czasem transakcyjnym  $d > 2$ , ponieważ do XIX wieku, przed wynalezieniem radia i telefonu media nie były w stanie przekazywać informacji z szybkością tygodnia, a w przypadku książek, z powodu konieczności ich tłumaczenia, zdobywanie zagranicznych rynków trwało nawet kilka lat.

## 8. Podsumowanie

W pracy poddałem analizie dane z wielu rynków akcji, rynku walutowego oraz rynku fonograficznego reprezentującego rynki towarowe. Wychodząc od tradycyjnych pojęć przy użyciu metod wywodzących się z fizyki statystycznej, takich jak macierze korelacji, z teorii grafów - analiza sieciowa, drzewa minimalnej rozpiętości, oraz wprowadzając nowe pojęcia, takie jak czas życia korelacji, badałem strukturę tych rynków i relacje zachodzące pomiędzy handlowanymi na nich walorami (akcje, kursy wymiany walut, artykuły). Uzyskane przez mnie wyniki są na ogół zgodne z dotychczasowym stanem badań nad tymi rynkami, jednak przedstawione za pomocą nowych pojęć, rzucają dodatkowe światło na ich własności. Dzięki korelacjom między akcjami możemy bowiem odtworzyć strukturę rynków za pomocą drzewa minimalnej rozpiętości, dobrze odzwierciedlającego sektory i podsektory branżowe badanych spółek, a także podzielić korelacje na silne, słabe i ujemne. Nowe pojęcie jak czas życia korelacji, zdefiniowane jako czas przebywania współczynnika korelacji na silnym poziomie (powyżej 0,5), pozwala określić trwałość korelacji między akcjami i wyznaczać, na podstawie historii cen, optymalne okno czasowe potrzebne do jego badań (znane wcześniej ekonomistom i inwestorom jedynie w praktyce), a także ilościowy wpływ i wagę wydarzeń historycznych na trwałość współczynnika korelacji.

Badając rynek walutowy zwracam uwagę na znaczenie wyboru waluty bazowej,

względem której liczymy wartość innych walut, co rzecz jasna ma wpływ na niespotykaną na innych rynkach dużą różnorodność rozkładów korelacji między walutami i strukturę drzewa minimalnej rozpiętości. Dlatego do miary trwałości drzew używam współczynnika przetrwania, zdefiniowanego restrykcyjnie lub nierestykcyjnie, dzięki czemu możemy oszacować w jaki sposób rozpadają się połączenia w drzewie. Jego postać, obliczona dla różnych walut bazowych, odzwierciedla geograficzne zależności między państwami i w dużej mierze pokrywa się z już znanymi wynikami.

Rynek fonograficzny jako szczególny przypadek rynku towarowego pod względem statystycznym wykazuje tylko częściowe podobieństwo do rynków finansowych, bowiem artyści, których notowania śledzimy na listach sprzedaży, konkurują ze sobą o zasoby finansowe konsumentów. Stąd obserwowana na rynku płytowym przewaga korelacji słabych i ujemnych. Warto zaznaczyć, że mimo słabo skorelowanej natury tego rynku, drzewa minimalnej rozpiętości całkiem dobrze odzwierciedlają podobieństwa stylistyczne pomiędzy artystami.

Na statystyczne własności rynku fonograficznego ma wpływ dynamika sprzedaży. Jej model stochastyczny udało się opracować w dwóch wersjach, różniących się założeniami: wersji stacjonarnej i niestacjonarnej, uwzględniającej warunki rynkowe w sposób bardziej kompletny. W wyniku konfrontacji wyników symulacji z rzeczywistością wykryłem, że przedświadczeni wzrost sprzedaży płyt nie wynika z zagęszczenia nowych premier płytowych, co byłoby wnioskiem bardziej intuicyjnym, biorąc pod uwagę założenie, że grupa konsumentów muzyki jest stała. Wzrost ten jest bowiem znacznie większy niż przewidywania modelu opartego o stałą grupę konsumentów, dochodzi dodatkowa, zewnętrzna grupa konsumentów, odpowiadająca za główną część wzrostu sprzedaży. Grupa ta zna rynek pobieżnie i dlatego jest zainteresowana jedynie płytami największych gwiazd z różnych gatunków muzycznych. W ten sposób w drzewie minimalnej rozpiętości zamiast stylistycznego sektora muzyki pop, powstaje sektor gwiazdorski, podczas gdy pozostałe gatunki muzyczne posiadają osobne sektory. Można dodać w tym miejscu, że zaproponowany w pracy model stochastyczny pozwala dosyć dokładnie przewidzieć ilościowo trajektorię sprzedaży danego artysty, mając do dyspozycji dane o sprzedaży poprzednich jego płyt. Ta zdolność prognostyczna modelu wynika ze specyfiki rynku fonograficznego, która odróżnia go od rynków finansowych, gdzie prognozowanie jest nadzwyczaj trudne, a często wręcz niemożliwe. Warto też podkreślić, że kolektywne



efekty zaobserwowane na rynku fonograficznym mają charakter emergentny i stanowią mogą przesłankę, że rynek ten ma charakter układu złożonego.

Prognozowanie rozprzestrzeniania się popularności jest możliwe także w sytuacji, kiedy lokalne rynki oddziałują na siebie – a tak jest w przypadku pojedynczych utworów (singli). Skuteczne okazało się statystyczne wykrywanie najbardziej prawdopodobnych geograficznych ścieżek zakażeń (w różnych erach historycznych – analogowej, cyfrowej, internetowej) i sprawdzenie, że zjawisko to może zachodzić także w mniejszej skali pojedynczego rynku – na przykładzie Wielkiej Brytanii (gdzie analiza widma wartości własnych macierzy korelacji sprzedażowych między utworami pozwala zauważyć, że klasyczny cykl życia produktu jest dominującym scenariuszem zdobywania popularności przez utwór). O ile Wielka Brytania jest zazwyczaj krajem inicjującym popularność piosenek w Europie, to w samej Wielkiej Brytanii źródłem inicjującym popularność utworu jest Londyn. Co ciekawe, zarówno w skali tego kraju, jak i całej zachodniej Europy, wpływ na ścieżki infekcji miały podobieństwa kulturowe między krajami jak i ich położenie geograficzne, bowiem (zwłaszcza w erze przedinternetowej) na podstawie opóźnień między opanowaniem kolejnych rynków, można wyróżnić klastry krajów, częściowo pokrywające się z ich odrębnością kulturową i geograficzną.

Wyniki te stały się inspiracją aby zaproponować w pracy epidemiologiczny model agentowy, pozwalający prześledzić różne warianty w zależności od struktury zmieniającego się rynku globalnego, w oparciu o jego własności sieciowe i parametry kontrolne jak atrakcyjność utworu (agresywność wirusa), czas pozostawania węzła w stanie zakażenia (czas transakcyjny) i wybór miejsca startu (peryferyjny bądź nie). Co ciekawe, wśród bogatych wariantów, model przewiduje możliwość istnienia na lokalnym rynku prognozy epidemicznego, który w modelach SIR opartych na bezskalowej architekturze sieciowej raczej nie występuje. Wpływ na ten wynik ma założenie, że prawdopodobieństwo zarażenia węzła jest odwrotnie proporcjonalne do jego krotności (odbiorcy posiadający większy dostęp do muzyki są bardziej odporni na przyjmowanie proponowanych im losowo utworów, niż odbiorcy z małym dostępem), co przekłada się na fakt, że zauważenie piosenki na rynku innym, niż macierzysty zwiększa jej szanse na dalszą popularność w kolejnych krajach. Popularność globalna – we wszystkich krajach – jest jednak zjawiskiem wyjątkowym, dlatego ważnym odkryciem jest, że dopiero założenie odrębności kulturowej (która prowadzi do podziału globalnej sieci na klastry)

daje największą zgodność modelu z rzeczywistością (zwłaszcza tą z ery analogowej i cyfrowej, gdy rynek globalny jeszcze nie zależał od Internetu).

Mimo, że proponowane w pracy modele są komplementarne i dają dużą zgodność z rzeczywistością, ich rozwinięcie jest możliwe przy wprowadzeniu nowych parametrów, które już w sposób konkretny wpływają na przebieg popularności (elementów promocji, dystrybucji, reakcji mediów, jakości utworu, etc.). Dotyczy to także analizowanych w pierwszych rozdziałach rynków finansowych i wpływu na nie konkretnych wydarzeń historycznych. Dlatego czas życia korelacji jako miara statystyczna może mieć zastosowanie także w innych dziedzinach. Interesująca byłaby też próba analitycznego uzyskania krzywej sprzedaży jako funkcji zależnej od kilku parametrów (np. jakości płyty), której maksima w kolejnych krokach zależą od całkowitej sprzedaży poprzedniej płyty.

Uzyskane w pracy wyniki i ich zgodność z danymi empirycznymi potwierdzają użyteczność modeli stosowanych w fizyce statystycznej do badań układów złożonych, a w szczególności rynków. Modele agentowe pozwalają na odtworzenie struktury rynku, składającego się z poziomu krajowego i globalnego, z których każdy ma dość odmienne własności, a także na odtworzenie głównych aspektów dynamiki sprzedaży. Bardzo pomocne są tutaj też proste statyczne modele sieciowe, oparte na drzewach minimalnej rozpiętości, dzięki którym można wydobyć nowe, nietrywialne aspekty rynków finansowych i fonograficznych. Procesy stochastyczne, modelujące ewolucję wyników sprzedaży, pozwalają w pewnym zakresie na prognozowanie tych wyników. Jest to jeszcze jeden przykład uniwersalności metod tworzonych w fizyce do badań nad układami, w których podstawą jest oddziaływanie między elementami składowymi, ale które tylko pozornie wydają się przekraczać obszar zainteresowania standardowo rozumianej fizyki.

## Literatura

- [1] A.K. Wróblewski *Historia Fizyki*, Warszawa, PWN (2007), 28-33
- [2] W. Tatarkiewicz *Historia Filozofii* t.3, PWN (1991)
- [3] Z. Burda, J. Jurkiewicz, M.A. Nowak *Signal and Noise in Correlation Matrix* 2003, Acta Phys. Polon. B 34
- [4] Bachelier *Théorie de la spéculation* praca doktorska (1900)
- [5] P. Erdi *Complexity Explained (Springer Complexity)*, Springer-Verlag (2007)
- [6] S. J. Gould, N. Eldredge *Paleobiology* 3 (1977) 115-151
- [7] R. Poczobut *System-struktura-emergencja*, [w:] M. Heller, J. Mączka (red.), *Struktura i emergencja*, Kraków-Tarnów: PAU-OBI, s. 11-38 (2006)
- [8] P. Bak, C.Tang, K. Wiesenfeld *Self-organized criticality: an explanation of 1/f noise* Physical Review Letters 59 (4): 381–384 (1987).
- [9] D.L. Turcotte, J.B. Rundle *Self-organized complexity in the physical, biological, and social sciences*, Proc. Nat. Acad. Sci. USA, 99, Supplement 1, 2463-2465, (2002).
- [10] M. E. J. Newman *Power laws, Pareto distributions and Zipf's law*. Contemporary Physics 46, 323 (2005).
- [11] S.Lloyd *Measures of complexity: a nonexhaustive list*. IEEE Control Systems Magazine 09/2001; 21(4):7-8 (2001)
- [12] J. Kwapien, S. Drożdż *Physical Approach to Complex Systems* Physics Reports 515 (2012) 115-226
- [13] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, B.I. Shraiman *Fractal measures and their singularities: the characterization of strange sets* Phys. Rev. A 33 (2), 1141, (1986) 3249
- [14] J. Wishart *The generalised product moment distribution in samples from a normal*

*multivariate population*, *Biometrika* 20A (1–2): 32–52 (1928)

- [15] A. M. Sengupta, P. P. Mitra *Distributions of singular values for some random matrices* *Phys. Rev. E* 60, 3389 (1999)
- [16] L. Laloux, P. Cizeau, J.P. Bouchaud, M. Potters, *Random Matrix Theory and Financial Correlations*, *Phys. Rev. Lett.* 83, 1467 (1999)
- [17] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amara, T. Guhr, H. E. Stanley, *Random Matrix Approach To Cross Correlations In Financial Data*, *Physical Review E*, (2002) 65
- [18] AL Barabási, R Albert *Emergence of scaling in random networks* *Science* 286 (5439), 509-512 1999
- [19] A. Fronczak, P. Fronczak *Świat sieci złożonych. Od fizyki do Internetu*. PWN (2009)
- [20] R. Percacci, A. Vespignani, *Scale-free behavior of the Internet* , *Eur. Phys. J. B* 32 (2003) 411–414.
- [21] R. Guimerà, S Mossa, A Turttschi, *The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles*, *LAN Amaral Proceedings of the National Academy of Sciences* 102 (22), (2005) 7794-7799
- [22] R. Tanaka *Scale-rich metabolic networks*. *Phys. Rev. Lett.* 94 (2005) 168101
- [23] J. Kwapien, S. Gworek, S. Drożdż, A. Górski, *Analysis of a network structure of the foreign currency exchange market*, *Journal of Economic Interaction and Coordination* (2009)
- [24] M.E.J. Newman, *Scientific collaboration networks. Network construction and fundamental results*, *Phys. Rev. E* 64 (2001) 016131
- [25] S. Redner, *How popular is your paper? An empirical study of the citation distribution*, *Eur. Phys. J. B* 4 (1998) 131–134.
- [26] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, *The architecture of complex weighted networks*, *Proc. Natl. Acad. Sci. USA* 101 (2004) 3747–3752
- [27] M. Krawczyk, L. Muchnik, A. Mańka-Krasoń, K. Kułakowski *Line graphs as social networks*, *Physica A* 390 (2011) 2611–2618
- [28] J. Travers, S. Milgram *An Experimental Study of the Small World Problem*, *Sociometry* 32 (1969) 425–443
- [29] [www.facebook.com](http://www.facebook.com)
- [30] D.J. Watts, S.H. Strogatz *Collective Dynamics of SmallWorld Networks*, *Nature* 393 (1998) 440–442.
- [31] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang *Complex networks: Structure and dynamics*, *Phys. Rep.* 424 (2006) 175–308.
- [32] R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin *Resilience of the Internet to Random Breakdowns*, *Phys. Rev. Lett.* 85, 4626-4628 (2000).
- [33] B.C.J. Moore *An Introduction to Psychology of Hearing*, Brill (Leiden 2012)
- [34] H. Markowitz *Portfolio Selection: efficient diversification of Investment*, John Wiley, New

York, (1959)

- [35] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley, *Universal and Non-Universal Properties of Cross-Correlations in Financial Time Series*, Phys. Rev. Lett. 83, 1471-1474 (1999).
- [36] Maslov *Measures of globalization based on cross-correlations of world financial indices*, Physica A 301, 397-406 (2001)
- [37] S. Drożdż, F. Ruf, J. Speth, M. Wójcik *Imprints of log-periodic self-similarity in the stock market*, Eur. Phys. J. B 10 (1999) 589
- [38] R. N. Mantegna *Hierarchical Structure In Financial Markets*, Eur. Phys. J. B., (1999)
- [39] M. Mezard, G. Parisi, M. A. Virasoro *Spin glass theory and beyond*, World Scientific, Singapore (1987)
- [40] R. Rammal, G. Toulouse, M. A. Virasoro *Ultrametricity for physicists*, Review Of Modern Physics, t. 58 (1986) 765-788
- [41] Joseph. B. Kruskal *On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem*, Proceedings of the American Mathematical Society, Vol 7, No. 1 (1956) 48–50
- [42] D.B. West *Introduction To Graph Theory*, Prentice-Hall, Englewood Cliffs NJ (1996)
- [43] J.P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto *Dynamics Of Market Correlations: Taxonomy And Portfolio Analysis*, Physical Review E, 68, 056110, (2003)
- [44] G. Bonanno, *Networks of equities in financial markets*, Eur. Phys. J. B. 38, 363-371 (2004)
- [45] E.J. Elton, M.J. Gruber *Modern Portfolio Theory and Investment Analysis*, J. Wiley and sons, New York (1995)
- [46] B. Toth, J. Kertesz *Accurate estimator of correlations between asynchronous signals*, Physica A 388 (2009) 1696-1705
- [47] K. Jajuga, inf. osobista, 2003
- [48] A. Buda, A. Jarynowski *Life-time of correlations and its applications vol. 1*, Wydawnictwo Niezalezne, Wrocław (2010)
- [49] T. Epps *Comovements in stock prices in the very short run*, J. Amer. Statist. Assoc. 74 (1979) 291-298
- [50] S. Galluccio, G. Caldarelli, M. Marsili, Y.C. Zhang *Scaling in currency exchange*. Physica A 245(3-4), (1997) 423-436
- [51] T. Mizuno, S. Kurihara, M. Takayasu, H. Takayasu *Analysis of high-resolution foreign exchange data of USD-JPY for 13 years* Physica A: Statistical Mechanics and its Applications 324 (1), 296-302, 59, (2003)
- [52] R. Gençay, B. Whitcher F. Celcuk, *Scaling Properties of Foreign Exchange Volatility*, Physica A, v289, 249-266. (2001)
- [53] A. Fisher, L. Calvet, B. Mandelbrot *Multifractality of Deutschemark/US dollar exchange rates*, Cowles Foundation Discussion Paper 1166, (1997).

- [54] F. Schmitt, D. Schertzer, S. Lovejoy *Multifractal analysis of foreign exchange data* Appl. Stoch. Mod. D. A. 15 (1999) 29–53.
- [55] M. McDonald, O. Suleman, S. Williams, S. Howinson, N. F. Johnson *Detecting a currency's dominance or dependence using foreign exchange network trees*, Phys. Rev. E72 (2005)
- [56] T. Mizuno, H. Takayasu and M. Takayasu *Correlation networks among currencies*, Physica A: Statistical Mechanics and its Applications, Vol.364, (2006), 336-342.
- [57] E.F. Fama *The Behavior Of Stock Market Prices* Journal Of Business 38, (1965) 34-105
- [58] K. Ivanova, M. Ausloos *False EUR Exchange Rates vs DKK, CHF, JPY and USD what is the strong currency?* Empirical Science Of Financial Fluctuations The Advent Of Econophysics, pp 51-60 (2002)
- [59] A. Buda *Historia rocka, popu i hip-hopu według krytyków 1974-2006*, Wrocław (2006)
- [60] *The IFPI Report Recording industry in numbers*, IFPI, 2010
- [61] *The IFPI Digital Music Report*, IFPI, 2011
- [62] *The IFPI Report Recording industry in numbers*, IFPI, 2012
- [63] H. E. Hurst, Trans. Am. Soc. Civ. Eng, 116: 770 (1951)
- [64] L. Lacasa, B. Luque, F. Ballesteros, J. Lique, J.C. Nuno *From Time Series To Complex Networks: Visibility Graphs*, PNAS vol. 105 no 13., (2008), 4972-4975
- [65] A. Buda *Does pop music exist? Hierarchical structure in phonographic markets*, Physica A: Statistical Mechanics and its Applications (2012) 5153
- [66] A. Buda, A. Jarynowski *Network structure of phonographic market with similarities between artists*, Acta Physica Polonica A 123 (3) (2013)
- [67] A. Buda *Encyklopedia hip-hopu*, Wydawnictwo Niezależne (2005)
- [68] R.N. Mantegna, H.E. Stanley *An Introduction To Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, 1999.
- [69] A. Jarynowski *Critical Phenomena In Nature*, Study Report, Wrocław University 2007  
[<http://people.su.se/~ajary/krachylin.pdf>]
- [70] I. Simonsen *Measuring anti-correlations in the Nordic electricity spot market by wavelets*, Physica A 322 (2003)
- [71] S. Ghoshghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Nature 381 (1996) 767–770.
- [72] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, E. Stanley *Inverse Cubic Law For The Distribution Of Stock Price Variations*, European Physics Journal B3, 139-140 (1998)
- [73] M.M. Dacorogna, U. A. Muller, R. J. Nagler, R. B. Olsen, O. V. Picket *A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market*, Journal Of International Money And Finance (1993), t 12, s 413-438
- [74] R.N. Mantegna, H.F. Stanley *Turbulence And Financial Markets*, Nature (1996), t. 383, s. 587-588
- [75] M. Cristelli, L. Pietronero, A. Zaccaria *Critical Overview of Agent-Based Models for*

*Economics*. (2011), arXiv 1101.1847

- [76] W. Paul, J. Baschnagel *Stochastic Processes. From Physics to Finance*, Springer (2000)
- [77] .J. P. Bouchaud, R. Cont, *A Langevin Approach To Stock Market Fluctuations And Crashes*, Physica A (1998), t. B6, 543-550
- [78] T. Bollerslev *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 31(1986), 307–327
- [79] B.B. Mandelbrot, A. Fisher, L. Calvet *A Multifractal Model of Asset Returns*, Cowles Foundation Discussion Papers 1164 (1997)
- [80] Zhang, Yi-Cheng *Toward a Theory of Marginally Efficient Markets* Physica A,269: (1999) 30–44
- [81] D.Sornette *Critical market crashes*. Physics Reports 378, 1–98.(2003)
- [82] J. Kwapień, S. Drożdż, J. Speth *Time scales involved in emergent market coherence*, Physica A 337, (2004) 231-242.
- [83] J. Janczura, R. Weron *An empirical comparison of alternate regime-switching models for electricity spot prices*, Energy Economics 32 (5), pp. 1059-1073 (2010).
- [84] J.B. Rundle, D.L. Turcotte, R. Shcherbakov, W. Klein, C Sammis *Micro and macroscopic models of rock fracture*. Rev. Geophys 41 (4),1019, 258, (2003)
- [85] L. Gulia, S. Wiener, Geophys. Res. Lett. 37, L10305 (2010)
- [86] P. Bak, C. Tang, K. Wiesenfeld *Self-organized criticality* Phys. Rev. A38, 364 (1988)
- [87] H. J. S. Feder, J.Feder *Self-organized criticality in a stick-slip process* Phys. Rev. Lett., 66:2669-2672 (1991)
- [88] W. Klein et al *Complexity and the Physics of Earthquakes*, Am. Geophysical Union (2000)
- [89] A. Buda, A. Jarynowski, *Dynamics Of Phonographic Markets. Record Industry Stochastic Model And Empirical Data*, Acta Physica Polonica B (PS), 2013
- [90] J. Stark, *Product Lifecycle Management: 21st Century Paradigm for Product Realisation*, Springer, ISBN 978-1-85233-810-7 (2004).
- [91] M.E.J. Newman, *Analysis of weighted networks*, Phys. Rev. E 70, 056131 (2004).
- [92] A. Jarynowski, A. Buda, *Diffusion Paths Between Product Life-Cycles In European Phonographic Markets*, Control And Cybernetics, 45 (2) (2016)
- [93] A. Buda, A. Jarynowski, *Exploring patterns in European singles charts*, 2015 Second European Network Intelligence Conference, 135-139 (2015)
- [94] R. Inglehart, C. Welzel, *Modernization, Cultural Change and Democracy*, New York and Cambridge: Cambridge University Press, (2005)
- [95] A.L. Barabasi, *Scale-free networks: a decade and beyond*, Science 412-413, 2009
- [96] G. Thomas Jr, *Building the buzz in the hive mind*, Journal of Consumer Behaviour 4 (1): 64–72. doi:10.1002/cb.158., 2006
- [97] A. Saumell-Mendiola, M.A. Serrano, M. Boguna, *Epidemic spreading*

*interconnected networks*, Phys. Rev. E 86, 026106 (2012).

[98] S. Shai, S. Dobson, *Effect of resource constraints on intersimilar coupled networks*, Phys. Rev. E 86, 066120 (2012)

[99] A. Buda, K. Kuźmich, *How to plug out of the networks in jeopardy of ASF, Covid-19, social media or markets toxicity*, E-methodology No. 2020 (7), (2020)

[100] S. Jung, H. Li, *Global production, circulation, and consumption of Gangnam Style*, Int. J. Commun. 8, 2790-2810 (2014)

[101] M. Dickison, S. Havlin, H.E. Stanley, *Epidemics on interconnected networks*, Phys. Rev. E 85, 066109 (2012)

[102] G. Siudem, J. Hołyst, *Diffusion on hierarchical systems of weakly-coupled networks*, Physica A 513, 675-686 (2019)

[103] A. Buda, J. Kwapień, *Agent-Based Modeling of the global phonographic market*, Physica A: Statistical Mechanics and its Applications, Volume 598 (2022)