MONOGRAFIE

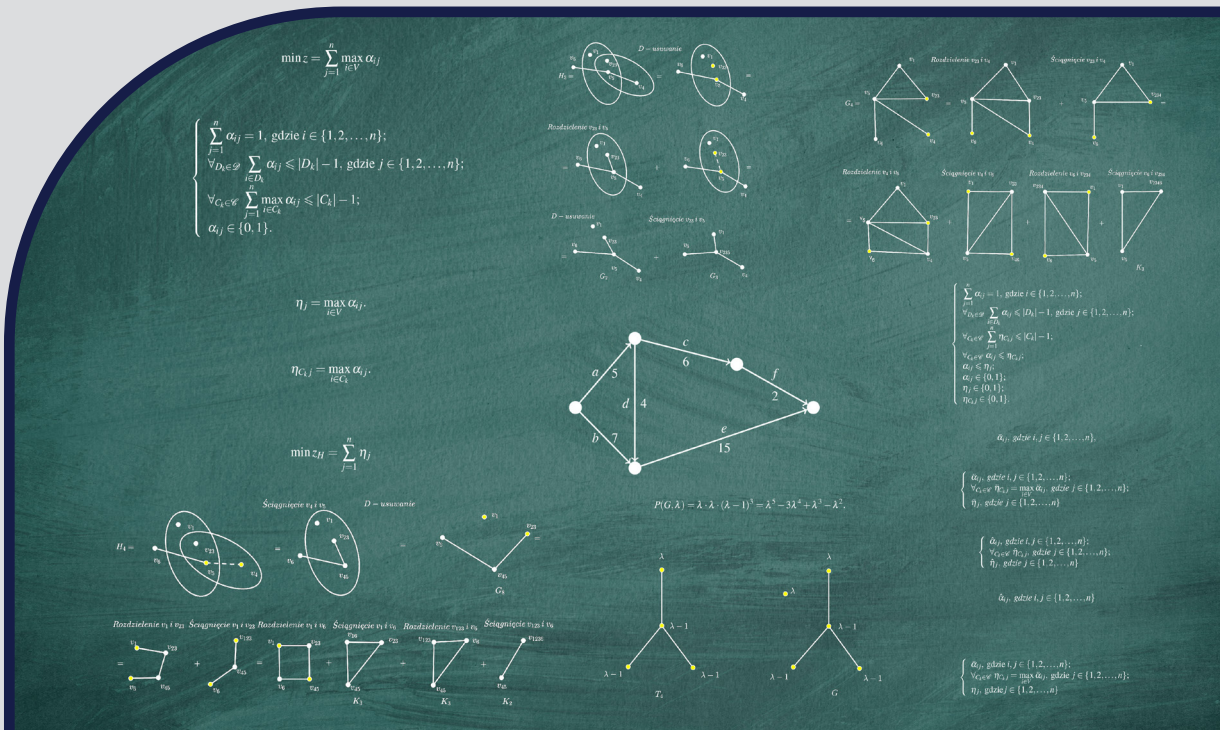
POLITECHNIKA LUBELSKA

Matematyczne miscellanea

Tom 4

Wybrane narzędzia modelowania matematycznego

Redakcja
Izolda Gorgol



Matematyczne miscellanea

Tom 4

Wybrane narzędzia modelowania matematycznego

Matematyczne miscellanea

Tom 4

Wybrane narzędzia modelowania matematycznego

Redakcja
Izolda Gorgol

WYDAWNICTWO
POLITECHNIKI
LUBELSKIEJ

Lublin 2024

R E C E N Z E N C I :

prof. dr hab. **Stepan Shakhno**, Narodowy Uniwersytet im. Iwana Franki we Lwowie, Ukraina

dr hab. inż. **Elżbieta Sidorowicz**, prof. uczelni, Uniwersytet Zielonogórski

R E D A K T O R :

dr **Izolda Gorgol**, ORCID: 0000-0003-4394-7387
Politechnika Lubelska – ROR: https://ror.org/024zjzd49

Redaktor prowadzący: Elżbieta Nazaruk
Korekta językowa: zespół redakcyjny
Skład i łamanie: Izolda Gorgol

Zdjęcie na okładce zostało stworzone przez Katarzynę Pełkę-Smętek z ilustracji znajdujących w monografii.

Publikacja wydana za zgodą **Rektora Politechniki Lubelskiej**

ISBN: 978-83-7947-613-8 (wersja drukowana)
ISBN: 978-83-7947-620-6 (wersja elektroniczna)
DOI: 10.35784/9788379476206
Wydawca: Wydawnictwo Politechniki Lubelskiej
www.wpl.pollub.pl
ul. Nadbystrzycka 36C, 20-618 Lublin
tel. (81) 538-46-59

WYDAWNICTWO
POLITECHNIKI
LUBELSKIEJ

Nakład: 50 egz.

Spis treści

Przedmowa

Współczesna matematyka, będąca fundamentem wielu dziedzin nauki i technologii, odgrywa kluczową rolę w rozwiązywaniu problemów, zarówno teoretycznych, jak i praktycznych. Wśród jej różnorodnych gałęzi, teoria grafów i teoria prawdopodobieństwa zajmują szczególne miejsce, oferując narzędzia i metody przydatne do analizy skomplikowanych systemów i procesów. Niniejsza monografia, będąca kolejnym, czwartym tomem serii Matematyczne Miscellanea, ma na celu przedstawienie kilku nowoczesnych narzędzi z tych dwóch dynamicznie rozwijających się dziedzin oraz ukazanie ich zastosowań w różnych obszarach.

Teoria grafów umożliwia modelowanie i analizę relacji oraz powiązań między różnymi obiektami. Dzięki swojej uniwersalności, znalazła ona szerokie zastosowanie w informatyce, biologii, logistyce, a także w analizie sieci społecznych i telekomunikacyjnych. Grafy pozwalają na reprezentowanie i zrozumienie złożonych struktur, takich jak sieci komputerowe, systemy transportowe czy interakcje białek w komórkach. Do modelowania jeszcze bardziej złożonych zależności i relacji używa się hipergrafów, które są uogólnieniem klasycznych grafów.

Z kolei teoria prawdopodobieństwa dostarcza narzędzi do modelowania i analizy zjawisk losowych, co jest nieodzowne w kontekście niepewności i zmienności rzeczywistości. Jej zastosowania obejmują szeroki wachlarz dziedzin, od statystyki i ekonomii, przez fizykę i inżynierię, aż po biologię i medycynę. Dzięki teorii prawdopodobieństwa możliwe jest przewidywanie przyszłych zdarzeń, ocena ryzyka oraz optymalizacja procesów w warunkach niepewności. Teoria ta pozwala prognozować niezawodności urządzeń technicznych, rozwój zjawisk gospodarczych, w tym przewidywać zachowania rynków finansowych, modelować oczekiwane straty i ryzyka kredytowe.

W pierwszym rozdziale zaprezentowano hipergrafy mieszane, jedną z nowszych struktur zdefiniowaną w celu modelowania złożonych zależności między obiektami. Po krótkiej prezentacji pojęć dotyczących klasycznych grafów i hipergrafów, zdefiniowano hipergraf mieszany, jego kolorowanie oraz przedstawiono dwa podejścia do tak zdefiniowanego kolorowania. Jednym z nich jest algorytm rozdzielania-ściągania, a drugim użycie programowania całkowitoliczbowego. W dalszej części rozdziału przedstawiono konkretne wykorzystanie kolorowania hipergrafu mieszanego przy wykonywaniu zapytań do bazy danych.

Drugi rozdział zawiera krótki opis czterech algorytmów grafowych, od najbardziej klasycznego algorytmu Dijkstry, przez algorytm genetyczny, po najnowsze specjalistyczne algorytmy cięcia pętla po pętli i wiszących ogrodów. Opisano, jak

można wykorzystać te algorytmy w modelowaniu zintegrowanego systemu miejskiej gospodarki wodnej.

Trzeci rozdział dotyczy zastosowania narzędzi teorii grafów do opracowania procesu realizacji różnego rodzaju projektów. Przedstawiono trzy metody grafowe stosowane w tym kontekście, są to: CPM (*critical path method*), PERT (*program evaluation and review technique*) oraz GERT (*graphical evaluation and review technique*). Wspólną cechą tych metod jest podział projektu na wiele dobrze zdefiniowanych i nienakładających się na siebie pojedynczych zadań oraz szeregowania ich w odpowiedniej kolejności z uwzględnieniem czasów ich wykonania. Najprostszą z nich jest najbardziej znana deterministyczna metoda CPM, w której z góry określona jest struktura całego przedsięwzięcia. Bardziej ogólna jest metoda PERT, w której uwzględnia się możliwość losowej zmiany czasu wykonania zadań w pewnym zakresie. Najbardziej zaawansowana jest metoda GERT, gdyż pozwala również na dynamiczną zmianę struktury w zależności od losowo występujących czynników. Metody te porównano na przykładzie opracowania harmonogramu realizacji konkretnego projektu budowlanego.

Czwarty rozdział stanowi wprowadzenie do przestrzeni podliniowej wartości oczekiwanej. Jest to nowe narzędzie powstałe w odpowiedzi na fakt, że klasyczne modele prawdopodobieństwa mogą okazać się niewystarczające w warunkach niepewności prawdopodobieństwa. Wiele zjawisk zachodzi w takich warunkach, a modele budowane przy użyciu liniowej i addytywnej wartości oczekiwanej ich nie uwzględniają. Prawdopodobieństwa nieaddytywne i nieaddytywne wartości oczekiwane lepiej nadają się do opisu takich zjawisk. W ostatnich latach teoria i metodologia nieliniowości zostały dobrze rozwinięte i spotkały się z dużym zainteresowaniem w niektórych dziedzinach, takich jak finanse czy pomiar i kontrola ryzyka. Model matematyczny podliniowej przestrzeni wartości oczekiwanej wprowadził w 2010 roku do literatury Peng jako naturalne rozszerzenie klasycznej liniowej przestrzeni, dostarczając tym samym nowych narzędzi do modelowania zjawisk zachodzących w warunkach niepewności. Konstrukcja tej przestrzeni opiera się na założeniu, że wartość oczekiwana E nie jest funkcjonałem liniowym, lecz podliniowym. Konsekwencją tego rozszerzenia jest konieczność przedefiniowania podstawowych pojęć, znanych z klasycznej przestrzeni prawdopodobieństwa, w taki sposób, aby spełniały one równoważne role w nowej rozszerzonej przestrzeni. Należą do nich między innymi: wartość średnia, wariancja, rozkład zmiennej losowej, prawdopodobieństwo, niezależność zmiennych losowych, ale także nowa wersja prawa wielkich liczb czy centralnego twierdzenia granicznego. Pojęcia takie jak mocna i słaba zbieżność zyskują w niej nowe interpretacje.

W piątym rozdziale szeroko opisano ciągłą procedurę aproksymacji stochastycznej w środowisku półmarkowowskim. Procedura aproksymacji stochastycznej (PAS)

stanowi podstawę teoretyczną optymalizacji konstrukcji ciągu zbieżnego do punktu równowagi w warunkach skomplikowanych wpływów środowiska zewnętrznego na funkcję regresji albo wpływów na układ stochastyczny. Głównym problemem dotyczącym PAS jest jej zbieżność przy wykorzystaniu jak najmniejszej ilości informacji o funkcji regresji oraz charakterystyk tej funkcji w punkcie obserwacji. Ponadto jednym z najważniejszych zadań ogólnej teorii układów ewolucyjnych w środowisku stochastycznym jest znalezienie warunków stabilności takich układów. W tym rozdziale zaprezentowano zastosowanie algorytmów fazowego uśrednienia procesów półmarkowowskich, a także algorytmów dyfuzyjnego przybliżenia stochastycznych ewolucji, które są wykorzystywane przez metody martyngałowe we współczesnej teorii procesów losowych.

Szósty, ostatni rozdział monografii prezentuje, w jaki sposób można połączyć teorię grafów i rachunek prawdopodobieństwa w celu opisu tworzenia się i rozwoju sieci złożonej o specjalnych własnościach. Prekursorami takiego podejścia byli Erdős i Rényi, węgierscy naukowcy, którzy zdefiniowali graf losowy. Jednak wiele sieci występujących w rzeczywistości nie może być opisanych z użyciem ich modelu. W rozdziale przedstawiono model Bianconi–Barabásiego, będący uogólnieniem bardziej znanego modelu Barabásiego–Albert. Zasada preferencyjnego dołączania wierzchołków została w nim rozszerzona o parametr różnorodności adaptacyjnej opisujący atrakcyjność węzła w sieci. Umożliwia to indywidualizację traktowania poszczególnych węzłów w kontekście ich oddziaływania na całą sieć.

Mamy nadzieję, że monografia ta stanie się źródłem wiedzy i inspiracji zarówno dla studentów, jak i dla naukowców oraz praktyków zainteresowanych zastosowaniami matematyki, w szczególności teorii grafów i prawdopodobieństwa, w różnych dziedzinach.

Redaktor

Emilia Popławska-Panas [1]

Kolorowanie hipergrafów mieszanych

Streszczenie

W rozdziale opisano wybrane pojęcia teorii grafów, teorii hipergrafów oraz najważniejsze własności hipergrafów mieszanych. Przedstawiono algorytm kolorowania hipergrafów mieszanych oraz modelowanie kolorowania hipergrafów mieszanych za pomocą programowania całkowitoliczbowego. Wyznaczono wielomian chromatyczny, dolną oraz górną liczbę chromatyczną dla przykładów hipergrafów mieszanych modelujących wybrane problemy.

Słowa kluczowe: *hipergrafy mieszane, kolorowanie hipergrafów mieszanych*

Abstract

In the chapter, selected terms from graph theory, hypergraph theory and the most important properties of mixed hypergraphs are described. The coloring algorithm for mixed hypergraphs and modelling of mixed hypergraphs coloring with the use of integer programming are introduced. The chromatic polynomial as well as the lower and upper chromatic numbers are calculated for examples of mixed hypergraphs which model selected problems.

Keywords: *mixed hypergraphs, mixed hypergraphs coloring*

Wstęp

Tematem niniejszego rozdziału są hipergrafy mieszane powstałe w latach 90. XX wieku. Są one mocno związane z kolorowaniem grafów. Mówiąc o hipergrafach mieszanych nie sposób nie wspomnieć o F. Guthrie, który jest twórcą hipotezy o czterech kolorach, która obecnie jest już twierdzeniem o czterech kolorach. Twierdzenie to dotyczy zagadnienia kolorowania dowolnej mapy geograficznej za pomocą czterech kolorów tak, aby kraje graniczące ze sobą nie posiadały tego samego koloru.

Rozwój badań na temat grafów i ich kolorowania pozwolił na wprowadzenie uogólnienia pojęcia grafu – hipergrafu. Najważniejsze pojęcia związane z teorią grafów oraz teorią hipergrafów zostały opisane w 1973 roku przez C. Berge w pracy o tytule *Grafy i hipergrafy* [2]. Ponad 20 lat później zdefiniowano pojęcie antykrawędzi (obecnie C-krawędzi) oraz hipergrafu mieszanego. Hipergrafy mieszane oprócz rodziny klasycznych hiperkrawędzi zawierają również rodzinę antykrawędzi. Podczas kolorowania hipergrafu mieszanego antykrawędzie mają co najmniej dwa wierzchołki takich samych kolorów, zaś w hiperkrawędziach występują co najmniej

[1] Katedra Matematyki Stosowanej, Politechnika Lubelska

dwa wierzchołki różnych kolorów. Wprowadzenie antykrawędzi, a w konsekwencji hipergrafu mieszanego pozwoliło odkryć narzędzie do modelowania sytuacji, w których pojawiają są dwa kontrastowe względem siebie ograniczenia.

1. Wprowadzenie do teorii grafów i hipergrafów

1.1. Podstawowe pojęcia dotyczące grafów

Przedstawimy podstawowe pojęcia związane z grafami oraz hipergrafami, które są niezbędne do zrozumienia tematyki hipergrafów mieszanych i ich kolorowania.

Definicja 1.1. [6, 9] Grafem prostym G nazywamy uporządkowaną parę $G = (V,E)$, gdzie niepusty, skończony zbiór $V = \{v_1, v_2, \cdots, v_n\}$ nazywamy zbiorem wierzchołków grafu G, a zbiór E dwuelementowych podzbiorów zbioru V nazywamy zbiorem krawędzi.

W celu wskazania danego grafu G, do którego odnoszą się zbiory V oraz E będziemy używać oznaczeń $V(G)$ oraz $E(G)$.

Definicja 1.2. [9] Wierzchołki v_i i v_j nazywamy sąsiednimi, jeżeli istnieje krawędź, która je łączy.

Można także powiedzieć, że wierzchołki te są incydentne z tą krawędzią.

Definicja 1.3. [9] Grafem pełnym K_n o n-wierzchołkach nazywamy graf, w którym każde dwa wierzchołki są sąsiednie.

Definicja 1.4. [9] Ścieżką, która łączy wierzchołek v_0 z wierzchołkiem v_n nazywamy ciąg parami różnych wierzchołków $(v_0, v_1, \ldots, v_n)$, jeżeli dla każdego $i \in \{0, 1, \ldots, n-1\}$ istnieje krawędź, a wierzchołki v_i oraz v_{i+1} są końcami tej krawędzi.

Definicja 1.5. [9] Grafem spójnym nazywamy taki graf, w którym za pomocą ścieżki można połączyć ze sobą dwa dowolne wierzchołki.

Definicja 1.6. [9] Składową spójności grafu G nazywamy maksymalny podgraf spójny grafu G.

Definicja 1.7. [9] Cyklem nazywamy graf, złożony ze ścieżki $(v_0, v_1, \ldots, v_n)$, w którym dodatkowo istnieje krawędź łącząca wierzchołki v_0 i v_n.

Definicja 1.8. [9] Lasem nazywamy graf $G = (V,E)$, który nie posiada cykli.

Definicja 1.9. [9] Drzewem nazywamy graf spójny $T = (V,E)$, który nie posiada cykli.

W lesie składowymi spójności są drzewa.

Definicja 1.10. [9] Grafem dwudzielnym nazywamy graf $G = (V,E)$, w którym zbiór wierzchołków $V(G)$ może być podzielony na dwa zbiory rozłączne w taki sposób, aby każda krawędź G miała jeden wierzchołek ze zbioru A oraz jeden wierzchołek ze zbioru B.

Przykładem grafu dwudzielnego jest graf z rysunku 6(a), gdzie

$$A = \{v_1, v_2.v_3, v_4, v_5, v_6\} \text{ oraz } B = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}.$$

1.2. Kolorowanie i wielomian chromatyczny wybranych grafów

Definicja 1.11. [12] Właściwym (poprawnym) λ-kolorowaniem grafu G nazywamy odwzorowanie

$$c\colon V(G) \to \{1,2,\ldots,\lambda\}$$

spełniające warunek $c(v_i) \neq c(v_j)$ dla dowolnych, sąsiednich wierzchołków grafu G.

Elementy zbioru $\{1,2,\ldots,\lambda\}$ są nazywane kolorami lub barwami. Definicja 1.11 oznacza, że wierzchołki danej krawędzi nie mogą mieć takiego samego koloru.

Definicja 1.12. [12] Liczbą chromatyczną $\chi(G)$ dla grafu G nazywamy najmniejszą liczbę kolorów potrzebną do właściwego pokolorowania grafu.

Z terminem kolorowania grafu związane jest pojęcie wielomianu chromatycznego. Wielomian chromatyczny grafu został wprowadzony przez Birkhoffa w 1912 roku i od tamtej pory jest przedmiotem badań [4].

Definicja 1.13. [4, 12] Wielomianem chromatycznym $P(G,\lambda)$ grafu G nazywamy wielomian zmiennej λ określający liczbę λ-kolorowań grafu G.

Twierdzenie 1.1. [8] *Wielomian chromatyczny grafu pełnego K_n o n wierzchołkach jest postaci*

$$P(K_n,\lambda) = \lambda \cdot \big(\lambda - 1\big) \cdot \ldots \cdot \big(\lambda - n + 1\big) = \lambda^{(n)}. \tag{1.1}$$

Dowód. Załóżmy, że graf K_n jest grafem pełnym o n wierzchołkach. Wybieramy dowolny wierzchołek i kolorujemy go. Możemy to zrobić na λ sposobów. Następnie wybieramy kolejny wierzchołek i kolorujemy go. Teraz mamy $\lambda - 1$ możliwości, ponieważ wybrany wierzchołek jest sąsiedni do wierzchołka pierwszego. W analogiczny sposób kolorujemy pozostałe wierzchołki. Zauważmy, że n-ty wierzchołek możemy pokolorować na $\lambda - (n-1)$ sposobów, ponieważ jest on sąsiedni do pokolorowanych $(n-1)$ wierzchołków. Stąd otrzymujemy, że

$$P(K_n,\lambda) = \lambda \cdot \big(\lambda - 1\big) \cdot \ldots \cdot \big(\lambda - n + 1\big).$$

□

Twierdzenie 1.2. [12] *Wielomian chromatyczny grafu G ma następującą postać*

$$P(G,\lambda) = \sum_{i=\chi(G}^{n} r_i(G) \cdot \lambda^{(i)}, \tag{1.2}$$

gdzie $r_i(G)$ jest liczbą możliwych kolorowań grafu G przy użyciu i kolorów (nie licząc permutacji kolorów), zaś $\lambda^{(i)}$ jest wielomianem chromatycznym grafu pełnego K_i.

Twierdzenie 1.3. [8] *Wielomian chromatyczny dowolnego drzewa T_n o n wierzchołkach jest postaci*

$$P(T_n,\lambda) = \lambda \cdot (\lambda-1)^{n-1}. \tag{1.3}$$

Dowód. Załóżmy, że graf T_n jest drzewem o n wierzchołkach. Jako pierwszy wybieramy wierzchołek, który jest incydentny tylko z jedną krawędzią. Możemy go pokolorować na λ sposobów. Wybieramy drugi wierzchołek, który jest połączony krawędzią z wierzchołkiem pierwszym. Drugi wierzchołek możemy pokolorować na $(\lambda-1)$ sposobów. W kolejnych krokach wybieramy wierzchołek sąsiedni do wierzchołka już pokolorowanego, a następnie kolorujemy go, mając $(\lambda-1)$ możliwości. Zatem

$$P(T_n,\lambda) = \lambda \cdot (\lambda-1)^{n-1}.$$

□

Twierdzenie 1.4. [8] *Jeżeli graf G ma składowe spójności $G_1, G_2, \ldots, G_n$, to*

$$P(G,\lambda) = P(G_1,\lambda) \cdot P(G_2,\lambda) \cdot \ldots \cdot P(G_n,\lambda). \tag{1.4}$$

Dowód. Załóżmy, że graf G ma składowe spójności $G_1, G_2, \ldots, G_n$. Kolorowania składowych spójności są niezależne od siebie, ponieważ składowe są rozłączne. Zatem kolorowanie grafu G jest iloczynem kolorowań jego składowych spójności. Stąd

$$P(G,\lambda) = P(G_1,\lambda) \cdot P(G_2,\lambda) \cdot \ldots \cdot P(G_n,\lambda).$$

□

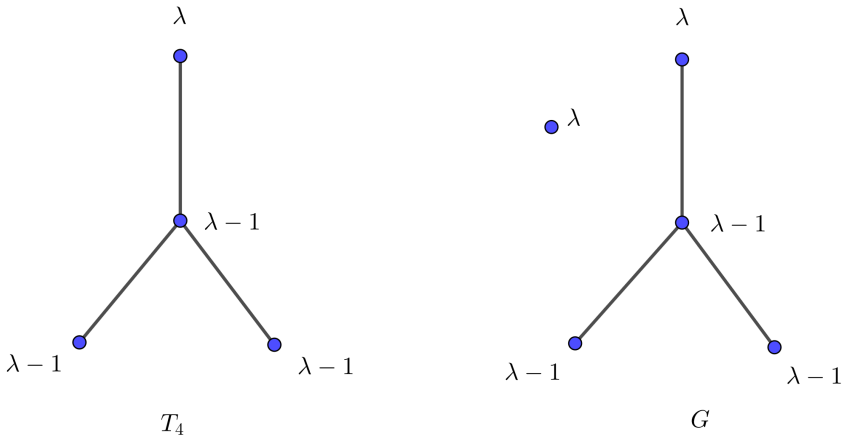
Na mocy twierdzenia 1.4 wielomian chromatyczny lasu jest zatem iloczynem wielomianów chromatycznych drzew wchodzących w jego skład.

Przykład 1.1. Na podstawie twierdzenia 1.3 wielomian chromatyczny drzewa T_4 zamieszczonego na rysunku 1(a) jest postaci

$$P(T_4,\lambda) = \lambda \cdot (\lambda-1)^3 = \lambda^4 - 3\lambda^3 + 3\lambda^2 - \lambda.$$

Graf G przedstawiony na rysunku 1(b) jest lasem, a jego wielomian chromatyczny na mocy twierdzeń 1.4, 1.3 ma następującą postać

$$P(G,\lambda) = \lambda \cdot \lambda \cdot (\lambda - 1)^3 = \lambda^5 - 3\lambda^4 + \lambda^3 - \lambda^2.$$



(a) Drzewo T_4 (b) Las G

Rysunek 1. Przykład drzewa T_4 oraz lasu G
Źródło: Opracowanie własne

1.3. Podstawowe pojęcia dotyczące hipergrafów

Definicja 1.14. [1,3,6] Hipergrafem H nazywamy uporządkowaną parę $H = (V, \mathscr{E})$, gdzie V jest zbiorem wierzchołków, a $\mathscr{E}$ jest rodziną podzbiorów zbioru V.

Hiperkrawędzie to elementy $E_1, E_2, ..., E_m$ rodziny $\mathscr{E}$.

Podobnie jak w przypadku grafów będziemy stosowali oznaczenia $V(H)$ oraz $E(H)$ w celu wskazania konkretnego hipergrafu H.

Definicja 1.15. [6] k-krawędzią nazywamy krawędź zawierającą k wierzchołków.

Definicja 1.16. [10] Rzędem hipergrafu H nazywamy liczbę jego wierzchołków $n(H) = |V(H)|$.

Definicja 1.17. [10] Rozmiarem hipergrafu H nazywamy liczbę jego krawędzi $m(H) = |\mathscr{E}(H)|$.

Definicja 1.18. [6] Podhipergrafem hipergrafu H_1 nazywamy hipergraf H_2, jeśli spełnione są warunki $V(H_1) \subseteq V(H_2)$ oraz $E(H_1) \subseteq E(H_2)$.

Definicja 1.19. [6] Podhipergrafem indukowanym krawędziowo hipergrafu H_1 nazywamy hipergraf H_2, jeżeli krawędzie ze zbioru $E(H_2)$ należą również do zbioru $E(H_1)$ oraz zbiór wierzchołków $V(H_2)$ zawiera tylko wierzchołki incydentne ze zbiorem krawędzi $E(H_2)$.

W podobny sposób możemy zdefinować podhipergraf indukowany wierzchołkowo. Możemy również mówić o podhipergrafie hipergrafu H indukowanym przez zbiór S, gdzie S jest zbiorem wierzchołków lub zbiorem krawędzi. Taki hipergraf oznaczamy przez $H[S]$.

2. Hipergrafy mieszane – własności i kolorowanie

2.1. Podstawowe pojęcia dotyczące hipergrafów mieszanych

Niech $V = \{v_1, v_2, \ldots, v_n\}$ dla $n \geqslant 1$ będzie dowolnym zbiorem skończonym. Niech $\mathscr{C} = \{C_1, C_2, \ldots, C_l\}$ oraz $\mathscr{D} = \{D_1, D_2, \ldots, D_m\}$, gdzie $I = \{1, 2, \ldots, l\}$ oraz $J = \{1, 2, \ldots, m\}$ będą rodzinami podzbiorów zbioru V takimi, że dla każdego z elementów rodziny $\mathscr{C}$ oraz $\mathscr{D}$ zachodzi warunek $|C \cup D| \geqslant 2$.

Definicja 2.1. [5] Hipergrafem mieszanym nazywamy trójkę $(V, \mathscr{C}, \mathscr{D})$, gdzie zbiór V jest zbiorem wierzchołków, $\mathscr{C}$ oraz $\mathscr{D}$ są rodzinami podzbiorów zbioru V nazywanych odpowiednio C-krawędziami oraz D-krawędziami.

Będziemy używali następujących oznaczeń:

— $V(H)$ – zbiór wierzchołków hipergrafu mieszanego H,
— $\mathscr{C}(H)$ – C-krawędzie hipergrafu mieszanego H,
— $\mathscr{D}(H)$ – D-krawędzie hipergrafu mieszanego H.

Niech zbiór $\{1, 2, \ldots, \lambda\}$ będzie zbiorem kolorów, gdzie $\lambda \in \mathbb{N}$.

Definicja 2.2. [5, 14] Właściwym λ-kolorowaniem hipergrafu mieszanego $H = (V, \mathscr{C}, \mathscr{D})$ nazywamy odwzorowanie $c\colon V \to \{1, 2, \ldots, \lambda\}$ spełniające warunki:

— każda C-krawędź ma co najmniej dwa wierzchołki tego samego koloru,
— każda D-krawędź ma co najmniej dwa wierzchołki różnych kolorów.

Hipergraf mieszany jest kolorowalny, jeśli ma co najmniej jedno kolorowanie, w przeciwnym razie jest niekolorowalny.

Z liczbą kolorów wykorzystanych podczas kolorowania mają związek pojęcia dolnej liczby chromatycznej i górnej liczby chromatycznej.

Definicja 2.3. [5] Dolną liczbą chromatyczną (ang. *lower chromatic number*) $\chi(H)$ hipergrafu mieszanego H nazywamy minimalną wartość λ, dla której istnieje właściwe λ-kolorowanie hipergrafu.

We właściwym λ-kolorowaniu nie wszystkich λ kolorów należy użyć. O ścisłym λ-kolorowaniu będziemy mówili, gdy wykorzystamy dokładnie λ kolorów.

Definicja 2.4. [5] Ścisłym i-kolorowaniem hipergrafu mieszanego H nazywamy takie właściwe i-kolorowanie, w którym użyto wszystkich i kolorów, gdzie $i \in \mathbb{N}$ oraz $1 \leqslant i \leqslant \lambda$.

Definicja 2.5. [5] Górną liczbą chromatyczną (ang. *upper chromatic number*) $\bar{\chi}(H)$ hipergrafu mieszanego H nazywamy największą liczbę i, dla której istnieje ścisłe i-kolorowanie.

Niech $Y \subset V$. Zbiór Y nazywamy monochromatycznym, jeśli wierzchołki do niego należące są takiego samego koloru. Jeżeli dla $y \in Y$ kolory $c(y)$ są parami różne, to zbiór Y będziemy nazywali polichromatycznym.

Każde ścisłe i-kolorowanie hipergrafu mieszanego H prowadzi do podziału zbioru wierzchołków V na i niepustych monochromatycznych podzbiorów $V_1, V_2, \ldots, V_i$ zwanych klasami kolorów, w taki sposób, że każda C-krawędź ma przynajmniej dwa wierzchołki w tym samym podzbiorze oraz każda D-krawędź ma przynajmniej dwa wierzchołki znajdujące się w różnych podzbiorach. Taki podział nazywamy podziałem wykonalnym (ang. *feasible partition*) hipergrafu mieszanego H.

Definicja 2.6. [5] Spektrum chromatycznym hipergrafu mieszanego H nazywamy wektor liczb całkowitych $R(H) = (r_1, r_2, \ldots, r_n)$, gdzie dla $k \in \{1, \ldots, n\}$ r_k jest liczbą możliwych k-kolorowań. Kolorowanie, które dzieli zbiór wierzchołków na te same podzbiory, traktujemy jako jedno kolorowanie.

Każdy kolorowalny hipergraf mieszany posiada dolną i górną liczbę chromatyczną, więc spektrum chromatyczne ma postać

$$R(H) = (0, \ldots, 0, r_\chi, \ldots, r_{\bar{\chi}}, 0, \ldots, 0). \tag{2.1}$$

Każdy podział wykonalny na i klas kolorów określa i ścisłych i-kolorowań różniących się między sobą ze względu na permutację kolorów. Stąd liczba ścisłych i-kolorowań jest równa $r_i \cdot i!$. Jeśli mamy $\lambda > i$ kolorów, to otrzymujemy $\binom{\lambda}{i}$ sposobów wyboru podzbioru i kolorów. Zatem liczba właściwych λ-kolorowań generowanych przez wszystkie możliwe podziały na i podzbiorów wynosi

$$\binom{\lambda}{i} r_i \cdot i! = \frac{\lambda \cdot (\lambda-1) \cdot (\lambda-2) \ldots (\lambda-i+1) \cdot (\lambda-i)!}{i! \cdot (\lambda-i)!} r_i \cdot i! = \lambda^{(i)} \cdot r_i.$$

Definicja 2.7. [5,12] Wielomianem chromatycznym $P(H,\lambda)$ hipergrafu mieszanego H nazywamy wielomian zmiennej λ określający liczbę właściwych λ-kolorowań hipergrafu mieszanego H. Ma on następującą postać

$$P(H,\lambda) = \sum_{i=\chi(H)}^{\bar{\chi}(H)} r_i(H) \cdot \lambda^{(i)}. \tag{2.2}$$

Wielomian chromatyczny (2.2) dla hipergrafu mieszanego jest uogólnieniem wielomianu chromatycznego (1.2) dla grafu.

Dla kolorowalnych hipergrafów mieszanych maksymalny stopień wielomianu chromatycznego jest równy górnej liczbie chromatycznej $\bar{\chi}$, zaś współczynnik przy najwyższej potędze wielomianu chromatycznego wynosi $r_{\bar{\chi}}$. Dla niekolorowalnego hipergrafu mieszanego H mamy

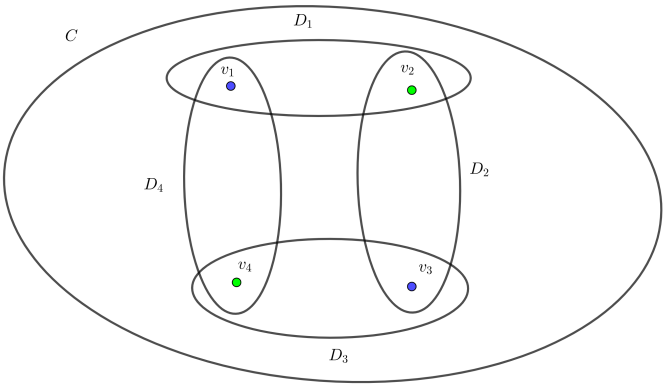
$$\chi(H) = 0, \quad \bar{\chi}(H) = 0, \quad R(H) = (0,0,\ldots 0), \quad P(H,\lambda) = 0.$$

Definicja 2.8. [5,12] Średnią liczbą chromatyczną hipergrafu mieszanego H (ang. *middle chromatic number*) nazywamy wielkość $\chi_m(H) = \frac{\bar{\chi}(H)+\chi(H)}{2}$, zaś wartość $b(H) = \bar{\chi}(H) - \chi(H) + 1$ nazywamy szerokością spektrum chromatycznego (ang. *breadth of the chromatic spectrum*).

Przykład 2.1. Rozważmy hipergraf mieszany $H_1 = (V, \mathscr{C}, \mathscr{D})$, gdzie

$$V(H_1) = \{v_1, v_2, v_3, v_4\},$$
$$\mathscr{C}(H_1) = \{\{v_1, v_2, v_3, v_4\}\},$$
$$\mathscr{D}(H_1) = \{D_1, D_2, D_3, D_4\} = \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_4, v_1\}\}.$$

Hipergraf mieszany H_1 oraz jego właściwe kolorowanie został przedstawiony na rysunku 2. Zauważmy, że krawędź D_1 ma wierzchołek v_1 w kolorze niebieskim oraz wierzchołek v_2 w kolorze zielonym. Krawędź D_2 ma wierzchołek v_2 w kolorze zielonym oraz wierzchołek v_3 w kolorze niebieskim. Krawędź D_3 ma niebieski wierzchołek v_3 oraz zielony wierzchołek v_4, a krawędź D_4 posiada zielony wierzchołek v_4 oraz niebieski wierzchołek v_1. C-krawędź ma dwa zielone wierzchołki v_2 i v_4 oraz dwa niebieskie v_1 i v_3. Każda z D-krawędzi ma co najmniej dwa wierzchołki o różnych kolorach, zaś krawędź C ma co najmniej dwa wierzchołki w tym samym kolorze, więc zostało znalezione właściwe kolorowanie. Jest to kolorowanie optymalne, ponieważ została znaleziona minimalna liczba użytych kolorów. Zatem dolna liczba chromatyczna wynosi $\chi(H_1) = 2$. Opisane kolorowanie jest ścisłym 2-kolorowaniem, ponieważ obydwa kolory zostały użyte. Na rysunku 3 zostało przedstawione ścisłe 3-kolorowanie hipergrafu mieszanego H_1. Wierzchołek v_4 został pokolorowany na czerwono.



Rysunek 2. Optymalne kolorowanie hipergrafu mieszanego H_1
Źródło: Opracowanie własne

Kolorowanie nadal pozostaje właściwe, ale nie jest już optymalne. Górna liczba chromatyczna hipergrafu mieszanego H_1 wynosi $\bar{\chi}(H_1) = 3$. Nie istnieje właściwe 4-kolorowanie hipergrafu mieszanego H_1. Gdybyśmy np. wierzchołek v_3 pokolorowali nowym kolorem, to założenie, że C-krawędź ma co najmniej dwa wierzchołki o takim samym kolorze nie byłoby spełnione.

2-kolorowanie dzieli zbiór wierzchołków na dwa podzbiory

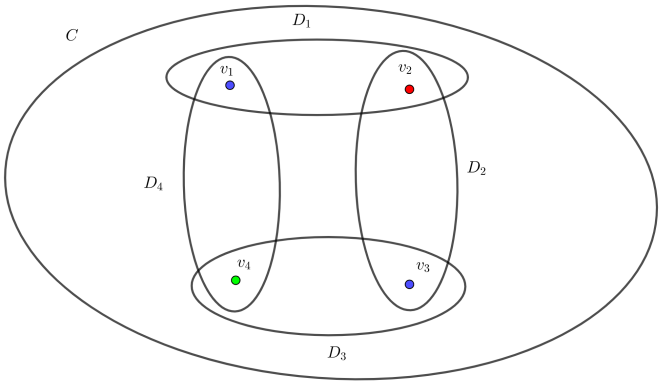
$$V_1 = \{v_1, v_3\},\ V_2 = \{v_2, v_4\}.$$

Kolorowanie przy użyciu dokładnie trzech kolorów może podzielić zbiór wierzchołków na podzbiory

$$V_1 = \{v_1, v_3\},\ V_2 = \{v_2\},\ V_3 = \{v_4\} \quad \text{lub} \quad V_1 = \{v_1\},\ V_2 = \{v_3\},\ V_3 = \{v_2, v_4\}.$$

Z tego powodu istnieje dokładnie jedno ścisłe 2-kolorowanie oraz dwa ścisłe 3-kolorowania, a spektrum chromatyczne przyjmuje postać $R(H_1) = (0,1,2,0)$. Średnia liczba chromatyczna wynosi $\chi_m(H) = \frac{2+3}{2} = 2.5$, zaś szerokość spektrum chromatycznego jest równa $b(H) = 3-2+1 = 2$.

Definicja 2.9. [12] Dla dowolnego hipergrafu mieszanego $H = (V, \mathscr{C}, \mathscr{D})$ D-hipergrafem H_D nazywamy częściowy podhipergraf mieszany (ang. *partial mixed subhypergraph*) taki, że $H_D = (V, \emptyset, \mathscr{D}) = (V, \mathscr{D})$.



Rysunek 3. 3-kolorowanie hipergrafu mieszanego H_1
Źródło: Opracowanie własne

Definicja 2.10. [12] Dla dowolnego hipergrafu mieszanego $H = (V, \mathscr{C}, \mathscr{D})$ C-hipergrafem H_C nazywamy częściowy podhipergraf mieszany (ang. *partial mixed subhypergraph*) taki, że $H_C = (V, \mathscr{C}, \emptyset) = (V, \mathscr{C})$.

Zauważmy, że klasyczny graf to D-hipergraf z krawędziami o rozmiarze 2.

Definicja 2.11. [12] Odwrotnością chromatyczną hipergrafu mieszanego $H = (V, \mathscr{C}, \mathscr{D})$ nazywamy hipergraf mieszany $\bar{H} = (V, \mathscr{C}_1, \mathscr{D}_1)$, jeśli $\mathscr{C}_1 = \mathscr{D}$ oraz $\mathscr{D}_1 = \mathscr{C}$. Podczas inwersji problem znalezienia dolnej liczby chromatycznej $\chi(H)$ „odwraca się" w problem znalezienia górnej liczby chromatycznej $\bar{\chi}(\bar{H})$ i odwrotnie. Problemy te są kombinatorycznie dualne względem siebie.

Dla hipergrafu mieszanego $H = (V, \mathscr{C}, \mathscr{D})$ oznaczmy przez $\mathscr{E}$ rodzinę $\mathscr{E} = \mathscr{C} \cup \mathscr{D}$. Przez hipergraf pierwotny hipergrafu mieszanego będziemy rozumieli hipergraf $H' = (V, \mathscr{E})$. Dowolna ścieżka zawarta w hipergrafie pierwotnym H' jest mieszaną ścieżką w hipergrafie H. C-ścieżką (D-ścieżką) nazywamy ścieżkę zawartą w C-hipergrafie (D-hipergrafie).

Poniżej zaprezentowaliśmy kilka własności właściwego kolorowania hipergrafu mieszanego:

— Każda permutacja kolorów tworzy nowe właściwe kolorowanie.
— D-krawędzie nie są zbiorami monochromatycznymi.
— C-krawędzie nie są zbiorami polichromatycznymi.

— Jeżeli $\mathscr{C} = \emptyset$, to kolorowanie hipergrafu mieszanego $H = (V, \emptyset, \mathscr{D})$ jest kolorowaniem hipergrafu.
— Dla dowolnego hipergrafu mieszanego H częściowe hipergrafy mieszane są kolorowalne.
— Trywialnym kolorowaniem C-hipergrafu jest wykorzystanie jednego koloru dla wszystkich wierzchołków.
— Trywialnym kolorowaniem D-hipergrafu jest wykorzystanie $n(H_D)$ kolorów.
— Dla dowolnego hipergrafu mieszanego H prawdziwe są następujące zależności

$$\chi(H_C) = 1,\ r_1 = 1,\ \bar{\chi}(H_D) = n,\ r_n(H_D) = 1.$$

— Dla kolorowalnego hipergrafu mieszanego H prawdziwa jest nierówność

$$1 \leqslant \chi(H_D) \leqslant \chi(H) \leqslant \bar{\chi}(H) \leqslant \bar{\chi}(H_C) \leqslant n.$$

Jeśli dla hipergrafu mieszanego H nie zachodzi nierówność

$$\chi(H_D) \leqslant \bar{\chi}(H_C), \tag{2.3}$$

to hipergraf mieszany nie jest kolorowalny. Nierówność 2.3 nazywa się pierwszym warunkiem kolorowalności hipergrafów mieszanych.

2.2. Hipergrafy mieszane jednoznacznie kolorowalne

Definicja 2.12. [1,12,13] Hipergrafem jednoznacznie kolorowalnym (ang. *uniquely colorable*, w skrócie uc) nazywamy hipergraf mieszany H, jeśli ma on dokładnie jedno ścisłe kolorowanie, nie licząc permutacji kolorów.

Hipergraf mieszany jest jednoznacznie kolorowalny, jeśli istnieje dokładnie jeden podział jego zbioru wierzchołków na klasy kolorów. Zauważmy, że „jednoznaczne kolorowanie" oznacza „jednoznaczny podział" na odpowiednią liczbę klas kolorów.

Warto zwrócić uwagę, że jeśli hipergraf mieszany H jest uc hipergrafem, to

$$\chi(H) = \bar{\chi}(H) = \chi\ ,\ r_\chi = 1 \text{ oraz } R(H) = (0,\ldots,1,\ldots,0).$$

Zatem wielomian chromatyczny, zgodnie z (2.2), ma postać

$$P(H,\lambda) = \sum_{i=\chi(H)}^{\bar{\chi}(H)} r_i(H) \cdot \lambda^{(i)} = r_\chi(H) \cdot \lambda^{(\chi)} = \lambda \cdot (\lambda - 1) \cdot \ldots \cdot (\lambda - \chi + 1). \tag{2.4}$$

2.3. Algorytm rozdzielania-ściągania

Do wyznaczania kolorowania hipergrafu mieszanego, dolnej liczby chromatycznej, górnej liczby chromatycznej, spektrum chromatycznego oraz wielomianu chromatycznego stosujemy algorytm rozdzielania-ściągania (ang. *splitting-contraction*).

Podczas wyznaczania wielomianu chromatycznego $P(H,\lambda)$ oraz spektrum chromatycznego $R(H)$ hipergrafu mieszanego $H=(V,\mathscr{C},\mathscr{D})$, gdzie $\mathscr{C}=\{C_1,C_2,\ldots,C_l\}$, $\mathscr{D}=\{D_1,D_2,\ldots,D_m\}$, $I=\{1,2,\ldots,l\}$ oraz $J=\{1,2,\ldots,m\}$, stosujemy następujące zasady:

(a) **Zasada eliminacji**
Niech $V'\subseteq V$ będzie zbiorem takim, że $|V'|\geqslant 2$. Jeśli $\{v,u\}\in\mathscr{D}$ dla każdej pary wierzchołków $u,v\in V'$ oraz $V'=C\in\mathscr{C}$, to H jest niekolorowalny. Podobnie, jeśli dla każdej pary wierzchołków $v,u\in V'$ połączonych C-ścieżką zawierającą C-krawędzie o rozmiarze 2 oraz $V'=D\in\mathscr{D}$, to H jest niekolorowalny.

(b) **Zasada C-usuwania** (ang. *C-clearing*)
Jeśli $C_i\subseteq C_j$, to $P(H,\lambda)=P(H-C_j,\lambda)$ oraz $R(H)=R(H-C_j)$, gdzie $i,j\in I$.

(c) **Zasada D-usuwania** (ang. *D-clearing*)
Jeśli $D_i\subseteq D_j$, to $P(H,\lambda)=P(H-D_j,\lambda)$ oraz $R(H)=R(H-D_j)$, gdzie $i,j\in J$.
W zasadach (b) oraz (c) usuwamy nadkrawędzie.

(d) **Zasada rozdzielania** (ang. *splitting*)
Jeśli $\{v_l,v_k\}\notin\mathscr{D}$ i $\{v_l,v_k\}\notin\mathscr{C}$, to

$$P(H,\lambda)=P(H_1,\lambda)+P(H_2,\lambda),\quad R(H)=R(H_1)+R(H_2),\ \text{gdzie}$$
$$H_1=(V,\mathscr{C},\mathscr{D}_1),\ \mathscr{D}_1=\mathscr{D}\cup\{v_l,v_k\},$$
$$H_2=(V,\mathscr{C}_1,\mathscr{D}),\ \mathscr{C}_1=\mathscr{C}\cup\{v_l,v_k\}.$$

(e) **Zasada ściągania** (ang. *contraction*)
Jeśli $C_t=\{v_l,v_k\}$, dla $t\in I$ i $v_k,v_l\in V$ oraz $C_t\neq D_s$ dla $s\in J$, to $P(H,\lambda)=P(H_1,\lambda)$, $R(H)=R(H_1)$, $H_1=(V_1,\mathscr{C}^1,\mathscr{D}^1)$, $V_1=(V\backslash\{v_l,v_k\}\cup\{u\})$, gdzie u jest nowym wierzchołkiem.
Jeśli $v_l\in D_j$ lub $v_k\in D_j$, $j\in J$, to $D_j^1=(D_j\backslash\{v_l,v_k\})\cup\{u\}$, w przeciwnym wypadku $D_j^1=D_j$.
Jeśli $v_l\in C_i$ lub $v_k\in C_i$, $i\in J$, $i\neq t$, to $C_i^1=(C_i\backslash\{v_l,v_k\})\cup\{u\}$, w przeciwnym wypadku $C_i^1=C_i$.

$$\mathscr{C}^1=\mathscr{C}-C_t.$$

Za pomocą wymienionych zasad algorytm sprowadza wielomian chromatyczny hipergrafu mieszanego do sumy wielomianów chromatycznych grafów pełnych. Otrzymane liczby grafów pełnych odpowiadają wartościom r_i (liczby właściwych kolorowań hipergrafu mieszanego H przy użyciu i kolorów z pominięciem permutacji kolorów) [12].

Algorytm rozdzielania-ściągania [12]

W algorytmie Y jest zbiorem pomocniczym, zbiór Z „zbiera" grafy pełne powstające podczas wykonywania kolejnych kroków, zaś zbiór L to lista ścisłych kolorowań hipergrafu mieszanego H.

WEJŚCIE: Na wejściu mamy dowolny hipergraf mieszany $H = (V, \mathscr{C}, \mathscr{D})$ ze zbiorem wierzchołków $V = \{v_1, v_2, \ldots, v_n\}$.

WYJŚCIE: Na wyjściu otrzymujemy listę L wszystkich ścisłych kolorowań, spektrum chromatyczne $R(H)$, wielomian chromatyczny $P(H,\lambda)$, dolną liczbę chromatyczną $\chi(H)$ oraz górną liczbę chromatyczną $\bar{\chi}(H)$.

1. Niech $L = Z = Y = \emptyset$, $R(H) = (0,0,\ldots,0)$, $P(H,\lambda) = 0$, $\chi(H) = 0$, $\bar{\chi}(H) = 0$. Dodajemy H do zbioru Y.
2. Sprawdzamy warunek eliminacji dla każdego elementu z Y. Usuwamy niekolorowalny hipergraf mieszany z Y.
3. Dla elementów z Y przeprowadzamy C-usuwanie i D-usuwanie, jeśli to możliwe.
4. Przeprowadzamy ściąganie, łącząc odpowiednie wierzchołki, jeśli to możliwe.
5. Wykonujemy rozdzielanie w każdym elemencie Y. Przenosimy otrzymane grafy pełne z Y do Z. Jeśli rozdzielanie zostało wykonane przynajmniej raz, przechodzimy do kroku 2.
6. Tworzymy listę L wszystkich ścisłych kolorowań, używając wierzchołków z otrzymanych grafów pełnych znajdujących się w Z.
7. Wyznaczamy spektrum chromatyczne $R(H)$, zliczając wszystkie grafy pełne mające dokładnie i wierzchołków, gdzie $i \in \{1,2,\ldots,n\}$.
8. Obliczamy wielomian chromatyczny, korzystając z zależności (2.2).
9. Wyznaczamy dolną liczbę chromatyczną $\chi(H)$ oraz górną liczbę chromatyczną $\bar{\chi}(H)$ w oparciu o spektrum chromatyczne $R(H)$.
10. Zwracamy listę ścisłych kolorowań L, wektor $R(H)$, wielomian chromatyczny $P(H,\lambda)$, liczby $\chi(H)$ oraz $\bar{\chi}(H)$.

Algorytm jest poprawny dla dowolnego hipergrafu mieszanego $(H, \mathscr{C}, \mathscr{D})$. Poprawność algorytmu wynika z zasad eliminacji, C-usuwania, D-usuwania, rozdzielania oraz ściągania. Każdemu wierzchołkowi grafu pełnego znajdującego się w zbiorze Z odpowiada monochromatyczny podzbiór odpowiednich wierzchołków.

2.4. Kolorowanie hipergrafu mieszanego a programowanie całkowitoliczbowe

Problem kolorowania hipergrafu mieszanego można modelować za pomocą programowania liniowego całkowitoliczbowego.

Niech dany będzie hipergraf mieszany $H = (V, \mathscr{C}, \mathscr{D})$, gdzie $V = \{v_1, v_2, \ldots, v_n\}$, $\mathscr{C} = \{C_1, C_2, \ldots, C_l\}$, $\mathscr{D} = \{D_1, D_2, \ldots, D_m\}$ dla $n, m, l \geqslant 1$. Kolory będziemy oznaczać liczbami ze zbioru $\{1, \ldots, n\}$. Niech $A = [\alpha_{ij}]$ będzie macierzą, w której

$$\alpha_{ij} = \begin{cases} 1, & \text{jeśli wierzchołek } v_i \text{ jest pokolorowany } j\text{-tym kolorem,} \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Warunkiem tego, aby każdy wierzchołek otrzymał dokładnie jeden kolor jest

$$\sum_{j=1}^{n} \alpha_{ij} = 1, \text{ gdzie } i \in \{1, 2, \ldots, n\}. \tag{2.5}$$

W celu właściwego kolorowania D-krawędzi potrzebujemy wprowadzić ograniczenia dane warunkiem

$$\forall_{D_k \in \mathscr{D}} \sum_{i \in D_k} \alpha_{ij} \leqslant |D_k| - 1, \text{ gdzie } j \in \{1, 2, \ldots, n\}. \tag{2.6}$$

Lewa strona nierówności w warunku (2.6) określa liczbę wierzchołków incydentnych z krawędzią D_k, które zostały pokolorowane j-tym kolorem. Natomiast prawa strona tej nierówności zapewnia, że co najmniej dwa wierzchołki krawędzi D_k mają różne kolory.

Dla właściwego kolorowania C-krawędzi wprowadzamy ograniczenia dane warunkiem

$$\forall_{C_k \in \mathscr{C}} \sum_{j=1}^{n} \max_{i \in C_k} \alpha_{ij} \leqslant |C_k| - 1. \tag{2.7}$$

Lewa strona nierówności (2.7) jest równa liczbie różnych kolorów użytych w C-krawędzi C_k, ponieważ

$$\max_{i \in C_k} \alpha_{ij} = \begin{cases} 1, & \text{jeśli } j\text{-ty kolor został użyty do kolorowania wierzchołka } v_i, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Prawa strona nierówności (2.7) gwarantuje, że co najmniej dwa wierzchołki każdej C-krawędzi mają ten sam kolor.

Zatem właściwe kolorowanie hipergrafu mieszanego H przy użyciu co najwyżej n kolorów może być modelowane za pomocą warunków (2.5). (2.6). (2.7), gdzie $\alpha_{ij} \in \{0, 1\}$.

Model programowania całkowitoliczbowego liniowego dla problemu znalezienia dolnej liczby chromatycznej $\chi(H)$ może być sformułowany następująco

$$\min z = \sum_{j=1}^{n} \max_{i \in V} \alpha_{ij} \tag{2.8}$$

z ograniczeniami

$$\begin{cases} \sum_{j=1}^{n} \alpha_{ij} = 1, \text{ gdzie } i \in \{1,2,\ldots,n\}; \\ \forall_{D_k \in \mathscr{D}} \sum_{i \in D_k} \alpha_{ij} \leqslant |D_k| - 1, \text{ gdzie } j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}} \sum_{j=1}^{n} \max_{i \in C_k} \alpha_{ij} \leqslant |C_k| - 1; \\ \alpha_{ij} \in \{0,1\}. \end{cases} \tag{2.9}$$

Dla każdego koloru j wprowadzamy nową zmienną

$$\eta_j = \max_{i \in V} \alpha_{ij}.$$

Dla każdej krawędzi $C_k \in \mathscr{C}$ i każdego koloru j również wprowadzamy nową zmienną

$$\eta_{C_k j} = \max_{i \in C_k} \alpha_{ij}.$$

Po wprowadzeniu nowych zmiennych problem programowania liniowego całkowitoliczbowego sprowadza się do warunku

$$\min z_H = \sum_{j=1}^{n} \eta_j \tag{2.10}$$

z ograniczeniami

$$\begin{cases} \sum_{j=1}^{n} \alpha_{ij} = 1, \text{ gdzie } i \in \{1,2,\ldots,n\}; \\ \forall_{D_k \in \mathscr{D}} \sum_{i \in D_k} \alpha_{ij} \leqslant |D_k| - 1, \text{ gdzie } j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}} \sum_{j=1}^{n} \eta_{C_k j} \leqslant |C_k| - 1; \\ \forall_{C_k \in \mathscr{C}} \alpha_{ij} \leqslant \eta_{C_k j}; \\ \alpha_{ij} \leqslant \eta_j; \\ \alpha_{ij} \in \{0,1\}; \\ \eta_j \in \{0,1\}; \\ \eta_{C_k j} \in \{0,1\}. \end{cases} . \tag{2.11}$$

Twierdzenie 2.1. [11]
Jeśli

$$\ddot{\alpha}_{ij}, \textit{ gdzie } i,j \in \{1,2,\ldots,n\},$$

jest optymalnym rozwiązaniem warunków (2.8) i (2.9), to

$$\begin{cases} \ddot{\alpha}_{ij}, \textit{ gdzie } i,j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}} \ \ddot{\eta}_{C_k j} = \max\limits_{i \in V} \ddot{\alpha}_{ij}, \textit{ gdzie } j \in \{1,2,\ldots,n\}; \\ \ddot{\eta}_j, \textit{ gdzie } j \in \{1,2,\ldots,n\} \end{cases} \tag{2.12}$$

jest optymalnym rozwiązaniem warunków (2.10) i (2.11).
Jeśli

$$\begin{cases} \hat{\alpha}_{ij}, \textit{ gdzie } i,j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}} \ \hat{\eta}_{C_k j}, \textit{ gdzie } j \in \{1,2,\ldots,n\}; \\ \hat{\eta}_j, \textit{ gdzie } j \in \{1,2,\ldots,n\} \end{cases} \tag{2.13}$$

jest optymalnym rozwiązaniem warunków (2.10) i (2.11), to

$$\hat{\alpha}_{ij}, \textit{ gdzie } i,j \in \{1,2,\ldots,n\}$$

jest optymalnym rozwiązaniem warunków (2.8) i (2.9).

Dowód. Niech $\ddot{\alpha}_{ij}$, gdzie $i,j \in \{1,2,\ldots,n\}$ będzie optymalnym rozwiązaniem zagadnienia opisanego warunkami (2.8) i (2.9). Wtedy

$$\begin{cases} \ddot{\alpha}_{ij}, \text{ gdzie } i,j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}} \ \eta_{C_k j} = \max\limits_{i \in V} \ddot{\alpha}_{ij}, \text{ gdzie } j \in \{1,2,\ldots,n\}; \\ \eta_j, \text{ gdzie} j \in \{1,2,\ldots,n\} \end{cases}$$

spełnia warunki (2.11). Jeśli

$$\begin{cases} \hat{\alpha}_{ij}, \text{ gdzie } i,j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}}\ \hat{\eta}_{C_k j}, \text{ gdzie } j \in \{1,2,\ldots,n\}; \\ \hat{\eta}_j, \text{ gdzie} j \in \{1,2,\ldots,n\} \end{cases}$$

jest optymalnym rozwiązaniem warunków (2.10) i (2.11), to

$$\widehat{z_0} = \sum_{j=0}^{n} \hat{\eta}_j \leqslant \sum_{j=1}^{n} \max \ddot{\alpha}_{ij} = \ddot{z}. \tag{2.14}$$

Pokażemy, że ostra nierówność w zależności (2.14) nie jest możliwa. Jeśli by tak było, to $\ddot{\alpha}_{ij}$, gdzie $i,j \in \{1,2,\ldots,n\}$, nie byłoby optymalnym rozwiązaniem (2.8) i (2.9), ponieważ $\hat{\alpha}_{ij}$ spełniają warunki (2.9), co wynika z zależności

$$\begin{cases} \sum\limits_{j=1}^{n} \hat{\alpha}_{ij} = 1, \text{ gdzie } i \in \{1,2,\ldots,n\}; \\ \forall_{D_k \in \mathscr{D}} \sum\limits_{i \in D_k} \hat{\alpha}_{ij} \leqslant |D_k| - 1, \text{ gdzie } j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}} \sum\limits_{j=1}^{n} \max\limits_{i \in C_k} \hat{\alpha}_{ij} \leqslant \sum\limits_{j=1}^{n} \hat{\eta}_{C_k j} \leqslant |C_k| - 1; \\ \hat{\alpha}_{ij} \in \{0,1\}; \\ \hat{\eta}_{C_k j} \in \{0,1\}. \end{cases} \tag{2.15}$$

Wartość funkcji celu dla tego rozwiązania wynosi

$$z_0 = \sum_{j=1}^{n} \max \hat{\alpha}_{ij} \leqslant \sum_{j=1}^{n} \hat{\eta}_j < \sum_{j=1}^{n} \max \ddot{\alpha}_{ij} = \ddot{z}.$$

Jest to sprzeczne z faktem, że $\ddot{\alpha}_{ij}$, gdzie $i,j \in \{1,2,\ldots,n\}$, jest optymalnym rozwiązaniem (2.8), (2.9).

Drugą część udowadniamy analogicznie. Niech teraz

$$\begin{cases} \hat{\alpha}_{ij}, \text{ gdzie } i,j \in \{1,2,\ldots,n\}; \\ \forall_{C_k \in \mathscr{C}}\ \hat{\eta}_{C_k j}, \text{ gdzie } j \in \{1,2,\ldots,n\}; \\ \hat{\eta}_j, \text{ gdzie} j \in \{1,2,\ldots,n\} \end{cases} \tag{2.16}$$

będzie optymalnym rozwiązaniem warunków (2.10), (2.11). Łatwo sprawdzić, że liczby $\hat{\alpha}_{ij}$ spełniają warunki (2.7) (patrz (2.15)). Jeśli założymy, że $\hat{\alpha}_{ij}$, gdzie

$i, j \in \{1, 2, \ldots, n\}$, nie jest optymalnym rozwiązaniem warunku (2.8), to

$$\sum_{j=1}^{n} \hat{\alpha}_{ij} > \sum_{j=1}^{n} \max \ddot{\alpha}_{ij}$$

i otrzymujemy sprzeczność z faktem, że $\hat{\alpha}_{ij}$, $\hat{\eta}_{C_k j}$, $\hat{\eta}_j$ jest optymalnym rozwiązaniem (2.10), (2.11), ponieważ

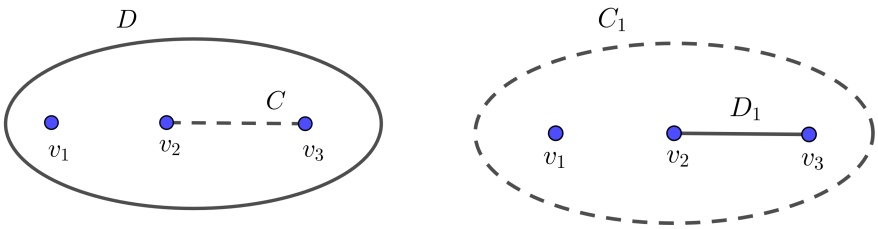
$$\sum_{j=0}^{n} \ddot{\eta}_{C_k j} = \sum_{j=1}^{n} \max_{i \in C_k} \ddot{\alpha}_{ij} < \sum_{j=1}^{n} \max_{i \in C_k} \hat{\alpha}_{ij}.$$

□

Przykład 2.2. Niech dany będzie hipergraf $H = (V, \mathscr{C}, \mathscr{D})$, gdzie

$$V = \{v_1, v_2, v_3\},$$
$$\mathscr{C} = \{C\} = \{v_2, v_3\},$$
$$\mathscr{D} = \{D\} = \{v_1, v_2, v_3\}.$$

Hipergraf mieszany H został zamieszczony na rysunku 4(a).



(a) Hipergraf mieszany H (b) Hipergraf mieszany $\bar{H}$

Rysunek 4. Hipergraf mieszany H oraz jego odwrotność chromatyczna $\bar{H}$
Źródło: Opracowanie własne

W celu wyznaczenia właściwego kolorowania hipergrafu mieszanego H możemy rozważyć następujący problem programowania całkowitoliczbowego.

Hipergraf mieszany H ma trzy wierzchołki. Zgodnie z warunkiem (2.5) wprowadzamy takie ograniczenia, aby każdy wierzchołek otrzymał dokładnie jeden kolor

$$\begin{cases} \alpha_{11} + \alpha_{12} + \alpha_{13} = 1 \\ \alpha_{21} + \alpha_{22} + \alpha_{23} = 1 \\ \alpha_{31} + \alpha_{32} + \alpha_{33} = 1. \end{cases} \tag{2.17}$$

Dla właściwego kolorowania krawędzi D musi zachodzić warunek (2.6), więc

$$\begin{cases} \alpha_{11} + \alpha_{21} + \alpha_{31} \leqslant 2 \\ \alpha_{12} + \alpha_{22} + \alpha_{32} \leqslant 2 \\ \alpha_{13} + \alpha_{23} + \alpha_{33} \leqslant 2. \end{cases} \tag{2.18}$$

Dla właściwego kolorowania krawędzi C musi być spełniony warunek (2.7). Stąd mamy ograniczenia

$$\begin{cases} \eta_{C1} + \eta_{C2} + \eta_{C3} \leqslant 1 \\ \alpha_{21} \leqslant \eta_{C1} \\ \alpha_{22} \leqslant \eta_{C2} \\ \alpha_{23} \leqslant \eta_{C3} \\ \alpha_{31} \leqslant \eta_{C1} \\ \alpha_{32} \leqslant \eta_{C2} \\ \alpha_{33} \leqslant \eta_{C3} \\ \alpha_{11} \leqslant \eta_1 \\ \alpha_{12} \leqslant \eta_2 \\ \alpha_{13} \leqslant \eta_3 \\ \alpha_{21} \leqslant \eta_1 \\ \alpha_{22} \leqslant \eta_2 \\ \alpha_{23} \leqslant \eta_3 \\ \alpha_{31} \leqslant \eta_1 \\ \alpha_{32} \leqslant \eta_2 \\ \alpha_{33} \leqslant \eta_3. \end{cases} \tag{2.19}$$

Aby wyznaczyć kolorowanie oraz dolną liczbę chromatyczną, musimy znaleźć rozwiązanie problemu

$$\min z_H = \eta_1 + \eta_2 + \eta_3 \tag{2.20}$$

z ograniczeniami danymi warunkami (2.17), (2.18), (2.19).

W programie *R* za pomocą pakietu `lpSolve` wyznaczyliśmy rozwiązanie problemu (2.20) z warunkami ograniczającymi (2.17), (2.18), (2.19). Wykorzystaliśmy w tym celu funkcję `lp`, dla której podaliśmy:

— `direction = "min"` – kierunek optymalizacji,
— `objective.in` – wektor składający się ze współczynników funkcji celu,
— `const.mat` – macierz współczynników warunków ograniczających,
— `const.dir` – wektor operatorów porównywania potrzebny do skonstuowania warunków ograniczających,
— `const.rhs` – wektor liczbowy prawych stron warunków ograniczających.

Dodatkowo podaliśmy `all.bins = TRUE`, zapewniając binarność zmiennych [15].

Otrzymaliśmy następujące wyniki:

— macierz A postaci

$$A = [\alpha_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

— oraz

$$\begin{cases} \eta_{C1} = 0 \\ \eta_{C2} = 1 \\ \eta_{C3} = 0 \\ \eta_1 = 1 \\ \eta_2 = 1 \\ \eta_3 = 0. \end{cases} \tag{2.21}$$

Na podstawie macierzy A widzimy, że w celu właściwego pokolorowania hipergrafu mieszanego H należy użyć 2 kolorów. Wyniki podane w (2.21) wskazują na to, że krawędź C jest pokolorowana jednym kolorem oznaczonym jako kolor drugi.

Zatem liczba 2 jest dolną liczbą chromatyczną.

Aby wyznaczyć górną liczbę chromatyczną, możemy wykorzystać odwrotność chromatyczną hipergrafu mieszanego H. Dokonując inwersji hipergrafu mieszanego H, dostajemy, że $\bar{H} = (V, \mathscr{C}_1, \mathscr{D}_1)$, gdzie

$$V = \{v_1, v_2, v_3\},$$

$$\mathscr{C}_1 = \{C_1\} = \{D\} = \{v_1, v_2, v_3\},$$

$$\mathscr{D}_1 = \{D_1\} = \{C\} = \{v_2, v_3\}.$$

Hipergraf mieszany $\bar{H}$ został przedstawiony na rysunku 4(b). W sposób analogiczny jak w przypadku wyznaczania kolorowania oraz dolnej liczby chromatycznej dla hipergrafu mieszanego H, za pomocą programowania całowitoliczbowego musimy znaleźć rozwiązanie problemu

$$\min z_{\bar{H}} = \bar{\eta}_1 + \bar{\eta}_2 + \bar{\eta}_3 \tag{2.22}$$

z ograniczeniami

$$\begin{cases} \bar{\alpha}_{11} + \bar{\alpha}_{12} + \bar{\alpha}_{13} = 1 \\ \bar{\alpha}_{21} + \bar{\alpha}_{22} + \bar{\alpha}_{23} = 1 \\ \bar{\alpha}_{31} + \bar{\alpha}_{32} + \bar{\alpha}_{33} = 1 \\ \bar{\alpha}_{21} + \bar{\alpha}_{31} \leqslant 1 \\ \bar{\alpha}_{22} + \bar{\alpha}_{32} \leqslant 1 \\ \bar{\alpha}_{23} + \bar{\alpha}_{33} \leqslant 1 \\ \eta_{C_1 1} + \eta_{C_1 2} + \eta_{C_1 3} \leqslant 2 \\ \bar{\alpha}_{21} \leqslant \eta_{C1} \\ \bar{\alpha}_{22} \leqslant \eta_{C2} \\ \bar{\alpha}_{23} \leqslant \eta_{C3} \\ \bar{\alpha}_{31} \leqslant \eta_{C1} \\ \bar{\alpha}_{32} \leqslant \eta_{C2} \\ \bar{\alpha}_{33} \leqslant \eta_{C3} \\ \bar{\alpha}_{11} \leqslant \bar{\eta}_1 \\ \bar{\alpha}_{12} \leqslant \bar{\eta}_2 \\ \bar{\alpha}_{13} \leqslant \bar{\eta}_3 \\ \bar{\alpha}_{21} \leqslant \bar{\eta}_1 \\ \bar{\alpha}_{22} \leqslant \bar{\eta}_2 \\ \bar{\alpha}_{23} \leqslant \bar{\eta}_3 \\ \bar{\alpha}_{31} \leqslant \bar{\eta}_1 \\ \bar{\alpha}_{32} \leqslant \bar{\eta}_2 \\ \bar{\alpha}_{33} \leqslant \bar{\eta}_3. \end{cases} \tag{2.23}$$

Modelując problem (2.22) z ograniczeniami (2.23) w programie R, otrzymaliśmy wyniki:

— macierz A postaci

$$A = [\bar{\alpha}_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

— oraz

$$\begin{cases} \eta_{C_1 1} = 1 \\ \eta_{C_1 2} = 1 \\ \eta_{C_1 3} = 0 \\ \bar{\eta}_1 = 1 \\ \bar{\eta}_2 = 1 \\ \bar{\eta}_3 = 0. \end{cases} \tag{2.24}$$

Na podstawie macierzy A możemy stwierdzić, że najmniejszą liczbą kolorów potrzebną do pokolorowania hipergrafu mieszanego $\bar{H}$ jest liczba $\chi(\bar{H}) = 2$.

Na podstawie wyników (2.24) możemy wnioskować, że krawędź C_1 jest pokolorowana przy użyciu dwóch kolorów.

Wyznaczając dolną liczbę chromatyczną dla hipergrafu mieszanego $\bar{H}$, będącego odwrotnością hipergrafu mieszanego H, otrzymaliśmy górną liczbę chromatyczną dla kolorowania hipergrafu H.

Podzbiory monochromatyczne wierzchołków dla kolorowania hipergrafów mieszanych H oraz $\bar{H}$ zostały zamieszczone w tabeli 1. Widzimy, że zarówno dla hipergrafu mieszanego H, jak i hipergrafu $\bar{H}$ istnieje dokładnie jedno ścisłe kolorowanie za pomocą dwóch kolorów. Zatem hipergrafy mieszane H oraz $\bar{H}$ są jednoznacznie kolorowalne.

Dla hipergrafu mieszanego H mamy

$$\chi(H) = \bar{\chi}(H) = 2 \quad \text{oraz} \quad R(H) = (0,1,0).$$

Korzystając z postaci (2.4) wielomianu chromatycznego dla hipergrafów mieszanych jednoznacznie kolorowalnych, otrzymujemy wielomian chromatyczny hipergrafu mieszanego H

$$P(H,\lambda) = r_2(H) \cdot \lambda^{(2)} = \lambda \cdot (\lambda - 1) = \lambda^2 - \lambda. \tag{2.25}$$

Tabela 1. Podzbiory monochromatyczne wierzchołków dla kolorowania hipergrafu mieszanego H oraz jego odwrotności chromatycznej $\bar{H}$

Hipergraf mieszany	Kolor 1	Kolor 2	Kolor 3
H	$\{v_1\}$	$\{v_2, v_3\}$	–
$\bar{H}$	$\{v_1, v_2\}$	$\{v_3\}$	–

Źródło: Opracowanie własne

3. Wyznaczanie kolorowania hipergrafu mieszanego modelującego wybrany problem

3.1. Zastosowanie kolorowania hipergrafów mieszanych

Wiele dziedzin nauki wykorzystuje hipergrafy mieszane do modelowania różnych problemów. Przy użyciu hipergrafu mieszanego $H = (V, \mathscr{C}, \mathscr{D})$ modeluje się terminy, które są połączeniem kontrastowych względem siebie ograniczeń opisanych za pomocą rodzin $\mathscr{C}$ oraz $\mathscr{D}$. W genetyce hipergrafy mieszane mogą pomóc w opisie

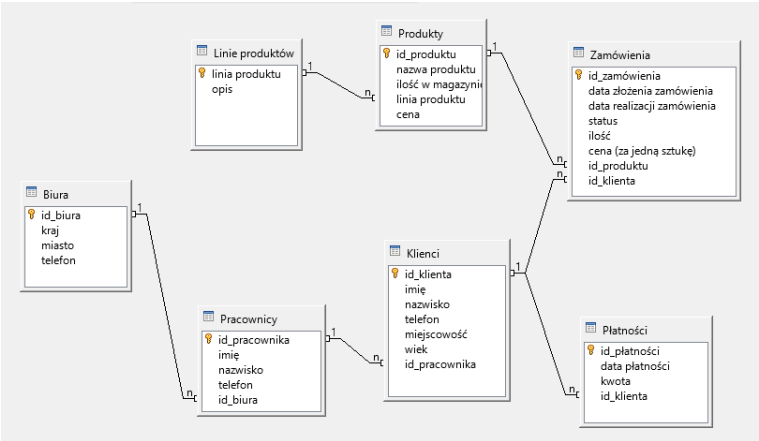
dziedziczenia cech recesywnych w populacji. Mogą być narzędziem pomocnym w interpretacji liczb naturalnych oraz zliczania obiektów. Innymi obszarami, w których stosuje się hipergrafy mieszane są obliczenia równoległe, biologia molekularna, alokacja zasobów oraz zarządzanie bazami danych. Więcej na temat wymienionych zastosowań można znaleźć w [7, 12].

3.2. Opis problemu zamodelowanego z wykorzystaniem hipergrafu mieszanego

Hipergrafy mieszane stosuje się m. in. do rozdzielenia zadań wymagających pewnych zasobów w taki sposób, aby zadania były wykonane w jak najkrótszym czasie.

Rozważmy zbiór $\{v_1, v_2, \ldots, v_n\}$, gdzie $n \in \mathbb{N}$, który reprezentuje pewne zadania do wykonania. Aby zadania mogły być wykonane, muszą być dostępne pewne zasoby. Niech $\{y_1, y_2, \ldots, y_m\}$, gdzie $m \in \mathbb{N}$, będzie zbiorem zasobów. Załóżmy, że czas jest mierzony w sposób dyskretny i każde z zadań wykonywane jest w danej jednostce czasu.

Jako konkretny przykład rozdzielenia zadań w najkrótszym możliwym czasie rozważymy wykonanie zapytań do bazy danych. Diagram bazy danych został przedstawiony na rysunku 5.



Rysunek 5. Diagram bazy danych
Źródło: Opracowanie własne

Zbiór tabel (plików) w bazie danych oznaczymy przez $Y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$. Przyporządkujmy elementy zbioru Y odpowiednim tabelom w bazie danych w następujący sposób:

— y_1 – Biura,
— y_2 – Pracownicy,

— y_3 – Klienci,
— y_4 – Płatności,
— y_5 – Zamówienia,
— y_6 – Produkty,
— y_7 – Linie produktów.

Zbiór zapytań do bazy danych niech będzie reprezentowany przez zbiór wierzchołków $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$ w następujący sposób:

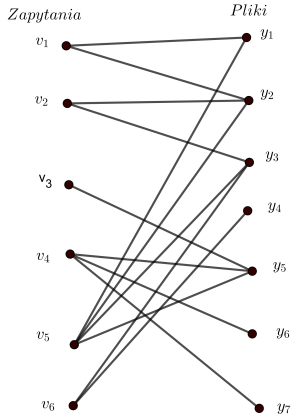
— v_1 – podać liczbę pracowników w każdym biurze (w tabelce ma się znajdować id_biura, miasto, liczba pracowników),
— v_2 – podać dla każdego pracownika liczbę klientów przez niego obsługiwanych (w tabelce ma się znajdować imię, nazwisko pracownika, liczba klientów),
— v_3 – podać liczbę zamówień dla każdego miesiąca w 2021 roku (w tabelce ma się znajdować miesiąc, liczba zamówień),
— v_4 – podać dla każdej linii produktu liczbę zamówień z nią związanych (w tabelce ma się znajdować linia produktów, liczba zamówień),
— v_5 – podać dla każdego biura liczbę zamówień oraz średni czas realizacji zamówienia (w tabelce ma się znajdować id_biura, miasto, średni czas realizacji zamówienia),
— v_6 – podać dla każdego klienta z Warszawy sumaryczną kwotę wydaną na zakupy (w tabelce ma się znajdować imię, nazwisko klienta, suma).

Przez $S_i \subseteq Y$ oznaczmy zbiór tych plików, które są potrzebne do realizacji zapytania v_i, gdzie $i \in \{1, 2, \ldots, 6\}$. Do wykonania zapytania v_1 potrzebujemy tabel: Biura oraz Pracownicy, więc $S_1 = \{y_1, y_2\}$. Do realizacji zapytania v_2 potrzebujemy tabel: Pracownicy i Klienci, stąd $S_2 = \{y_2, y_3\}$. Zapytanie v_3 zrealizuje się przy dostępie do tabeli Klienci, więc $S_3 = \{y_3\}$. Do wykonania zapytania v_4 potrzebujemy trzech tabel o nazwach: Zamówienia, Produkty, Linie produktów, zatem $S_4 = \{y_5, y_6, y_7\}$. Do realizacji zapytania v_5 potrzebujemy tabel: Biura, Pracownicy, Klienci, Zamówienia, więc $S_5 = \{y_1, y_2, y_3, y_5\}$. Ostanie zapytanie v_6 będzie wykonane przy dostępie do tabel: Klienci oraz Płatności, stąd $S_6 = \{y_3, y_4\}$. Związki między zapytaniami a tabelami możemy przedstawić w formie grafu dwudzielnego zaprezentowanego na rysunku 6(a).

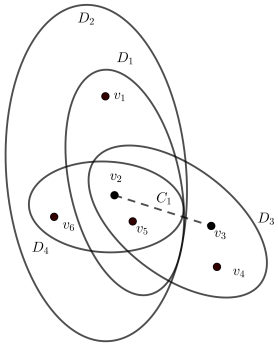
Zakładamy, że zapytania, które wymagają dostępu tych samych plików nie mogą być wykonane jednocześnie. Możemy skonstruować hipergraf mieszany, w którym D-krawędzie powstają w następujący sposób

$$S_1 \cap S_2 = \{y_2\}, \text{ więc } D_1 = \{v_1, v_2, v_5\},$$

$$S_1 \cap S_5 = \{y_1, y_2\} \text{ więc } D_1 = \{v_1, v_2, v_5\},$$



(a) Graf dwudzielny przedstawiający związki między zapytaniami i plikami w bazie danych



(b) Hipergraf mieszany H

Rysunek 6. Ilustracja problemu z użyciem grafu dwudzielnego i hipergrafu mieszanego
Źródło: Opracowanie własne

$$S_2 \cap S_5 = \{y_2, y_3\}, \text{ więc } D_2 = \{v_1, v_2, v_5, v_6\},$$
$$S_3 \cap S_5 = S_4 \cap S_5 = \{y_5\}, \text{ więc } D_3 = \{v_3, v_4, v_5\},$$
$$S_5 \cap S_6 = \{y_3\}, \text{ więc } D_4 = \{v_2, v_5, v_6\}.$$

Dodatkowo wprowadzamy wymaganie, że zapytanie v_2 oraz zapytanie v_3 mają być wykonane w tej samej jednostce czasu, gdyż potrzebujemy wyniku zapytań z tej samej chwili. Przy wykonywaniu zapytań w różnych jednostkach czasu mogłoby dojść do zmiany danych w bazie. W rezultacie otrzymujemy C-krawędź $C_1 = \{v_2, v_3\}$. Skonstruowany hipergraf mieszany H jest przedstawiony na rysunku 6(b). C-krawędzie dla odróżnienia od D-krawędzi będą oznaczane linią przerywaną.

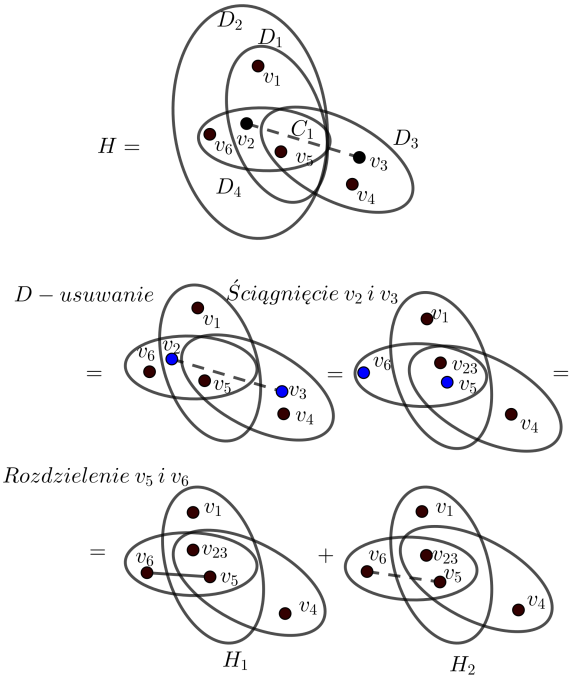
3.3. Kolorowanie skonstruowanego hipergrafu mieszanego za pomocą algorytmu rozdzielania-ściągania

W celu wyznaczenia optymalnej liczby jednostek czasu potrzebnych do wykonania wszystkich zadań, tzn. udzielenia odpowiedzi na zapytania, użyliśmy algorytmu rozdzielania-ściągania opisanego w podrozdziale 2.3.

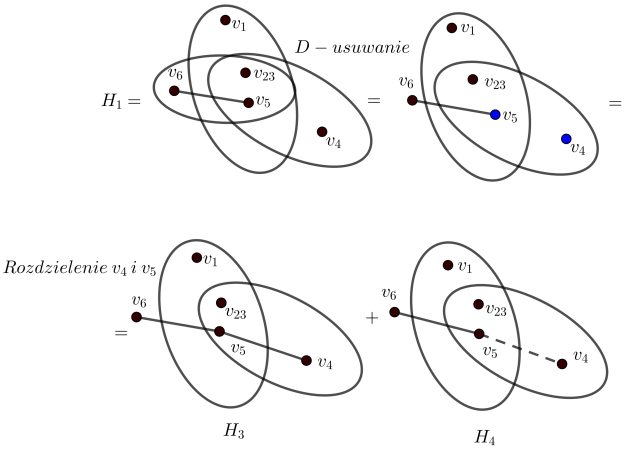
Przebieg algorytmu został zilustrowany na rysunkach 7–23.

Z rysunku 7 możemy odczytać, że w pierwszym kroku algorytmu dokonaliśmy D-usuwania krawędzi D_2, ponieważ $D_1 \subseteq D_2$. Następnie dokonaliśmy ściągnięcia wierzchołków v_2 i v_3. Wielomian chromatyczny hipergrafu mieszanego H jest równy sumie wielomianów chromatycznych H_1 oraz H_2. W kolejnym kroku wykonaliśmy rozdzielenie wierzchołków v_5 oraz v_6. Powstała nowa D-krawędź w hipergrafie H_1 oraz nowa C-krawędź w hipergrafie H_2.

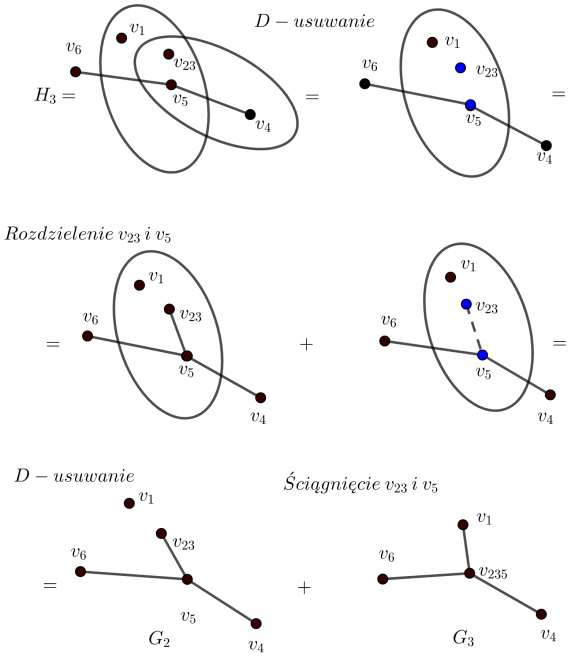
W celu ułatwienia obserwacji kolejnych kroków algorytmu pokazaliśmy przebieg algorytmu oddzielnie dla hipergrafu H_1 oraz H_2. Kolejne kroki algorytmu możemy obserwować, analizując pozostałe rysunki, na których zostały zamieszczone komentarze.



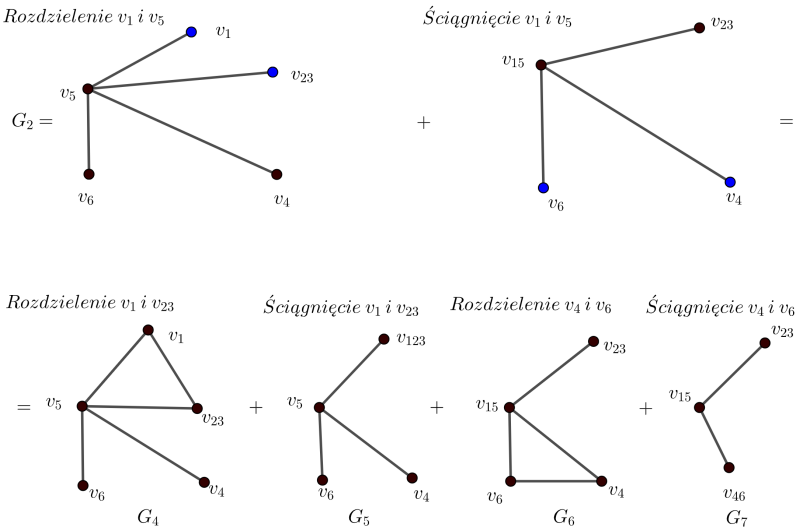
Rysunek 7. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H
Źródło: Opracowanie własne



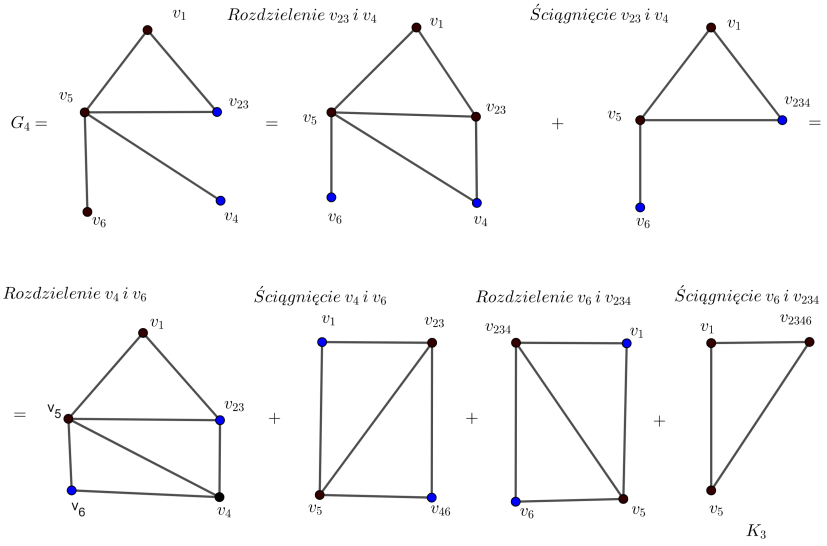
Rysunek 8. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu H_1
Źródło: Opracowanie własne



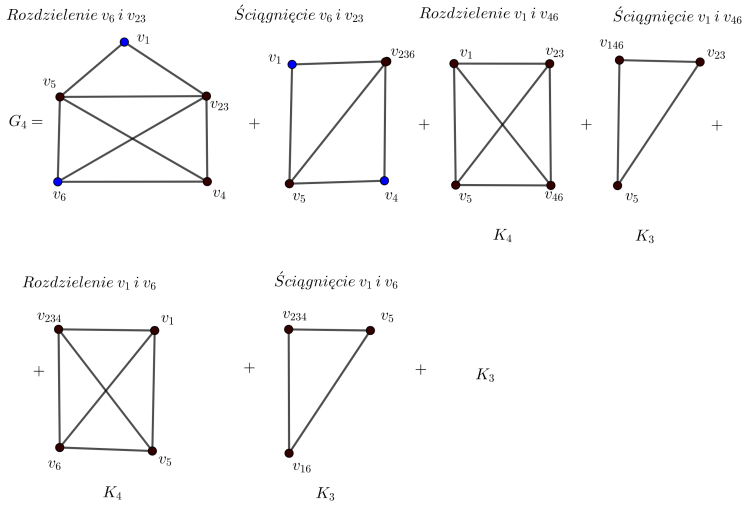
Rysunek 9. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu H_3
Źródło: Opracowanie własne



Rysunek 10. Przebieg algorytmu rozdzielania-ściągania dla grafu G_2
Źródło: Opracowanie własne

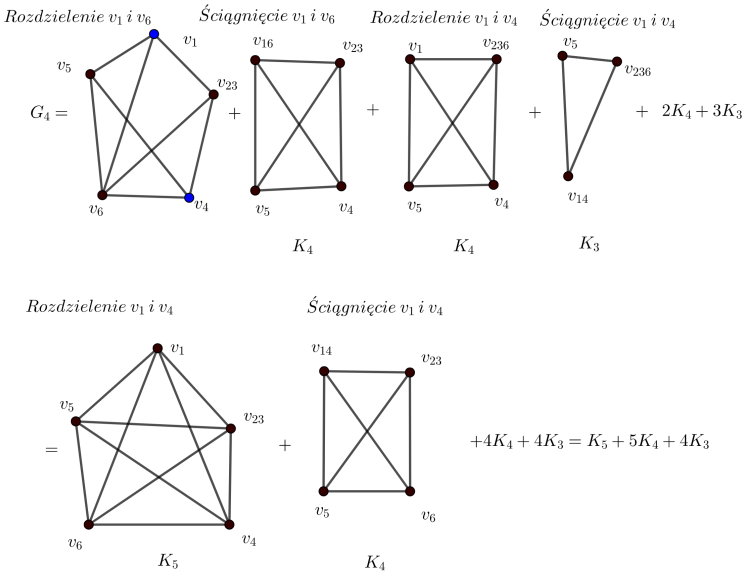


Rysunek 11. Przebieg algorytmu rozdzielania-ściągania dla grafu G_4 (część pierwsza)
Źródło: Opracowanie własne



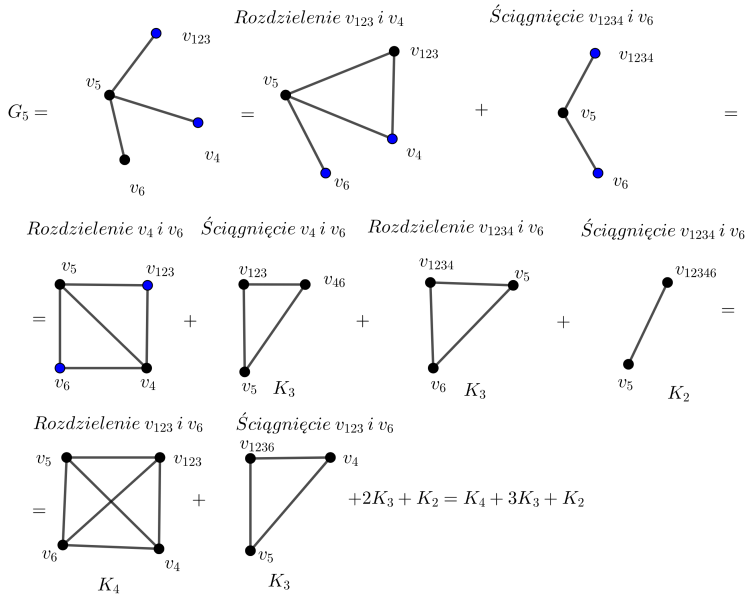
Rysunek 12. Przebieg algorytmu rozdzielania-ściągania dla grafu G_4 (część druga)

Źródło: Opracowanie własne

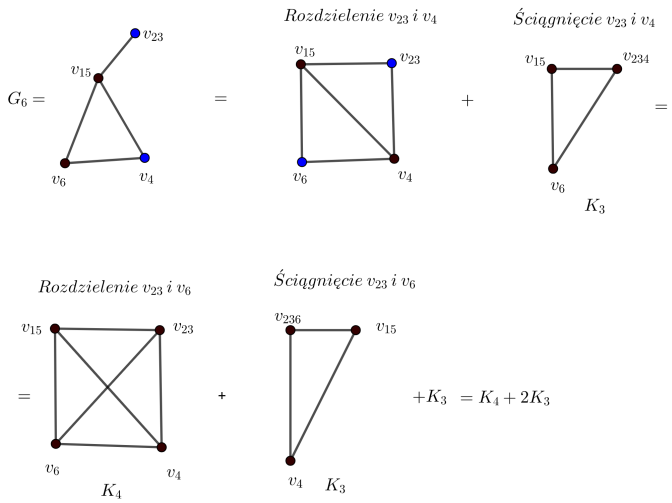


Rysunek 13. Przebieg algorytmu rozdzielania-ściągania dla grafu G_4 (część trzecia)

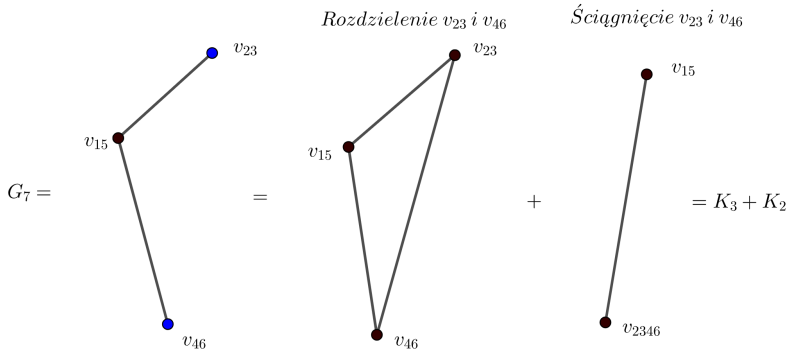
Źródło: Opracowanie własne



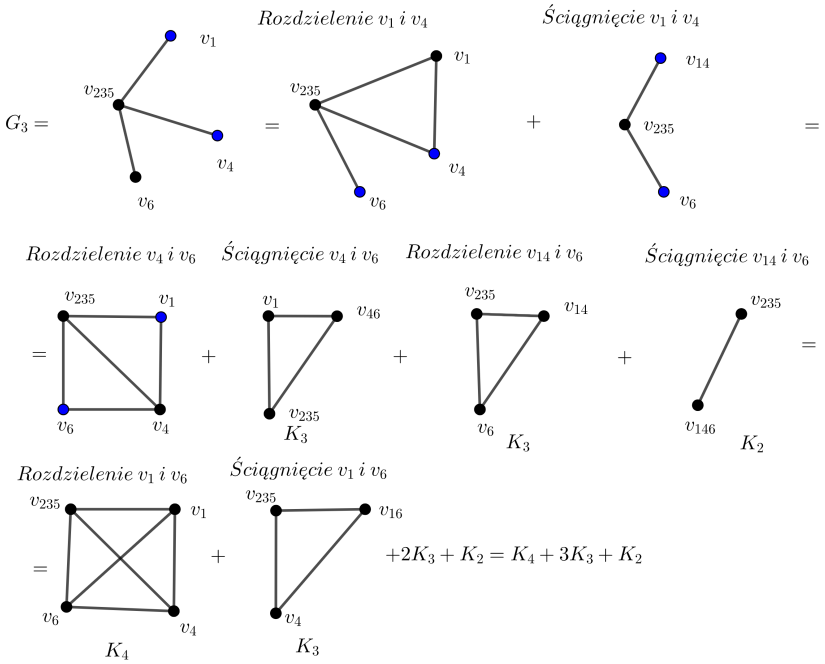
Rysunek 14. Przebieg algorytmu rozdzielania-ściągania dla grafu G_5
Źródło: Opracowanie własne



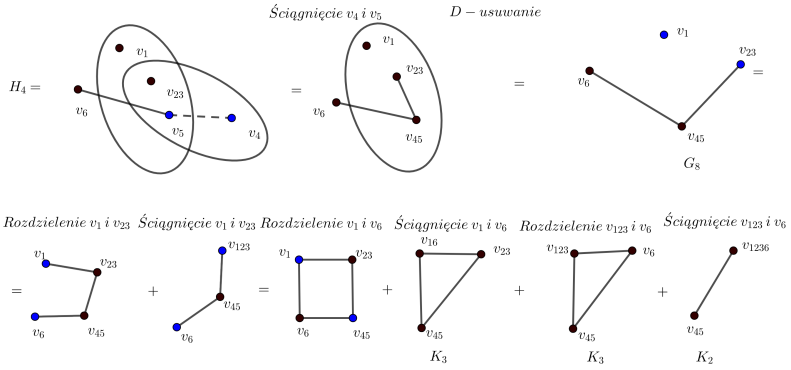
Rysunek 15. Przebieg algorytmu rozdzielania-ściągania dla grafu G_6
Źródło: Opracowanie własne



Rysunek 16. Przebieg algorytmu rozdzielania-ściągania dla grafu G_7
Źródło: Opracowanie własne

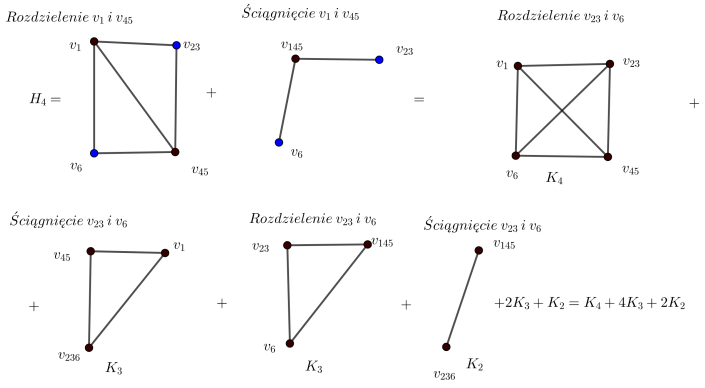


Rysunek 17. Przebieg algorytmu rozdzielania-ściągania dla grafu G_3
Źródło: Opracowanie własne



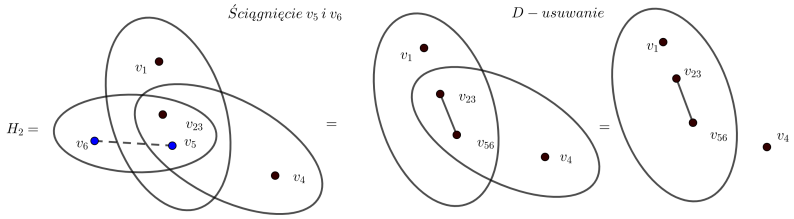
Rysunek 18. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H_4 (część pierwsza)

Źródło: Opracowanie własne



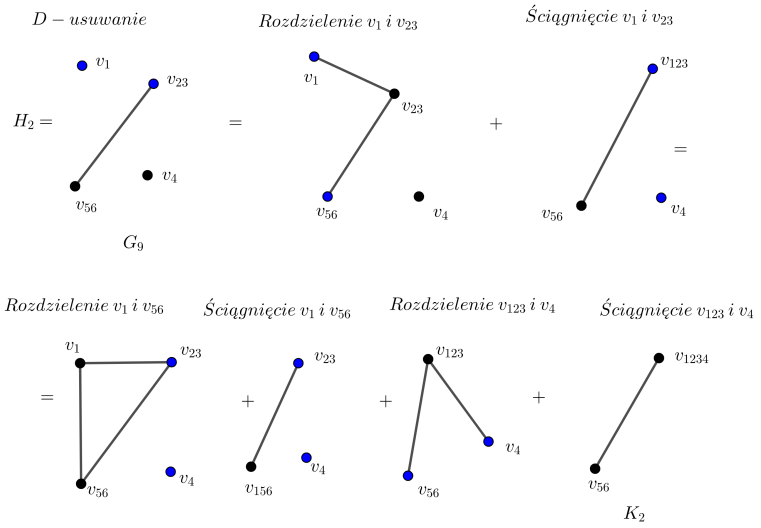
Rysunek 19. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H_4 (część druga)

Źródło: Opracowanie własne



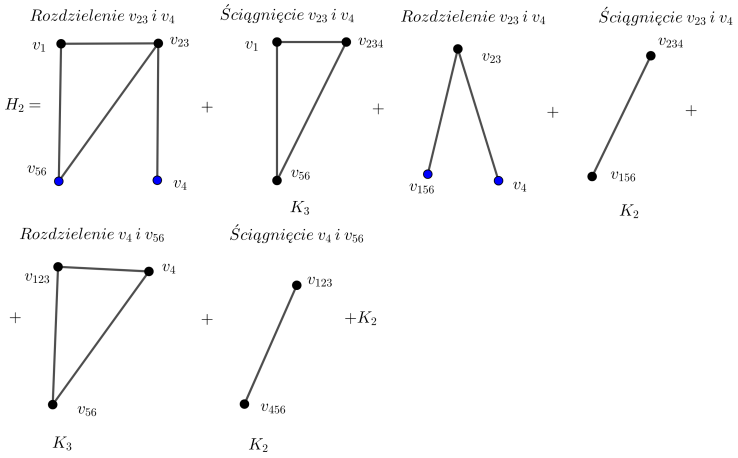
Rysunek 20. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H_2 (część pierwsza)

Źródło: Opracowanie własne



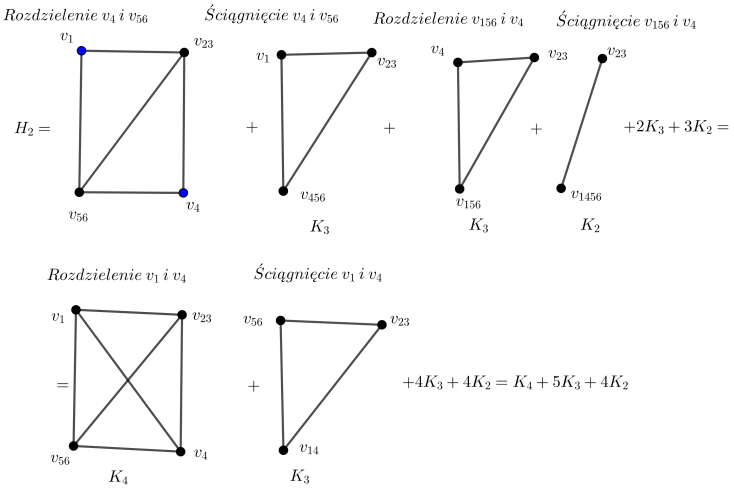
Rysunek 21. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H_2 (część druga)

Źródło: Opracowanie własne



Rysunek 22. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H_2 (część trzecia)

Źródło: Opracowanie własne



Rysunek 23. Przebieg algorytmu rozdzielania-ściągania dla hipergrafu mieszanego H_2 (część czwarta)

Źródło: Opracowanie własne

Liczby grafów pełnych o 2, 3, 4, 5 wierzchołkach zwrócone przez algorytm zostały zaprezentowane w tabeli 2.

Tabela 2. Liczby grafów pełnych zwróconych przez algorytm

Liczba wierzchołków grafu pełnego K_i	**Liczba grafów pełnych r_i**
2	9
3	22
4	10
5	1

Źródło: Opracowanie własne

Aby wyznaczyć wielomiany chromatyczne grafów pełnych, korzystamy z twierdzenia 1.1. Otrzymujemy wówczas, że:

— wielomian chromatyczny grafu pełnego K_2 wynosi

$$P(\lambda, K_2) = \lambda^{(2)} = (\lambda - 1)\lambda = \lambda^2 - \lambda,$$

— wielomian chromatyczny grafu pełnego K_3 wynosi

$$P(\lambda, K_3) = \lambda^{(3)} = (\lambda - 2)(\lambda - 1)\lambda = \lambda^3 - 3\lambda^2 + 2\lambda,$$

— wielomian chromatyczny grafu pełnego K_4 wynosi

$$P(\lambda, K_4) = \lambda^{(4)} = (\lambda - 3)(\lambda - 2)(\lambda - 1)\lambda = \lambda^4 - 6\lambda^3 + 11\lambda^2 - 6\lambda,$$

— wielomian chromatyczny grafu pełnego K_5 wynosi

$$P(\lambda, K_5) = \lambda^{(5)} = (\lambda - 4)(\lambda - 3)(\lambda - 2)(\lambda - 1)\lambda =$$

$$= \lambda^5 - 10\lambda^4 + 35\lambda^3 - 50\lambda^2 + 24\lambda.$$

Wielomian chromatyczny $P(\lambda, H)$ wyznaczony zgodnie z zależnością (2.2) przyjmuje postać

$$P(\lambda, H) = \sum_{i=2}^{5} r_i(H) \cdot \lambda^{(i)} = 1 \cdot (\lambda^5 - 10\lambda^4 + 35\lambda^3 - 50\lambda^2 + 24\lambda) +$$

$$+10 \cdot (\lambda^4 - 6\lambda^3 + 11\lambda^2 - 6\lambda) + 22 \cdot (\lambda^3 - 3\lambda^2 + 2\lambda) + 9 \cdot (\lambda^2 - \lambda) =$$

$$= \lambda^5 - 10\lambda^4 + 35\lambda^3 - 50\lambda^2 + 24\lambda + 10\lambda^4 - 60\lambda^3 + 110\lambda^2 - 60\lambda + 22\lambda^3 - 66\lambda^2 +$$

$$+44\lambda + 9\lambda^2 - 9\lambda = \lambda^5 - 3\lambda^3 + 3\lambda^2 - \lambda.$$

Mając wyznaczone przez algorytm liczby grafów pełnych, zamieszczone w tabeli 2, możemy łatwo wyznaczyć dolną liczbę chromatyczną $\chi(H) = 2$ i górną liczbę chromatyczną $\bar{\chi}(H) = 5$.

Spektrum chromatyczne zgodnie z zależnością (2.1) przyjmie postać

$$R(H) = (0, 9, 22, 10, 1, 0).$$

Wielomian chromatyczny $P(H, \lambda)$ moglibyśmy również wyznaczyć w oparciu o powstające podczas wykonywania algorytmu grafy

$$P(H,\lambda) = P(H_1,\lambda) + P(H_2,\lambda) = P(H_3,\lambda) + P(H_4,\lambda) + P(H_2,\lambda) =$$

$$= P(G_2,\lambda) + P(G_3,\lambda) + P(G_8,\lambda) + P(G_9,\lambda).$$

W kolejnych krokach doszliśmy do grafów: lasu G_2 oraz drzewa G_3, które są zamieszczone na rysunku 9. Otrzymaliśmy również las G_8 przedstawiony na rysunku 18 oraz las G_9 umieszczony na rysunku 21. Zatem wielomian chromatyczny hipergrafu mieszanego H będzie sumą wielomianów chromatycznych powstających drzew i lasów. Wielomany chromatyczne dla lasów i drzew można wyznaczyć za pomocą twierdzeń 1.3 oraz 1.4.
Mamy:

— $P(G_2,\lambda) = \lambda \cdot \lambda \cdot (\lambda - 1)^3$,
— $P(G_3,\lambda) = \lambda \cdot (\lambda - 1)^3$,
— $P(G_8,\lambda) = \lambda \cdot \lambda \cdot (\lambda - 1)^2$,
— $P(G_9,\lambda) = \lambda \cdot \lambda \cdot \lambda \cdot (\lambda - 1)$.

Ostatecznie otrzymujemy, że

$$P(H,\lambda) = (\lambda^2 - \lambda)(\lambda^3 - 2\lambda^2 + \lambda + \lambda + \lambda^2 - 2\lambda + 1 + \lambda^2 - \lambda + \lambda^2) =$$

$$= \lambda^5 + \lambda^4 - 2\lambda^3 + \lambda^2 - \lambda^4 - \lambda^3 + 2\lambda^2 - \lambda = \lambda^5 - 3\lambda^3 + 3\lambda^2 - \lambda.$$

Uzyskany w ten sposób wielomian chromatyczny potwierdza poprawność wyniku otrzymanego za pomocą użytego wcześniej algorytmu rozdzielania-ściągania dla hipergrafu mieszanego.

Dodatkowo, wyznaczanie wielomianu chromatycznego przy wykorzystaniu zależności (2.2) przeprowadziliśmy w programie Maxima. Maxima jest programem wspierającym wykonywanie obliczeń symbolicznych oraz numerycznych [16]. W tym programie zaimplementowano funkcję obliczającą wielomian chromatyczny hipergrafu mieszanego H, która zwróciła identyczną postać wielomianu chromatycznego jak w przypadku poprzednich metod.

W tabeli 3 przedstawiliśmy podzbiory monochromatyczne wierzchołków dla odpowiednich kolorów. Widzimy, że dla kolorowania za pomocą $\bar{\chi}(H) = 5$ kolorów istnieje jedno ścisłe kolorowanie.

Tabela 3. Podzbiory monochromatyczne wierzchołków kolorowania hipergrafu mieszanego H

Kolor 1	**Kolor 2**	**Kolor 3**	**Kolor 4**	**Kolor 5**
$\{v_1, v_2, v_3, v_4, v_6\}$	$\{v_5\}$			
$\{v_2, v_3, v_4, v_6\}$	$\{v_1, v_5\}$			
$\{v_1, v_4, v_6\}$	$\{v_2, v_3, v_5\}$			
$\{v_1, v_2, v_3, v_6\}$	$\{v_4, v_5\}$			
$\{v_1, v_4, v_5\}$	$\{v_2, v_3, v_6\}$			
$\{v_1, v_2, v_3, v_4\}$	$\{v_5, v_6\}$			
$\{v_1, v_5, v_6\}$	$\{v_2, v_3, v_4\}$			
$\{v_4, v_5, v_6\}$	$\{v_1, v_2, v_3\}$			
$\{v_1, v_4, v_5, v_6\}$	$\{v_2, v_3\}$			
$\{v_1\}$	$\{v_2, v_3, v_4, v_6\}$	$\{v_5\}$		
$\{v_1, v_4, v_6\}$	$\{v_2, v_3\}$	$\{v_5\}$		
$\{v_1, v_6\}$	$\{v_2, v_3, v_4\}$	$\{v_5\}$		
$\{v_1, v_4\}$	$\{v_2, v_3, v_6\}$	$\{v_5\}$		
$\{v_1, v_2, v_3\}$	$\{v_4, v_6\}$	$\{v_5\}$		
$\{v_1, v_2, v_3, v_4\}$	$\{v_6\}$	$\{v_5\}$		
$\{v_1, v_2, v_3, v_6\}$	$\{v_4\}$	$\{v_5\}$		
$\{v_1, v_5\}$	$\{v_2, v_3, v_4\}$	$\{v_6\}$		
$\{v_1, v_5\}$	$\{v_2, v_3, v_6\}$	$\{v_4\}$		
$\{v_1, v_5\}$	$\{v_2, v_3\}$	$\{v_4, v_6\}$		
$\{v_1\}$	$\{v_4, v_6\}$	$\{v_2, v_3, v_5\}$		
$\{v_1, v_4\}$	$\{v_6\}$	$\{v_2, v_3, v_5\}$		
$\{v_1, v_6\}$	$\{v_4\}$	$\{v_2, v_3, v_5\}$		
$\{v_1, v_6\}$	$\{v_4, v_5\}$	$\{v_2, v_3\}$		
$\{v_6\}$	$\{v_4, v_5\}$	$\{v_1, v_2, v_3\}$		
$\{v_1\}$	$\{v_4, v_5\}$	$\{v_2, v_3, v_6\}$		
$\{v_6\}$	$\{v_1, v_4, v_5\}$	$\{v_2, v_3\}$		
$\{v_1\}$	$\{v_2, v_3, v_4\}$	$\{v_5, v_6\}$		
$\{v_1, v_2, v_3\}$	$\{v_4\}$	$\{v_5, v_6\}$		
$\{v_1\}$	$\{v_2, v_3\}$	$\{v_4, v_5, v_6\}$		
$\{v_4\}$	$\{v_2, v_3\}$	$\{v_1, v_5, v_6\}$		
$\{v_1, v_4\}$	$\{v_2, v_3\}$	$\{v_5, v_6\}$		

Kolor 1	Kolor 2	Kolor 3	Kolor 4	Kolor 5
$\{v_1\}$	$\{v_2, v_3\}$	$\{v_4, v_6\}$	$\{v_5\}$	
$\{v_1\}$	$\{v_2, v_3, v_4\}$	$\{v_5\}$	$\{v_6\}$	
$\{v_1, v_6\}$	$\{v_2, v_3\}$	$\{v_5\}$	$\{v_4\}$	
$\{v_1\}$	$\{v_2, v_3, v_6\}$	$\{v_5\}$	$\{v_4\}$	
$\{v_1, v_4\}$	$\{v_2, v_3\}$	$\{v_5\}$	$\{v_6\}$	
$\{v_4\}$	$\{v_1, v_2, v_3\}$	$\{v_5\}$	$\{v_6\}$	
$\{v_4\}$	$\{v_1, v_5\}$	$\{v_2, v_3\}$	$\{v_6\}$	
$\{v_4\}$	$\{v_1\}$	$\{v_6\}$	$\{v_2, v_3, v_5\}$	
$\{v_1\}$	$\{v_6\}$	$\{v_2, v_3\}$	$\{v_4, v_5\}$	
$\{v_1\}$	$\{v_5, v_6\}$	$\{v_2, v_3\}$	$\{v_4\}$	
$\{v_1\}$	$\{v_2, v_3\}$	$\{v_5\}$	$\{v_6\}$	$\{v_4\}$

Źródło: Opracowanie własne

3.4. Kolorowanie skonstruowanego hipergrafu mieszanego z wykorzystaniem programowania całkowitoliczbowego

Przeprowadziliśmy również w programie *R* modelowanie kolorowania hipergrafu mieszanego H za pomocą programowania całkowitoliczbowego danego warunkiem (2.10) z ograniczeniami (2.11).

Najpierw wyznaczyliśmy ograniczenia (2.11) dla krawędzi C_1 oraz D_1, D_2, D_3, D_4, gdzie $i, j \in \{1,2,3,4,5,6\}$. Funkcja celu przyjęła postać

$$\min z_H = \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 + \eta_6. \tag{3.1}$$

Otrzymaliśmy macierz A

$$A = [\alpha_{ij}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Stąd widzimy, że dolna liczba chromatyczna wynosi $\chi(H) = 2$.

Podsumowując, najmniejsza liczba jednostek czasu potrzebna do wykonania zadań wynosi 2.

Podsumowanie

Rozdział stanowi przegląd pojęć pochodzących z teorii grafów oraz teorii hipergrafów potrzebnych do zrozumienia problematyki kolorowania hipergrafów mieszanych.

W rozdziale opisano algorytm rozdzielania-ściągania wykorzystywany do wyznaczania kolorowania hipergrafów mieszanych. Pokazano, że kolorowanie hipergrafów mieszanych może być modelowane za pomocą programowania całkowitoliczbowego.

Dla opracowanego przykładu hipergrafu modelującego problem alokacji zasobów zaprezentowano przebieg algorytmu kolorowania hipergrafów mieszanych. Algorytm ten sprowadza wielomian chromatyczny hipergrafu mieszanego do sumy wielomianów chromatycznych grafów pełnych. Za pomocą zwróconych przez algorytm liczb grafów pełnych został wyznaczoney wielomian chromatyczny oraz dolna i górna liczba chromatyczna. Dolna liczba chromatyczna była odpowiedzią na pytanie dotyczące optymalnej liczby jednostek czasu potrzebnych do wykonania wszystkich zadań przy przyjęciu pewnych ograniczeń. Wielomian chromatyczny został również wyznaczony w oparciu o powstające podczas wykonywania algorytmu drzewa i lasy. W obu podejściach otrzymano identyczne wyniki. Dolną liczbę chromatyczną obliczono modelując kolorowanie hipergrafu mieszanego H za pomocą programowania całkowitoliczbowego.

Literatura

[1] Allagan J.A., Slutzky D., *Chromatic Polynomials of some Mixed Hypergraphs*, „Australasian Journal of Combinatorics", 2014, vol. 58(1), p. 197–213.

[2] Berge C., *Graphs and Hypergraphs*, North-Holland Mathematical Library, Amsterdam, 1976.

[3] Berge C., *Hypergraphs: Combinatorics of Finite Sets*, North-Holland Mathematical Library, Amsterdam, 1989.

[4] Birkhoff G.D., *A Determinant Formula for the Number of Ways of Coloring a Map*, „Annals of Mathematics", 1912, vol. 14(1).

[5] Voloshin V., *The Mixed Hypergraphs*, „Computer Science Journal of Moldova", 1993, vol.1(1), p. 45–52.

[6] Łazuka E., *Wielomian chromatyczny jako narzędzie klasyfikacji hipergrafów* [W:] *Matematyczne miscellanea*, t. 3, red. I. Gorgol, Wydawnictwo Politechniki Lubelskiej, Lublin, 2018.

[7] Popławska E., *Hipergrafy mieszane i ich kolorowanie*, [W:] *Koła naukowe etapem rozwoju kompetencji zawodowych i naukowych*, red. B. Buraczyńska, Wydawnictwo Politechniki Lubelskiej, Lublin, 2023.

[8] Read R.C., *An Introduction to Chromatic Polynomials*, „Journal of Combinatorial Theory", 1968, vol. 4, p. 52–71.

[9] Wilson R., *Wprowadzenie do teorii grafów*, PWN, Warszawa, 2007.

[10] Tomescu I., *Chromatic Coefficients of Linear Uniform Hypergraphs*, „Journal of Combinatorial Theory", 1998, vol. 72, p. 229–235.
[11] Lozovanu D., Voloshin V.I., *Integer Programming Models for Colorings of Mixed Hypergraphs*, „Computer Science Journal of Moldova", 2000, vol. 8(1), p. 64–74.
[12] Voloshin V., *Coloring Mixed Hypergraphs: Theory, Algorithms and Applications*, American Mathematical Society, Providence, 2002.
[13] Tuza Z., Voloshin V.I., Zhou H., *Uniquely Colorable Mixed Hypergraphs*, „Discrete Mathematics", 2002, vol. 248, p. 221–236.
[14] Zhang R., *Chromatic Polynomials of Hypergraphs and Mixed Hypergrapghs*, National Institute of Education NanYang Technological University, 2020.
[15] Berkelaar M., Csárdi G., *Interface to Lpsolve v. 5.5 to Solve Linear/Integer Programs*, https://cran.r-project.org/web/packages/lpSolve/lpSolve.pdf, [dostęp: 15.05.2022r].
[16] *A Computer Algebra System*, https://maxima.sourceforge.io/, [dostęp: 15.05.2022r].

Anna Futa [1], **Magdalena Jastrzębska** [2]

Zastosowanie algorytmów grafowych do modelowania zintegrowanego systemu miejskiej gospodarki wodnej

Streszczenie

Stosowanie metod wykorzystujących grafy do modelowania różnorodnych zagadnień inżynierskich jest znane od kilkudziesięciu lat, natomiast zastosowanie algorytmów grafowych do modelowania zagadnień miejskiej gospodarki wodnej jest podejściem nowym. W rozdziale przedstawiono wybrane algorytmy teorii grafów służące do modelowania zintegrowanych systemów miejskiej gospodarki wodnej. Część z przedstawionych algorytmów wywodzi się bezpośrednio z teorii grafów, niektóre z nich zostały opracowane na podstawie innych nauk, m.in. inżynierii środowiska czy genetyki, w celu rozwiązywania konkretnych problemów inżynierskich.

Słowa kluczowe: *modelowanie grafami, algorytmy grafowe, analiza zintegrowana, modelowanie sieci, woda miejska*

Abstract

The application of methods using graphs to model a variety of engineering issues has been known for several decades, but the application of graph algorithms to model the urban water management issues is a completely new approach. The chapter presents selected graph theory algorithms for modeling integrated urban water management systems. Some of the presented algorithms are directly derived from graph theory, while others were developed from other sciences, including environmental engineering or genetics, to solve specific engineering problems.

Keywords: *graph modeling, graph algorithms, integrated analysis, network modeling, municipal water*

Wstęp

Od końca XX wieku zmienia się postrzeganie systemów inżynieryjnych związanych z zaopatrzeniem w wodę i odprowadzaniem ścieków na terenach zurbanizowanych. Układy te dawniej rozpatrywano oddzielnie, jako niezależne układy, w których zachodzą zjawiska fizyczne związane z transportem wody, ścieków i osadów. Obecnie obserwuje się tendencję do całościowego analizowania tych systemów oraz traktowania poszczególnych ich części jako jednego zintegrowanego systemu miejskiej gospodarki wodnej (ZSMGW). Można to zaobserwować w pracach wielu autorów, którzy postulują konieczność podjęcia wysiłków w celu opracowania kompleksowej metodologii projektowania, analizy i działania ZSMGW [17, 22, 24].

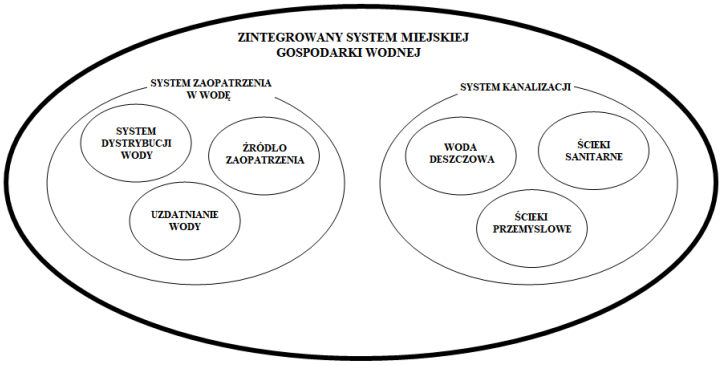
[1] Katedra Matematyki Stosowanej, Politechnika Lubelska
[2] Katedra Matematyki Stosowanej, Politechnika Lubelska

W związku z tym poszukuje się narzędzi, które będą mogły analizować zintegrowany system całościowo, jak również w obszarze poszczególnych podsystemów. Takim narzędziem matematycznym są algorytmy grafowe. Niektóre algorytmy powstały w ramach teorii grafów, inne zaś w wyniku rozwiązania konkretnych problemów inżynierskich.

Pierwszą nauką stosowaną, która wykorzystała narzędzia teorii grafów była elektrotechnika (prawo Kirchhoffa, topologiczne metody analizy obwodów). W ostatnich latach nastąpił intensywny rozwój zastosowań teorii grafów w informatyce, zarówno w kontekście metod (bazy danych, sztuczna inteligencja), jak i algorytmów kombinatorycznych [28]. Obecnie nowym, popularnym trendem jest zastosowanie narzędzi teorii grafów w naukach społecznych. Coraz więcej dziedzin wykorzystuje grafy jako obiekty uniwersalne, które pomagają opisywać różne zagadnienia z matematyczną precyzją, modelować je, a następnie dostarczać rozwiązania za pomocą metod i algorytmów grafowych.

1. Zintegrowany system miejskiej gospodarki wodnej

Zintegrowany system miejskiej gospodarki wodnej (ZSMGW) to złożona sieć składająca się z połączonych ze sobą rur i węzłów każdego podsystemu [8]. ZSMGW stanowi niezbędną infrastrukturę miejską i ma bezpośredni wpływ na gospodarkę publiczną, zdrowie i środowisko [17, 23]. ZSMGW składa się z dwóch zasadniczych części: systemu zaopatrzenia w wodę i systemu kanalizacji (por. rys. 1).



Rysunek 1. Zintegrowany system miejskiej gospodarki wodnej

System zaopatrzenia w wodę składa się z trzech podstawowych elementów: źródła zaopatrzenia, uzdatniania lub przetwarzania wody oraz dystrybucji wody

do społeczeństwa i przemysłu. Woda pochodząca ze źródła kierowana jest do oczyszczalni za pomocą akweduktów i rur. Proces ten zachodzi przy wykorzystaniu ciśnienia lub przepływu w otwartym kanale. Po oczyszczeniu woda jest natychmiast transportowana do systemu dystrybucyjnego, może być również przekazywana poprzez zbiorniki retencyjne [10, 12, 20, 25].

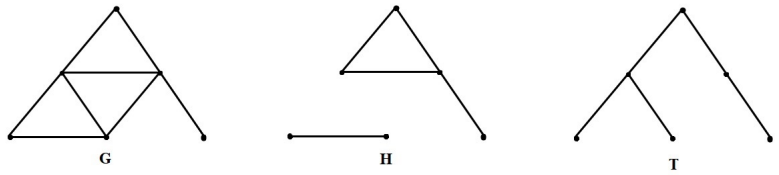
System kanalizacji to sieć rur, studzienek, przepompowni i obiektów z nimi związanych. Dzięki niemu ścieki są przekazywane z miejsca ich powstania do miejsca oczyszczania i unieszkodliwiania [6,21]. Natężenie przepływu ścieków różni się w zależności od miejsca, czynników ekonomicznych i charakteru publicznego, charakterystyki przedsiębiorstw przemysłowych na badanym obszarze, zużycia wody, warunków pogodowych i rodzaju kanalizacji. Istnieją trzy rodzaje systemów kanalizacyjnych: ścieki sanitarne (z łazienek, toalet, kuchni itp.), ścieki przemysłowe oraz woda deszczowa.

2. Opis wybranych algorytmów opartych na teorii grafów

Tradycyjne zintegrowane systemy miejskiej gospodarki wodnej opierają się na scentralizowanej infrastrukturze sieciowej. Ostatnio coraz częściej pojawia się dyskusja na temat sieci kanalizacyjnych w miastach. Najnowsze badania sugerują zastąpienie systemów scentralizowanych systemami zdecentralizowanymi lub hybrydowymi. W związku z tym potrzebne są narzędzia i metody do szacowania i optymalizacji sieci wodno-ściekowych o dowolnym stopniu centralizacji. W tym celu wykorzystuje się wiele algorytmów bazujących na teorii grafów.

W związku z tym przypomnijmy kilka podstawowych faktów i definicji związanych z tą teorią. Przez graf G rozumiemy trójkę zawierającą zbiór wierzchołków $V(G)$, zbiór krawędzi $E(G)$ oraz relację łączącą z każdą krawędzią dowolne dwa wierzchołki (niekoniecznie różne), nazwane punktami końcowymi. Graf prosty to taki graf, który ma nie więcej niż jedną krawędź pomiędzy dowolnymi dwoma wierzchołkami i żadna z krawędzi nie zaczyna się ani nie kończy w tym samym wierzchołku, czyli jest to graf bez krawędzi wielokrotnych i pętli. Ścieżka to graf prosty, którego wierzchołki można uporządkować w taki sposób, że dwa wierzchołki są połączone krawędzią, jeśli są kolejne w ciągu. Podgrafem danego grafu G nazywamy graf H spełniający inkluzje $V(H) \subseteq V(G)$ i $E(H) \subseteq E(G)$, gdzie przyporządkowanie punktów końcowych do krawędzi w grafie H jest identyczne jak w grafie G. Graf G nazywamy spójnym, jeżeli każda para wierzchołków jest połączona ścieżką. W przeciwnym razie G jest niespójny. Cykl to ścieżka, w której pierwszy i ostatni wierzchołek są połączone krawędzią. Graf nazywamy acyklicznym, jeśli nie zawiera żadnego cyklu. Spójny graf acykliczny nazywany jest drzewem. Podgrafem rozpinającym grafu G nazywamy podgraf ze zbiorem wierzchołków $V(G)$. Drzewo

rozpinające to podgraf rozpinający, będący drzewem [26,27]. Rysunek 2 przedstawia przykład grafu spójnego G, jego przykładowy podgraf rozpinający H, który jest niespójny, oraz przykładowe drzewo rozpinające T.



Rysunek 2. Przykłady grafów

2.1. Algorytm cięcia pętla po pętli

Algorytm cięcia pętla po pętli (ang. *the loop-by-loop algorithm*) został wprowadzony w 2013 roku przez A. Hanghighi'ego w celu zaprojektowania wykonalnych układów kanałów kanalizacyjnych na podstawie grafu bazowego [14]. Pierwszym krokiem tego algorytmu jest wprowadzenie nieskierowanego grafu bazowego do konfiguracji obejmującej wszystkie możliwe połączenia i rury. W grafie bazowym przez m jest oznaczona liczba krawędzi (liczba rur), n oznacza liczbę wierzchołków (włazów), zaś NL to liczba pętli. Innymi słowy, aby zbudować drzewo rozpinające z grafu bazowego, który ma NL pętli, należy wyciąć NL krawędzi (po jednej krawędzi w każdej pętli). Według Hanghighi'ego przez pętle rozumiemy cykle w grafie. Wszystkie wierzchołki, krawędzie i pętle zawarte w grafie bazowym są dowolnie oznaczane liczbami. Graf ten jest opisany macierzą B, która zawiera m wierszy i $NL+3$ kolumn.

Struktura macierzy B:

— pierwsza kolumna zawiera nazwy ścieków,
— kolumny od 2 do $NL+1$ składają się ze wskaźników kanalizacji w pętli, które wskazują, czy kanalizacja znajduje się w pętli, czy nie,
— w końcowych kolumnach od $NL+2$ do $NL+3$ znajdują się nazwy studzienek kanalizacyjnych (nazwy zakończeń kanalizacyjnych).

Aby zbudować układ typu drzewo, należy w grafie bazowym wyciąć jeden kanał danej pętli. Po wybraniu rury do cięcia, cięcie można przeprowadzić zarówno w studzience górnej, jak i dolnej. W związku z tym można wymienić dwa parametry decyzyjne dotyczące uruchomienia każdej pętli, które zawierają nazwę przecinanej rury oraz miejsce przecięcia. Są one oznaczone odpowiednio przez α i β. Po uruchomieniu pętli modyfikuje się graf bazowy, zaś macierz B wymaga odpowiedniej modyfikacji (co opisano szczegółowo w [3, 14]). Po wprowadzeniu powyższych

modyfikacji rozważana jest kolejna pętla i proces trwa do momentu otwarcia wszystkich NL pętli. Ostatecznie, na podstawie określonych parametrów α i β, tworzony jest układ kanalizacji obejmujący m kanałów, $n+NL$ studzienek i niezawierający pętli. Następnie wyznacza się kierunki kanalizacji w stronę studzienki wylotowej w oparciu o zasadę, że „z każdej studzienki, z wyjątkiem wylotu, wychodzi dokładnie jeden kanał". Kierunek każdej kanalizacji zmienia się w macierzy B w taki sposób, że górne studzienki umieszczane są w kolumnie $NL+1$, a dolne w kolumnie $NL+2$, por. [3, 4, 14].

2.2. Algorytm wiszących ogrodów

Algorytm wiszących ogrodów został wprowadzony w 2019 roku przez A. Hanghighi'ego, A.E. Bakhshipour'a i innych w celu stworzenia wszystkich możliwych układów kanalizacyjnych i zbadania różnych stopni decentralizacji [3]. Nazwa „algorytm wiszących ogrodów" wywodzi się z faktu, że algorytm ten „tnie" główne wiszące drzewo na wiele wiszących drzew o mniejszych rozmiarach. Algorytm ten tworzy zdecentralizowane układy przy użyciu następujących elementów:

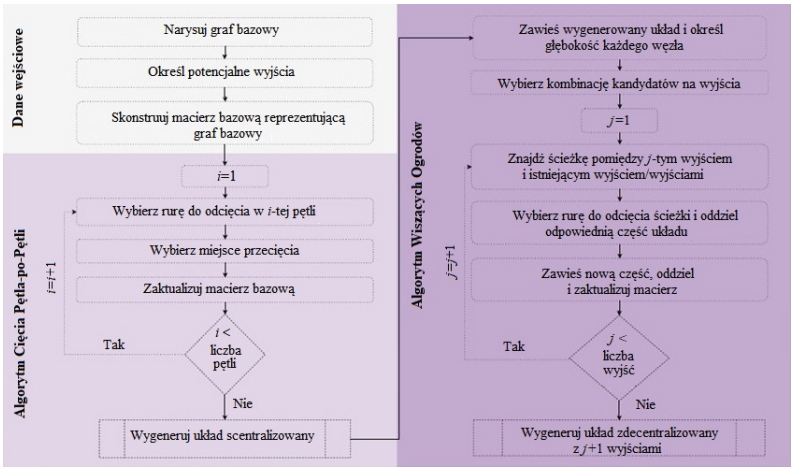
— tropiciel – sprawdzenie odległości pomiędzy sugerowanym dodatkowym nowym korzeniem a istniejącym(i) korzeniem(ami),
— separator – wybór miejsca odcięcia tej odległości i podzielenie grafu na dwie części,
— konstruktor macierzy – wyszukanie znaczących węzłów i rur we wszystkich częściach i utworzenie macierzy H (która opiera się na węzłach, podczas gdy macierz B opiera się na rurach) dla każdego z nowych drzew.

Macierz H zawiera n wierszy i $n+2$ kolumn. Struktura macierzy H jest następująca:

— pierwsza kolumna składa się z poziomów węzła,
— druga kolumna zawiera nazwy węzłów,
— kolumny od 3 do $n+2$ zawierają informację o połączeniu węzła, wskazując, czy węzeł jest powiązany z innymi, czy nie.

Stosując tę metodę, można podzielić każde drzewo skierowane na dwa drzewa, również skierowane. Wprowadzając nowy korzeń do systemu, algorytm najpierw sprawdza, do której części należy nowy korzeń, a następnie wykorzystuje opisane powyżej procedury do podziału drzewa. Proces ten można powtarzać do momentu, aż wszystkie możliwe korzenie (i wszystkie możliwe połączenia) znajdą się w ostatecznym układzie. Ostatnim etapem jest wykorzystanie wszystkich modyfikacji w rzeczywistym układzie w obrębie grafu bazowego [3–5].

Algorytm cięcia pętla po pętli tworzy układ scentralizowany. Z drugiej strony, algorytm wiszących ogrodów wykorzystuje utworzony układ do wygenerowania układu zdecentralizowanego. Proces ten przedstawiono na rysunku 3.



Rysunek 3. Algorytm służący do tworzenia układów zdecentralizowanych [3]

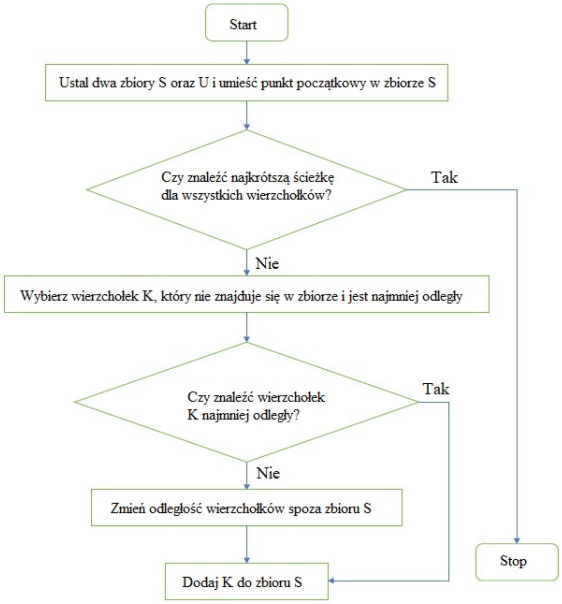
2.3. Algorytm Dijkstry

Algorytm Dijkstry (ang. *Dijkstra's algorithm*) został wprowadzony w 1959 roku przez E.W. Dijkstrę w celu znalezienia najkrótszej ścieżki pomiędzy wierzchołkami w grafie [11]. Algorytm ten rozpoczyna się w wybranym wierzchołku (węźle źródłowym) i eksploruje graf w celu znalezienia najkrótszej ścieżki łączącej ten wierzchołek i wszystkie pozostałe wierzchołki grafu. Procedura podąża najkrótszą zidentyfikowaną ścieżką pomiędzy dowolnym wierzchołkiem a wierzchołkiem źródłowym, a następnie zmienia te wielkości w przypadku odkrycia krótszej odległości. Gdy algorytm zidentyfikuje najkrótszą ścieżkę od wierzchołka źródłowego do innego wierzchołka, wówczas ten drugi jest oznaczony jako „odwiedzony" i dołączony do ścieżki. Procedura trwa do momentu wstawienia każdego wierzchołka grafu do ścieżki [18].

Ograniczeniem algorytmu Dijkstry jest to, że może on działać wyłącznie z grafami, które mają dodatnie wagi. Wynika to z faktu, że w trakcie całej procedury należy sumować wagi krawędzi tak, aby znaleźć najkrótszą ścieżkę. Algorytm nie będzie działać poprawnie, jeśli w grafie pojawi się waga ujemna. Kiedy wierzchołek zostanie oznaczony jako „odwiedzony", rzeczywista odległość prowadząca do tego wierzchołka zostanie opisana przez „najkrótszą ścieżkę" prowadzącą do osiągnięcia tego wierzchołka.

Schemat działania algorytmu Dijkstry przedstawiono na rysunku 4. Na rysunku tym zbiór S stanowi zbiór wszystkich wierzchołków, dla których nie znaleziono najkrótszej ścieżki, czyli zawiera on jedynie punkt początkowy. Zbiór U zawiera

wierzchołki inne niż punkt początkowy. Więcej informacji na temat algorytmu Dijkstry można znaleźć w [9, 18, 28].



Rysunek 4. Schemat blokowy algorytmu Dijkstry [18]

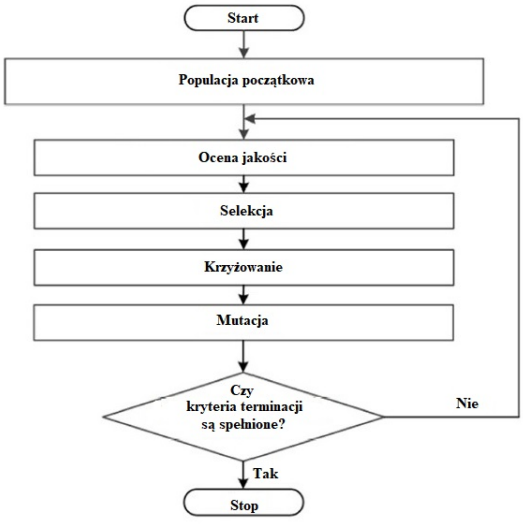
2.4. Algorytm genetyczny

W 1975 roku J. Holland i jego współpracownicy zaproponowali algorytm genetyczny (ang. *genetic algorithm*). Jego głównym celem było zbadanie procedury adaptacyjnej niektórych metod naturalnych i opracowanie metod symulujących ten proces [15]. Algorytmy genetyczne to powszechne metodologie poszukiwania sztucznej ewolucji, których istotą są mechanizmy ewolucji i genetyki populacyjnej. Metody te opierają się na naturalnej, wysoce wydajnej optymalizacji technologii ewolucyjnych, które bazują na preferowanym doświadczeniu i reprodukcji najlepiej dopasowanych uczestników populacji, zachowaniu populacji złożonej z różnych członków, dziedziczeniu informacji genetycznej od rodziców i okazjonalnych mutacjach genów.

Podstawowy algorytm genetyczny składa się z pięciu etapów: inicjalizacji, selekcji, krzyżowania, mutacji i terminacji [2].

— Inicjalizacja – zbiór genów składa się z populacji proponowanych rozwiązań. Ta metoda tworzy losowe sekwencje z poszczególnych rozwiązań w celu utworzenia populacji wstępnej. Inicjalizacja odbywa się w sposób losowy, aby uwzględnić całą gamę możliwych rozwiązań w obszarze poszukiwań.
— Selekcja – z żywej populacji wybierane są konkretne genomy w celu wychowania kolejnego pokolenia. Pojedyncze rozwiązania, zwane rozwiązaniami dopasowanymi, są wybierane przy użyciu funkcji dopasowania w ramach procedury opartej na dopasowaniu. Do oceny dopasowania wszystkich rozwiązań stosuje się metody wyboru.
— Krzyżowanie – ten operator genetyczny przyjmuje co najmniej jedno rozwiązanie rodzicielskie i tworzy rozwiązanie potomne. Pobierane są geny chromosomów rodzicielskich i powstaje więcej potomstwa. W szczególności operator wybiera losowy punkt przecięcia.
— Mutacja – operator ten modyfikuje pojedynczy bit w losowy sposób w rzeczywistym potomku utworzonym z operatora krzyżowania. Metoda ta wymaga stworzenia nowego przypuszczenia poprzez mutację dostępnego. Łańcuch nadrzędny jest przekształcany w tabelę składającą się z ciągów nadrzędnych.
— Zakończenie – cały proces jest powtarzany do momentu odkrycia zanotowanego rozwiązania lub populacji n-iteracji, która nie uległa zmianie.

Schemat działania algorytmu przedstawiono na rysunku 5. Więcej informacji na temat algorytmu genetycznego można znaleźć w [2].



Rysunek 5. Schemat blokowy standardowego algorytmu genetycznego [2]

3. Zastosowanie wybranych algorytmów do optymalizacji zintegrowanego systemu miejskiej gospodarki wodnej

3.1. Algorytm cięcia pętla po pętli

Algorytm cięcia pętla po pętli wprowadzono w celu utworzenia możliwych układów kanalizacji na podstawie grafu bazowego. Za pomocą tej metody regularnie rozwiązywane są całkowite ograniczenia problemu instalacji kanalizacyjnej. Optymalny układ osiągany jest przy użyciu funkcji celu i zastosowaniu prostego algorytmu genetycznego. Następnie ustalana jest konfiguracja kanalizacji, a opisy pomp i rur są projektowane przy użyciu dyskretnego różnicowego modelu programowania dynamicznego. Algorytm cięcia pętla po pętli jest szczególnie pomocny przy projektowaniu miejskich systemów odwadniających na terenach płaskich. Ponadto algorytm ten jest efektywny obliczeniowo, a także prosty w zastosowaniu i włączeniu do rozwiązań optymalizacyjnych [3, 14].

3.2. Algorytm wiszących ogrodów

Algorytm wiszących ogrodów jest przystosowany do spełnienia wszystkich ograniczeń związanych z projektowaniem miejskich sieci rur kanalizacyjnych. Algorytm wiszących ogrodów został wprowadzony w celu wytworzenia i optymalizacji (zde)centralizowanych systemów odwadniających w miastach dla obszarów stromych i płaskich [3]. Algorytm ten był używany przez Hanghighi'ego i innych w odniesieniu do Ahvaz, miasta w Iranie. Pomimo tego, że ich wyniki pokazały, że zasugerowana metoda może zarówno stworzyć rzeczywiste struktury, jak i dać niemal optymalne odpowiedzi, badacze jedynie uwzględnili koszty budowy jako funkcję celu [4, 5].

3.3. Algorytm Dijkstry

W wyniku burz duża ilość wody wypływa z kanalizacji poprzez przelew ogólnospławny. Następnie dla każdego źródła (węzła) wyznaczana jest droga do określonego przelewu (celu) w ściekach. Każda ścieżka ma określoną długość lub „koszt". Dla każdego węzła najtańsza ścieżka do połączonej struktury przelewowej wyznaczana jest przez algorytm Dijkstry. Implementacja tej metody wymaga rozłożenia kosztów pomiędzy dowolne dwa węzły. Ponieważ kanalizacja przepływa głównie grawitacyjnie, ujemne nachylenia zdecydowanie podnoszą koszty wydobycia. W związku z tym można zastosować zmodyfikowaną długość do wyrażenia kosztu pomiędzy dowolnymi dwoma węzłami [16, 19].

3.4. Algorytm genetyczny

Algorytm genetyczny został wprowadzony w celu optymalizacji projektowania sieci wód deszczowych. W modelu tym za zmienne decyzyjne przyjmuje się rzędne węzłowe sieci kanalizacyjnej. Do sprawdzenia rozwiązań eksperymentalnych podanych przez optymalizator algorytmu genetycznego wykorzystywany jest kod symulacji stanu ustalonego. Algorytmy te są zasadniczo budowane z myślą o nieograniczonych problemach optymalizacyjnych. Zastosowanie algorytmu genetycznego do zagadnień optymalizacji ograniczonej, czyli sieci wód deszczowych, wymaga przekształcenia zagadnienia podstawowego ograniczonego w zagadnienie optymalizacji nieograniczonej [1].

Algorytm genetyczny może być również stosowany jako narzędzie optymalizacyjne w odniesieniu do ścieków sanitarnych. Celem jest minimalizacja całkowitego kosztu, zminimalizowanie czasu przebywania wody w systemie, a także maksymalizacja metryki niezawodności sieci [13].

Algorytm genetyczny jest metodą stosowaną również w celu optymalizacji systemu zaopatrzenia w wodę. Model ten stanowi istotny element inteligentnej konfiguracji diagnostycznej stosowanej w lokalnej sieci wodociągowej. Główną rolą tego procesu jest wykrywanie i lokalizacja wycieków wody. W przypadku wejść konstrukcja wykorzystuje dane z czujników przepływu lub czujników ciśnienia zainstalowanych na sieci rurociągów, natomiast wyjściem jest część danych dotyczących wykrycia wycieku oraz jego położenia. Główną zaletą tego systemu jest możliwość oszacowania lokalizacji wycieku jedynie na podstawie skończonej liczby montowanych czujników [7].

Podsumowanie

Zaprezentowane algorytmy oparte na teorii grafów mają szerokie zastosowanie w modelowaniu zagadnień dotyczących zintegrowanych systemów miejskiej gospodarki wodnej. Główną rolą tych algorytmów jest optymalizacja procesów zachodzących w poszczególnych podsystemach ZSMGW, m.in. w sieciach kanalizacyjnych oraz w systemie zaopatrzenia w wodę. Możliwym kierunkiem dalszych badań może stać się wykorzystanie do modelowania zagadnień systemów gospodarki wodnej obiektów bardziej uniwersalnych niż grafy. Mianowicie są to hipergrafy, których właściwości i zalety wykorzystywane są w modelowaniu złożonych układów, m.in. w chemii, elektrotechnice i mechanice.

Literatura

[1] Afshar M.H., *Application of a Genetic Algorithm to Storm Sewer Network Optimization*, „Scientia Iranica", 2006, vol. 13(3), p. 234–244.

[2] Albadr M.A., Tiun S., Ayob M., AL-Dhief F., *Genetic Algorithm Based on Natural Selection Theory for Optimization Problems*, „Symmetry", 2020, vol. 12(11), p. 1758–1788.

[3] Bakhshipour A.E., Bakhshizadeh M., Dittmer U., Haghighi A., Nowak W., *Hanging Gardens Algorithm to Generate Decentralized Layouts for the Optimization of Urban Drainage Systems*, „Journal of Water Resources Planning and Management", 2019, vol. 145(9), p. 1–12.

[4] Bakhshipour A.E., Hespen J., Haghighi A., Dittmer U., Nowak W., *Integrating Structural Resilience in the Design of Urban Drainage Networks in Flat Areas Using a Simplified Multi-Objective Optimization Framework*, „Water", 2021, vol. 13, p. 269–290.

[5] Bakhshipour A.E., Dittmer U., Haghighi A., Nowak W., *Hybrid Green-blue-gray Decentralized Urban Drainage Systems Design, a Simulation-optimization Framework*, „Journal of Environmental Management", 2019, vol. 249, p. 1–13.

[6] Berko A., Zhuk V., Sereda I., *Modeling of Sewer Networks by Means of Directed Graphs*, „Environmental Problems", 2017, vol. 2(2), p. 97–100.

[7] Bi W., Dandy G.C., Maier H.R., *Improved Genetic Algorithm Optimization of Water Distribution System Design by Incorporating Domain Knowledge*, „Environmental Modelling and Software", 2015, vol. 69, p. 370–381.

[8] Butler D., Davies J.W., *Urban Drainage*, Spon Press, 2004.

[9] Cormen T.H., Leiserson C.E., Rivest R.L., Stein C., *Introduction to Algorithms*, The MIT Press, Cambridge, 2009.

[10] Deuerlein J., Piller O., Montalvoc I., *Improved Real-Time Monitoring and Control of Water Supply Networks by Use of Graph Decomposition*, „Procedia Engineering", 2014, vol. 89, p. 1276–1281.

[11] Dijkstra E.W., *A Note on Two Problems in Connexion with Graphs*, „Numerische Mathematik", 1959, vol. 1, p. 269–271.

[12] Giudicianni C., Di Nardo A., Di Natale M., Greco R., Santonastaso G.F., Scala A., *Topological Taxonomy of Water Distribution Networks*, „Water", 2018, vol. 10(4), p. 444–462.

[13] Hassan W.H., Attea Z.H., Mohammed S.S., *Optimum Layout Design of Sewer Networks by Hybrid Genetic Algorithm*, „Journal of Applied Water Engineering and Research", 2020, vol. 8(2), p. 1–17.

[14] Haghighi A., *Loop-by-Loop Cutting Algorithm to Generate Layouts for Urban Drainage Systems*, „Journal of Water Resources Planning and Management", 2013, vol. 139(6), p. 693–703.

[15] Holland J.H., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, The MIT Press, Cambridge, 1975.

[16] Huang J., James W., James W.R.C., *A Lifecycle Cost-based Design Optimization Model for Stormwater Management Systems*, „Journal of Water Management Modeling", 2005, p. 53–70.
[17] Hvitved-Jacobsen T., *Sewer Processes: Microbial and Chemical Process Engineering of Sewer Networks*, CRC Press, Boca Raton, Floryda, 2002.
[18] Li-sang Liu, Jia-feng Lin, Jin-xin Yao, Dong-wei He, Ji-shi Zheng, Jing Huang, Peng Shi, *Path Planning for Smart Car Based on Dijkstra Algorithm and Dynamic Window Approach*, „Wireless Communications and Mobile Computing", 2021, vol. 2021, p. 1–12.
[19] Meijer D., van Bijnen M., Langeveld J., Korving H., Post J., Clemens F., *Identifying Critical Elements in Sewer Networks Using Graph-Theory*, „Water", 2018, vol. 10(2), p. 136–164.
[20] Di Nardo A., Giudicianni C., Greco R., Herrera M., Santonastaso G.F., *Applications of Graph Spectral Techniques to Water Distribution Network Management*, „Water", 2018, vol. 10(1), p. 45–60.
[21] Pelda J., Holler S., *Methodology to Evaluate and Map the Potential of Waste Heat from Sewage Water by Using Internationally Available Open Data*, „Energy Procedia", 2018, vol. 149, p. 555–564.
[22] Pfister A., Stein A. Schlegel S., Teichgraber B., *An Integrated Approach for Improving the Wastewater Discharge and Treatment System*, „Water Science and Technology", 1998, vol. 37(1), p. 341–346.
[23] Reyes-Silva J.D., Zischg J., Klinkhamer C., Rao P.S.C., Sitzenfrei R., Krebs P., *Centrality and Shortest Path Length Measures for the Functional Analysis of Urban Drainage Networks*, „Applied Network Science", 2020, vol. 5(1), p. 1–14.
[24] Szeląg B., Drewnowski J., Łagód G., Majerek D., Dacewicz E., Fatone F., *Soft Sensor Application in Identification of the Activated Sludge Bulking Considering the Technological and Economical Aspects of Smart Systems Functioning*, „Sensors", 2020, vol. 20(7), p. 1941–1965.
[25] Ulusoy A.J., Stoianov I., Chazerain A., *Hydraulically Informed Graph Theoretic Measure of Link Criticality for the Resilience Analysis of Water Distribution Networks*, „Applied Network Science", 2018, vol. 3(31), p. 1–22.
[26] West D.B., *Introduction to Graph Theory*, Prentice Hall, Michigan, 2001.
[27] Wilson R.J., *Introduction to Graph Theory*, Longman, Harlow, 1998.
[28] Wojciechowski J., Pieńkosz K., *Grafy i sieci*, Wydawnictwo Naukowe PWN, Warszawa, 2013.

Wiktoria Bielak [1], **Ewa Łazuka** [2]

Wykorzystanie teorii grafów w harmonogramowaniu projektów budowlanych

Streszczenie

Rozdział poświęcony został wykorzystaniu teorii grafów w planowaniu projektów. Dzięki grafom możliwe jest konstruowanie dobrych modeli harmonogramów projektów, które uwzględniają czas realizacji zadań, zestawienie materiałów, zarządzanie jakością oraz analizę ryzyka. Modele grafowe pozwalają także na określenie zależności występujących w takich harmonogramach. W rozdziale opisano trzy najważniejsze metody stosowane przy planowaniu projektów: CPM, PERT oraz GERT, a następnie zaprezentowano ich działanie na przykładzie konkretnego harmonogramu projektu budowlanego.

Słowa kluczowe: *graf, projekt budowlany, CPM, PERT, GERT*

Abstract

The chapter is devoted to the use of graph theory in project planning. Thanks to the use of graphs, it is possible to construct good models of project schedules that take into account task completion time, bill of materials, quality management and risk analysis. Moreover, graph models help determine relationships occurring in such schedules. The chapter describes three most important methods used in project planning: CPM, PERT and GERT, and then presents how they work on the example of a specific construction project schedule.

Keywords: *graph, construction project, CPM, PERT, GERT*

Wstęp

Teoria grafów jest dziedziną matematyki zajmującą się badaniem struktur grafowych. Graf to pojęcie, które może być używane do opisywania szerokiej gamy zagadnień, w tym związków między obiektami, struktur sieciowych, problemów przydziału i wielu innych. Teoria grafów może być używana do opracowywania i rozwiązywania wielu problemów w różnych dziedzinach, m.in. w informatyce, matematyce, biologii i statystyce. Współczesne zastosowania teorii grafów obejmują wiele różnorodnych zagadnień, od analizy sieci społecznych po modelowanie procesów produkcyjnych i zarządzanie zasobami.

Celem rozdziału jest przedstawienie możliwości wykorzystania modelowania wybranych problemów inżynierskich za pomocą grafów w planowaniu projektów. Rozwój metod planowania projektów jest ważnym obszarem badawczym. Zarówno planowanie projektu, jak i planowanie rozwoju produktu uwzględnia czas realizacji

[1] Absolwentka studiów I stopnia na kierunku matematyka, Politechnika Lubelska
[2] Katedra Matematyki Stosowanej, Politechnika Lubelska

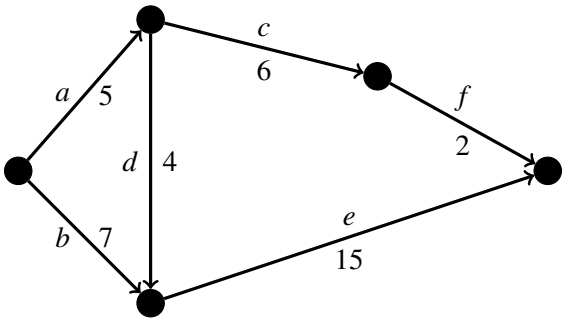
zadań, zestawienie materiałów, zarządzanie jakością, a także analizę ryzyka. Takie planowanie może być wspierane przez teorię grafów.

Najbardziej znane metody grafowe stosowane w tym kontekście to CPM (*critical path method*), PERT (*program evaluation and review technique*) oraz GERT (*graphical evaluation and review technique*). Wspólną cechą tych metod jest podział projektu na wiele dobrze zdefiniowanych i nienakładających się na siebie pojedynczych zadań, zwanych czynnościami. Ze względu na ograniczenia techniczne niektóre zadania muszą zostać ukończone przed rozpoczęciem innych. Oprócz tej relacji pierwszeństwa między czynnościami, każda czynność wymaga również określonego czasu, zwanego czasem trwania czynności. Biorąc pod uwagę listę działań w projekcie oraz czas ich trwania, możemy narysować ważony digraf przedstawiający projekt w następujący sposób: każda krawędź reprezentuje aktywność, a jej waga reprezentuje czas trwania aktywności. Wierzchołki reprezentują początki i końce działań i są nazywane zdarzeniami w projekcie. Każde działanie zaczynające się zdarzeniem i, a kończące zdarzeniem j będzie reprezentowane przez krawędź skierowaną (i, j).

Jeżeli mówimy o dużych, skomplikowanych i złożonych projektach, to przy stosowaniu wymienionych metod możemy mówić także o sieciach.

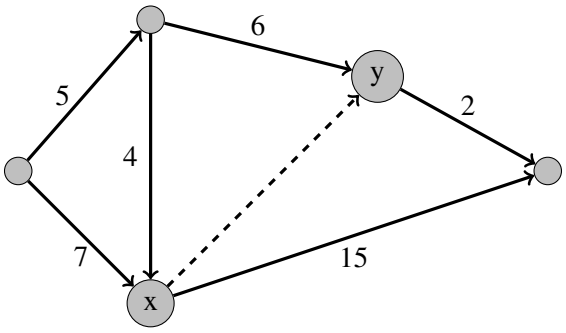
1. Tworzenie grafu aktywności

Tworzenie grafu odpowiadającego projektowi wyjaśnimy na podstawie przykładu zaczerpniętego z literatury [3]. Załóżmy, że mamy projekt składający się z sześciu działań a, b, c, d, e oraz f. Działania są skonstruowane w ten sposób, że a poprzedza c i d, b i d poprzedzają e, zaś c musi wystąpić przed f. Czas trwania działań a, b, c, d, e oraz f wynosi odpowiednio 5, 7, 6, 4, 15 i 2 dni. Graf aktywności tego projektu przedstawiono na rysunku 1.



Rysunek 1. Graf aktywności projektu
Źródło: Opracowanie własne

Można zauważyć, że graf aktywności musi być acykliczny. W przeciwnym razie wystąpiłaby sytuacja, w której żadna aktywność w grafie skierowanym nie mogłaby zostać zainicjowana. Na rysunku 1 widzimy, że wierzchołek oznaczający początek projektu musi mieć zerowy stopień wejściowy, ponieważ żadna aktywność nie poprzedza tego wierzchołka. Podobnie, wierzchołek oznaczający zakończenie projektu musi mieć zerowy stopień wyjściowy, ponieważ żadna aktywność nie następuje po tym wierzchołku.

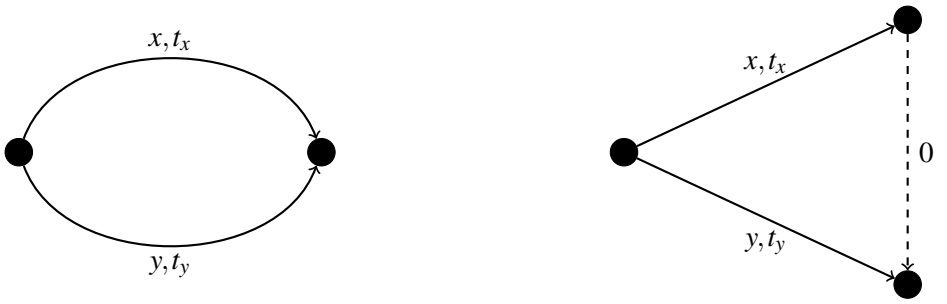


Rysunek 2. Aktywność fikcyjna w grafie aktywności projektu
Źródło: Opracowanie własne

W takim grafie aktywności może wystąpić tzw. aktywność fikcyjna. Należy wtedy wprowadzić założenie, że istnieje dodatkowe ograniczenie mówiące, że aktywność f nie może zostać rozpoczęta przed ukończeniem aktywności b i d. Taką sytuację uwzględnia się, rysując krawędź od wierzchołka x do y, tak jak na rysunku 2. Taka krawędź, która reprezentuje tylko relację pierwszeństwa, a nie żadne zadanie w projekcie, nazywana jest czynnością fikcyjną. Fikcyjne działania stają się konieczne, gdy istniejące działania nie są wystarczające do dokładnego przedstawienia wszystkich relacji pierwszeństwa. Aktywności fikcyjne mają zerowy czas trwania i są zwykle przedstawiane w postaci przerywanych linii.

Dwie równoległe krawędzie (tj. działania mające tego samego bezpośredniego poprzednika i tego samego bezpośredniego następnika) można zastąpić pojedynczą krawędzią, łącząc oba działania w jedno, tak jak na rysunku 3, gdzie x, y oznaczają działania, a t_x, t_y oznaczają odpowiednio czas trwania działania x i y. Jeśli jednak działania powinny być śledzone oddzielnie, należy utworzyć atrapę działania i atrapę zdarzenia, tak jak na rysunku 4.

Rysunek 3. Zastąpienie dwóch równoległych krawędzi jedną krawędzią
Źródło: Opracowanie własne



Rysunek 4. Zastąpienie dwóch równoległych krawędzi – atrapa działania i atrapa zdarzenia
Źródło: Opracowanie własne

Możemy założyć, że graf aktywności ma dokładnie jeden wierzchołek z zerowym stopniem wejściowym i dokładnie jeden wierzchołek z zerowym stopniem wyjściowym. Jeśli istnieje więcej niż jeden wierzchołek o zerowym stopniu wejściowym, można arbitralnie wybrać jeden z nich jako zdarzenie początkowe i narysować atrapy aktywności z tego wierzchołka do innych wierzchołków. Wierzchołki z zerowymi stopniami wyjściowymi są traktowane podobnie. Graf aktywności jest więc reprezentacją dwóch aspektów projektu: relacji pierwszeństwa między działaniami oraz czasu trwania tych działań.

W niektórych przypadkach planowania przydatne okazuje się połączenie metod planowania opartych na teorii grafów z metodą zarządzania ryzykiem. Niepewność i ryzyko związane z projektami pochodzą z różnych źródeł i muszą być odpowiednio modelowane. Podejmując decyzję o rozwoju projektu, należy wziąć pod uwagę kilka kryteriów jednocześnie. Najważniejsze z nich to:

— kryteria finansowe – projekt powinien pochłaniać ograniczone koszty,
— kryteria czasowe – projekt powinien zostać ukończony w określonym czasie,
— kryteria wyników – projekt powinien koncentrować się na określonych wynikach.

Aby móc zintegrować wielowymiarowe kryteria realizacji i oceny projektu, konieczne jest zagregowanie kryteriów oceny i połączenie ich ze stochastyczną metodą planowania projektu. Stochastyczne metody planowania pomagają tworzyć harmonogramy, które uwzględniają wszystkie możliwe rozwiązania przed określeniem ostatecznego harmonogramu. Typową metodą planowania stochastycznego jest metoda GERT. Natomiast w algorytmie PERT probabilistyczny charakter mają jedynie parametry opisujące czynności, a struktura grafu jest zdeterminowana, podczas gdy metoda CPM jest typową metodą deterministyczną.

2. Critical path method

Critical path method (CPM) [5, 9] to matematyczny algorytm planowania zestawu działań projektowych. Służy do znalezienia tzw. ścieżki krytycznej i opisania zależności pomiędzy działaniami. Narzędzie to pozwala określić, które zadania lub czynności, spośród wielu, które składają się na projekt, są „krytyczne" pod względem ich wpływu na całkowity czas realizacji projektu oraz jak najlepiej zaplanować wszystkie zadania, aby dotrzymać docelowego terminu przy minimalnych kosztach. Jest to ważne narzędzie do skutecznego zarządzania projektami. Metoda CPM pozwala na analizę bardzo zróżnicowanych rodzajów projektów, między innymi w budownictwie, rozwoju oprogramowania, projektach badawczych, rozwoju produktów, inżynierii i konserwacji instalacji. Każdy projekt ze współzależnymi działaniami może zastosować tę metodę planowania. Aby projekty mogły być analizowane dzięki metodzie CPM, niezbędne jest, by składały się z dobrze zdefiniowanego zbioru zadań, które mogą być uruchamiane i zatrzymywane w określonej kolejności niezależnie od siebie. Zadania muszą być uporządkowane i wykonywane w odpowiedniej sekwencji technologicznej.

Koncepcja CPM jest dość prosta i najlepiej można ją zilustrować za pomocą grafu projektu. Graf nie jest istotną częścią CPM, ponieważ istnieją programy komputerowe, które pozwalają na wykonanie niezbędnych obliczeń bez odwoływania się do grafów. Jednak są one cenne jako sposób na wizualne i wyraźne przedstawienie kompleksu zadań w projekcie i ich wzajemnych powiązań. Algorytm CPM wyznacza najdłuższą ścieżkę zaplanowanych działań do końca projektu oraz najwcześniejsze i najpóźniejsze rozpoczęcie i zakończenie każdego działania bez wydłużania projektu.

Metoda CPM ma trzy główne etapy:

1. podział projektu na operacje niezbędne do jego ukończenia oraz określenie sekwencyjnego powiązania operacji,
2. tworzenie szacunków czasu dla każdej operacji,

3. określenie najwcześniejszej możliwej daty rozpoczęcia, najwcześniejszej możliwej daty zakończenia, najpóźniejszej daty rozpoczęcia i zakończenia poszczególnych działań, a w konsekwencji całego projektu.

Każde działanie niezbędne do ukończenia projektu jest wymienione z unikalnym symbolem identyfikacyjnym (takim jak litera lub cyfra), czasem wymaganym do ukończenia tego zadania i jego zadaniami wstępnymi. Aby ułatwić tworzenie grafów i sprawdzanie niektórych rodzajów błędów w danych, zadania mogą być ułożone w kolejności technologicznej, to znaczy, że żadne nie pojawiają się na grafie, dopóki nie zostaną uwzględnieni ich poprzednicy. Kolejność technologiczna jest niemożliwa, jeśli w pewnych zadaniach występuje błąd cyklu (np. zadanie a poprzedza b, b poprzedza c, zaś c poprzedza a). Następnie każde zadanie jest rysowane w grafie jako okrąg, z jego symbolem identyfikacyjnym i czasem pojawiającym się wewnątrz okręgu. Relacje sekwencji są oznaczone strzałkami w każdym okręgu prowadzącym do jego bezpośrednich następców. Dla wygody wszystkie okręgi bez poprzedników są połączone z okręgiem oznaczonym „Start".

Graf jest szeregiem różnych ścieżek od punktu początkowego do końcowego. Czas wymagany do przejścia każdej ścieżki jest sumą czasów związanych ze wszystkimi zadaniami na ścieżce. Ścieżka krytyczna wyznacza możliwie najkrótszy czas niezbędny do ukończenia całego projektu. Koncepcja CPM zakłada, że co najmniej jedna ścieżka w grafie aktywności określa czas trwania całego projektu i jest ona jest ścieżką krytyczną, a działania na niej nazywane są działaniami krytycznymi. Metoda CPM jest metodą deterministyczną, która nie uwzględnia niepewności i prawdopodobieństwa ukończenia każdego działania w projekcie. Tylko poprzez znalezienie sposobów na skrócenie zadań wzdłuż ścieżki krytycznej można skrócić całkowity czas projektu.

Gdy znana jest kolejność działań, można przejść do wykonywania obliczeń czasu trwania całego projektu, rozpoczynając od zdarzenia początkowego na zdarzeniu końcowym kończąc. Chodzi o to, aby wyznaczyć najwcześniej ukończone zdarzenie w grafie oraz obliczyć najwcześniejszy czas rozpoczęcia i zakończenia działań. Początkowo jest dane $TE_{(i)} = ES_{(i,j)} = 0$,

$$EF_{(i,j)} = ES_{(i,j)} + t_{(i,j)}, \tag{2.1}$$

gdzie:

i, j – numery poszczególnych zdarzeń w projekcie,

$TE_{(i)}$ – najwcześniej ukończone zdarzenie i,

$ES_{(i,j)}$ – najwcześniejszy czas rozpoczęcia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$EF_{(i,j)}$ – najwcześniejszy czas zakończenia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$t_{(i,j)}$ – czas wymagany do wykonania aktywności, która rozpoczyna się zdarzeniem i, a kończy zdarzeniem j.

Po wykonaniu powyższych obliczeń przechodzimy do obliczania czasu, rozpoczynając od zdarzenia końcowego, a kończąc na zdarzeniu początkowym. Celem jest wyznaczenie najpóźniej ukończonego zdarzenia w grafie oraz obliczenie najpóźniejszego czasu rozpoczęcia i zakończenia działań.

$$LS_{(i,j)} = LF_{(i,j)} - t_{(i,j)}, \tag{2.2}$$

$$LF_{(i,j)} = TL_{(j)},$$

gdzie:

i, j – numery poszczególnych zdarzeń w projekcie,

$LS_{(i,j)}$ – najpóźniejszy czas rozpoczęcia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$LF_{(i,j)}$ – najpóźniejszy czas zakończenia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$t_{(i,j)}$ – czas wymagany do wykonania aktywności, która rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$TL_{(j)}$ – najpóźniej ukończone zdarzenie j.

Następnym krokiem jest znalezienie całkowitego przepływu (ang. *total float*) i wolnego przepływu (ang. *free float*) w celu określenia ścieżki krytycznej. Całkowity przepływ (TF) odnosi się do maksymalnego czasu opóźnienia, jakie może wystąpić podczas wykonywania danej czynności bez wpływu na całkowity czas trwania projektu. To czas, w którym ukończenie działania może zostać przesunięte bez wpływu na najwcześniejszy czas ukończenia projektu jako całości. Jest to różnica między najpóźniejszym czasem wykonania danej czynności a najwcześniejszym czasem rozpoczęcia zadania, bez wpływu na opóźnienie całego projektu. Całkowity przepływ pozwala na określenie, jak elastyczny jest harmonogram i jakie rezerwy czasowe są dostępne w projekcie. Ten krok jest najdłuższą ścieżką realizacji, która określa czas ukończenia projektu. Czynności na ścieżce krytycznej nie mają czasu wolnego ($TF = 0$), co oznacza, że opóźnienie w którymkolwiek z tych zadań opóźni cały projekt. Wolny przepływ (FF) odnosi się do maksymalnego czasu opóźnienia, jakie może wystąpić podczas wykonywania danej czynności bez wpływu na rozpoczęcie następnej czynności w ścieżce krytycznej. Jest to różnica między najwcześniejszym czasem wykonania czynności a najwcześniejszym czasem rozpoczęcia zadania, przy zachowaniu ciągłości na ścieżce krytycznej. Wolny przepływ pozwala na określenie,

jakie opóźnienia można tolerować w danej czynności bez wpływu na czas trwania całego projektu.

Całkowity przepływ oraz wolny przepływ wyrażają się następującymi wzorami

$$TF_{(i,j)} = LF_{(i,j)} - ES_{(i,j)} - t_{(i,j)}, \tag{2.3}$$

$$FF_{(i,j)} = EF_{(i,j)} - ES_{(i,j)} - t_{(i,j)}, \tag{2.4}$$

gdzie:

i, j – numery poszczególnych zdarzeń w projekcie,

$TF_{(i,j)}$ – całkowity przepływ działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$LF_{(i,j)}$ – najpóźniejszy czas zakończenia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$ES_{(i,j)}$ – najwcześniejszy czas rozpoczęcia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$t_{(i,j)}$ – czas wymagany do wykonania aktywności, która rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$FF_{(i,j)}$ – wolny przepływ działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j,

$EF_{(i,j)}$ – najwcześniejszy czas zakończenia działania, które rozpoczyna się zdarzeniem i, a kończy zdarzeniem j.

3. Program evaluation and review technique

Program evaluation and review technique (PERT) [1, 2, 6] jest to technika zarządzania projektami, która rozpoczyna się od zdefiniowania celów projektu. Pozwala na oszacowanie czasu trwania działania i obliczenie prawdopodobieństwa zakończenia projektu w wyznaczonym czasie. Cechą PERT odróżniającą go od innych metod jest wykorzystanie teorii prawdopodobieństwa jako narzędzia predykcyjnego w prognozowaniu prawdopodobnego wyniku konkretnych planów. PERT jest metodą, która opiera się na założeniu, że jedna ścieżka określa czas trwania projektu, ale czas trwania każdej czynności jest określony przez funkcję gęstości prawdopodobieństwa. Pierwszym krokiem w tej metodzie jest struktura podziału pracy, tak aby stworzyć odgórne podejście do planowania. Proces ten polega na podzieleniu całego projektu na mniejsze i łatwiejsze do zarządzania elementy. Wszystkie mniejsze elementy powinny być ze sobą powiązane. Relacje zależności istniejące między komponentami obrazuje się za pomocą grafu. Cały przebieg podziału i klasyfikacji trwa aż do osiągnięcia określonego poziomu szczegółowości. Taki graf jest podstawą

systemu PERT, ponieważ pokazuje plan ustanowiony dla osiągnięcia celów projektu, wzajemne powiązania oraz współzależności elementów projektu i priorytety planu. Graf jest graficzną reprezentacją planu projektu. W przypadku złożonych projektów najpierw opracowywana jest sieć główna przedstawiająca cały program. W tym przypadku złożoność struktury podziału pracy jest podstawą do ustalenia liczby i rodzaju podsieci. Każda podsieć jest rozwinięciem szczegółów określonej sieci głównej.

W metodzie PERT cały graf składa się ze zdarzeń i działań. Zdarzenia reprezentują początek lub zakończenie działania i nie pochłaniają czasu ani zasobów, są to chwilowe momenty w czasie. Są one unikalne i mogą wystąpić w grafie tylko raz. Nie możemy powrócić do zdarzenia, które zostało dokonane, co oznacza, że w grafie nie mogą występować pętle. Zdarzenia są reprezentowane za pomocą okręgów.

Działanie to zadanie w projekcie wymagające personelu i zasobów, odbywające się przez pewien czas. Składa się ono z procesów prowadzących od jednego zdarzenia do drugiego. Opisy działań muszą być tak skonstruowane, aby można było przypisać do nich odpowiedni personel i dokonać realistycznego oszacowania czasu trwania działania, a także zrozumieć jego cel w schemacie. Działanie jest reprezentowane w sieci przez strzałkę łączącą zdarzenia, natomiast działania fikcyjne (niepochłaniające czasu ani zasobów) są oznaczane strzałkami przerywanymi.

Do budowy grafu można zastosować konstrukcję *backward*, to znaczy zacząć budowę od zdarzenia końcowego do zdarzenia początkowego. Jednak taka metoda ma sens tylko w przypadku, gdy elementy projektu są dość dobrze znane i zdefiniowane. Taka konstrukcja wsteczna zmusza użytkownika do rozważenia, jakie działania i zdarzenia muszą nastąpić zanim cele projektu będą mogły zostać osiągnięte. Należy jednak pamiętać, że w przypadku złożonych projektów przedstawionych jako sieć przepływ sieci zawsze przesuwa się od lewej do prawej.

Po zbudowaniu i zatwierdzeniu grafu, kolejnym etapem metody PERT jest ustalenie przewidywanego czasu wykonania każdego etapu. Typową procedurą są trzy oszacowania czasu w celu uwzględnienia niepewności dotyczącej zakresu prac: czasu optymistycznego, najbardziej prawdopodobnego oraz pesymistycznego.

Czas optymistyczny (t_o) to najkrótszy czas, w którym może zostać wykonana dana czynność w najbardziej optymalnych warunkach. Zakładamy, że działania związane z daną czynnością będą postępowały z wyjątkową łatwością i szybkością, ale jest małe prawdopodobieństwo zakończenia tego działania w tak krótkim czasie.

Najbardziej prawdopodobny czas (t_m) to taki, który w ocenie estymatora będzie zajmował wykonanie działania w normalnych okolicznościach.

Czas pesymistyczny (t_p) zakłada, że wszystkie działania mogą skończyć się niepowodzeniem. Szacunek w tym wypadku powinien uwzględniać najbardziej niekorzystne warunki, w tym możliwość awarii, które wymagają ponownego

uruchomienia poprzednich działań od początku grafu. Tutaj również istnieje małe prawdopodobieństwo, że czynność zostanie wykonana w takim czasie.

Zakres między optymistycznymi i pesymistycznymi szacunkami jest miarą zmienności, która pozwala na statystyczne wnioskowanie o prawdopodobieństwie wystąpienia zdarzeń projektowych w określonym czasie.

Koncepcja PERT opiera się na założeniu, że czas trwania każdego działania jest zmienną losową, która zachowuje się zgodnie z pewnym znanym i identycznym rozkładem prawdopodobieństwa. Wybranym rozkładem jest rozkład beta. Te trzy szacunki czasów są powiązane w formie rozkładu prawdopodobieństwa beta z parametrami t_o i t_p jako punktami końcowymi i t_m jako najczęstszą wartością. Rozkład beta jest tutaj używany, ponieważ jest jednomodalny i niekoniecznie symetryczny, a są to własności, które wydają się pożądane dla rozkładu czasu trwania aktywności.

Metoda PERT opiera się na centralnym twierdzeniu granicznym i oblicza oczekiwany czas trwania projektu jako sumę oczekiwanych czasów trwania działań na ścieżce krytycznej, podczas gdy wariancja czasu trwania projektu jest obliczana jako suma wariancji działań na ścieżce krytycznej. Umożliwia to generowanie statystycznych szacunków prawdopodobieństwa ukończenia projektu w określonym terminie.

W oparciu o rozkład beta i trzy oszacowania czasu, oczekiwany czas T_e trwania aktywności jest obliczany przy użyciu następującego wzoru

$$T_e = \frac{t_o + 4t_m + t_p}{6}, \tag{3.1}$$

gdzie:

t_o – czas optymistyczny,

t_m – najbardziej prawdopodobny czas aktywności,

t_p – czas pesymistyczny.

Po obliczeniu oczekiwanego czasu, jaki upłynął dla danej czynności, możemy też uzyskać oszacowanie zmienności związanych z nią szacunków czasu.

Zgodnie z centralnym twierdzeniem granicznym można wykazać, że czas trwania projektu jest zgodny z rozkładem normalnym, jednoznacznie zdefiniowanym przez parametry obliczone na podstawie parametrów każdego działania.

Suma dużej liczby niezależnych zmiennych losowych ma rozkład normalny, więc w projekcie, który składa się z listy działań, czas trwania projektu będzie zbliżony do rozkładu normalnego z wariancją V_{T_e} określoną wzorem

$$V_{T_e} = \left(\frac{t_p - t_o}{6}\right)^2. \tag{3.2}$$

Następnie określa się odchylenie standardowe działań w grafie, sumując wariancje aktywności na ścieżce krytycznej według poniższego wzoru

$$\sigma_{T_e} = \sqrt{\sum V_{T_e}}. \tag{3.3}$$

Czas trwania projektu ma rozkład normalny z funkcją gęstości prawdopodobieństwa określoną wzorem

$$f(T_e) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{T_e - \mu}{\sigma}\right)^2\right). \tag{3.4}$$

Do obliczenia prawdopodobieństwa ukończenia projektu w terminie wykorzystywana jest standaryzowana zmienna losowa Z wyrażona wzorem

$$Z = \frac{T_s - \sum T_e}{\sigma_{T_e}}, \tag{3.5}$$

gdzie:

T_s – oczekiwany czas trwania projektu,

$\sum T_e$ – suma oczekiwanych czasów trwania działań na ścieżce krytycznej,

σ_{T_e} – odchylenie standardowe działań na ścieżce krytycznej.

Dla ustalonej wartości tej zmiennej, korzystając z tablic rozkładu normalnego standaryzowanego, odczytujemy prawdopodobieństwo ukończenia projektu w terminie oraz przewidywaną liczbę dni trwania projektu.

Modele CPM oraz PERT dotyczą projektów, których struktura logiczna jest zdeterminowana. Oznacza to, że dla takich projektów konstrukcje grafu aktywności oparte są na określonych zestawach czynności wchodzących w skład analizowanego projektu. Składa się on z ciągu czynności, które są ze sobą powiązane i muszą być wykonywane w określonej kolejności. Takie zależności pomiędzy zdarzeniami określają strukturę logiczną modelu grafowego, która jest zdeterminowana, ponieważ w trakcie realizacji takiego projektu wszystkie czynności przedstawione w grafie muszą być zrealizowane, aby cały projekt mógł być zakończony.

4. Graphical evaluation and review technique

Metoda *graphical evaluation and review technique* [7, 8, 11–14], powszechnie znana jako GERT, jest to stochastyczna technika analizy grafów stosowana w zarządzaniu projektami, która pozwala na warunkowe i probabilistyczne traktowanie zarówno logiki grafu, jak i szacowania czasu trwania działań. GERT umożliwia tworzenie pętli między zadaniami i uwzględnia węzły probabilistyczne, a tym samym bierze pod uwagę stopień niepewności w projektach.

W metodzie GERT każda czynność jest zilustrowana przez łuk, a każdy łuk jest opisany przez dwa atrybuty: czas wymagany do wykonania danej czynności i prawdopodobieństwo, że ta czynność zostanie wykonana. Łuki są połączone węzłami logicznymi w grafie, który reprezentuje analizowany projekt. Węzły w tym grafie mogą mieć jednocześnie różne typy wejść i wyjść po to, aby móc wyrazić losowość wystąpienia czynności wchodzących do węzła, jak również prawdopodobieństwo tego, że czynność biorąca początek w tym węźle może, lecz nie musi wystąpić.

W metodzie GERT istnieje sześć typów węzłów w zależności od charakterystyki wejściowej i wyjściowej strony węzła.

W grafie stosowane są trzy typy węzłów wejściowych:

1. węzeł „i" – dla zdarzeń, które zachodzą wtedy i tylko wtedy, gdy wszystkie czynności je poprzedzające zostaną zrealizowane, tzn. pierwsze „i" drugie „i" następne. Wszystkie łuki prowadzące do tego węzła muszą zostać zrealizowane, aby potwierdzić osiągnięcie węzła. Czas realizacji jest najdłuższym z czasów wykonania zadań prowadzących do tego węzła.
2. węzeł „lub" – dla zdarzeń poprzedzonych kilkoma czynnościami moment realizacji zdarzenia jest zapewniony, gdy co najmniej jedna z czynności poprzedzających je zostanie zrealizowana, ale mogą zostać zrealizowane również inne czynności, dla których to zdarzenie jest zdarzeniem końcowym. Ten rodzaj węzła oznaczony jest w grafie jako kwadrat obrócony o 45°. Co ważne, czas realizacji jest najkrótszym z czasów realizacji zadań prowadzących do tego węzła.
3. węzeł „albo" – dla zdarzeń, które zachodzą wtedy i tylko wtedy, gdy dokładnie jedna z czynności poprzedzających je zostanie zrealizowana. Ten rodzaj węzła jest oznaczony w grafie jako kwadrat obrócony o 45° z linią prostopadłą do przekątnej kwadratu.

W grafie stosowane są dwa typy węzłów wyjściowych:

1. „i" – wyjście deterministyczne dotyczące zdarzeń, których wystąpienie powoduje realizację każdej czynności następującej po takich zdarzeniach. Po takim węźle musi występować jedna lub kilka czynności. Moment zajścia zdarzenia oznacza możliwość rozpoczęcia wszystkich tych czynności, tzn. że jeżeli taki węzeł zostanie osiągnięty, projekt będzie kontynuowany ze wszystkimi łukami, które rozpoczynają się w danym węźle. Prawdopodobieństwo wykonania wszystkich tych łuków, czyli zrealizowania każdej czynności wychodzącej z danego węzła, jest równe 1.
2. „lub" – wyjście probabilistyczne dotyczące zdarzeń, których wystąpienie powoduje realizację co najmniej jednej czynności po nich następującej. Z kilku wychodzących czynności zrealizowana może być jedna lub kilka, tzn. że można wybrać tylko jeden łuk z możliwych alternatyw. W takim przypadku suma

prawdopodobieństw wszystkich czynności wychodzących z tego węzła jest równa 1.

Rysunek 5 przedstawia wszystkie powyżej opisane typy wierzchołków.

Wyjście z wierzchołków / Wejście do wierzchołków		deterministyczne „i”	probabilistyczne „lub”
Zdarzenie wystąpi, jeżeli skończone zostaną wszystkie czynności poprzedzające „i”.			
Zdarzenie wystąpi, jeżeli skończy się którakolwiek z czynności poprzedzających „lub”.			
Zdarzenie wystąpi, jeżeli skończy się jedna i tylko jedna z czynności wzajemnie wykluczających się „albo”.			

Rysunek 5. Typy wierzchołków występujące w metodzie GERT
Źródło: M. Trocki, P. Wyrozębski, *Planowanie przebiegu projektów* [11]

Jeśli graf składa się z węzłów, które mają tylko wejścia „i” oraz deterministyczne wyjścia „i”, wówczas struktura grafu jest zredukowana do typowego scenariusza metod CPM/PERT i może być odpowiednio traktowana jako struktura deterministyczna. Graf zbudowany metodą GERT składa się z jednego węzła początkowego i jednego lub kilku węzłów końcowych, co oznacza, że możliwe są różne warianty zakończenia projektu. Zdarzenie początkowe może być zdarzeniem probabilistycznym, natomiast zdarzenie końcowe zawsze jest zdarzeniem deterministycznym. Cechą charakterystyczną modeli GERT są wyżej wspomniane pętle tzw. samosprzężenia, a także sprzężenia zwrotne. Pętle wskazują na to, że pewne działania mogą być wykonane więcej niż jeden raz. W grafie GERT krawędzie wskazują zadania, dla których przydzielane są zasoby projektu, takie jak czas lub koszty oraz prawdopodobieństwo.

Aby utworzyć graf GERT, należy wykonać poniższe kroki:

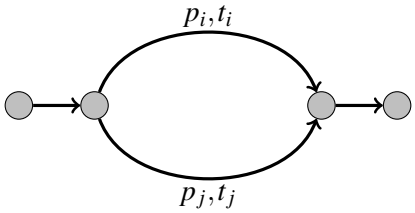
1. przekształcamy projekt w graf stochastyczny, wskazujemy zadania projektu i ustalamy relacje między nimi,
2. określamy dane dotyczące zadań projektu, ustalamy czas trwania zadań i ich prawdopodobieństwo,
3. określamy, które etapy projektu mogą być uproszczone lub zastąpione innymi etapami, co pozwoli uzyskać informacje na temat prawdopodobieństwa sukcesu projektu i przewidywanego całkowitego czasu projektu,
4. interpretujemy wyniki.

W przypadku techniki GERT, w celu przeprowadzenia analizy grafu, dokonuje się wielu kolejnych uproszczeń jego struktury. Prowadzą one do prostych połączeń czynności. Kolejne redukcje grafu polegają na stopniowym zastępowaniu łuków, rozgałęzień i pętli łukami zastępczymi. Po przeprowadzeniu odpowiedniej liczby etapów redukcji można otrzymać graf zredukowany do prostszej postaci lub, w niektórych przypadkach, nawet do pojedynczej czynności.

Poniżej przedstawiono charakterystyczne przypadki redukcji fragmentów grafu aktywności w metodzie GERT. Każdy łuk grafu jest opisany przez wektor $[p_i, t_i]$, gdzie:

— p jest prawdopodobieństwem realizacji łuku pod warunkiem, że węzeł, z którego on wychodzi, został zrealizowany,
— t oznacza czas trwania czynności odpowiadającej łukowi,
— i oznacza numer czynności w projekcie.

1. W pierwszym przypadku zakładamy, że projekt obejmuje dwie równolegle występujące czynności i oraz j połączone węzłami typu „i", które można zastąpić jednym łukiem r.



Rysunek 6. Przypadek pierwszy redukcji fragmentu grafu aktywności w metodzie GERT
Źródło: Opracowanie własne

W takiej sytuacji nowe prawdopodobieństwo dane jest wzorem

$$p_r = p_i \cdot p_j. \tag{4.1}$$

Gdy kilka czynności przebiega równolegle i wszystkie muszą być zrealizowane, aby projekt został ukończony, to oczekiwany czas trwania projektu t_r jest określany przez czynność o najdłuższym czasie trwania i dany wzorem

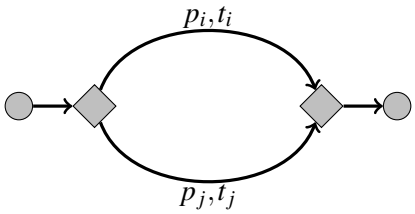
$$t_r = \max(t_i, t_j). \tag{4.2}$$

Natomiast w sytuacji, gdy kilka czynności przebiega równolegle, ale wystarczy, aby jedna z nich została zrealizowana, aby projekt został ukończony, to oczeki-

wany czas trwania projektu t_r określamy na podstawie czynności o najkrótszym czasie trwania i dany jest wzorem

$$t_r = \min(t_i, t_j). \tag{4.3}$$

2. Drugi przypadek jest podobny do pierwszego, ale z tym wyjątkiem, że zakładamy, że projekt obejmuje dwie czynności połączone węzłami typu „lub". Czynności występujące równolegle – i oraz j – można również zastąpić jednym łukiem r.



Rysunek 7. Przypadek drugi redukcji fragmentu grafu aktywności w metodzie GERT
Źródło: Opracowanie własne

Wtedy prawdopodobieństwo dane jest wzorem

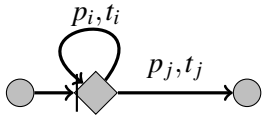
$$p_r = p_i + p_j, \tag{4.4}$$

a oczekiwany czas trwania projektu t_r uwzględniający węzły typu „lub" jest określony wzorem

$$t_r = \frac{p_i \cdot t_i + p_j \cdot t_j}{p_i + p_j}. \tag{4.5}$$

W tym przypadku spełnione jest również założenie, że wszystkie czynności wychodzące z węzła początkowego są od siebie niezależne.

3. Przypadek trzeci uwzględnia najbardziej charakterystyczną cechę metody GERT, czyli pętle. W tej sytuacji będzie to samosprzężenie.



Rysunek 8. Przypadek trzeci redukcji fragmentu grafu aktywności w metodzie GERT
Źródło: Opracowanie własne

W tym przypadku, gdy należy zredukować powstałą pętlę, prawdopodobieństwo realizacji łuku będziemy określać wzorem

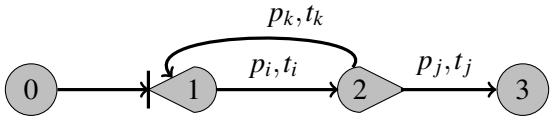
$$p_r = \frac{p_j}{1 - p_i}, \tag{4.6}$$

natomiast czas trwania czynności odpowiadającej temu łukowi jest dany wzorem

$$t_r = t_j + \frac{p_i \cdot t_i}{1 - p_i}. \tag{4.7}$$

Pętla pokazana w tym przypadku jako samosprzężenie jest stosowana w sytuacji, gdy podjęcie właściwych kroków, które w swoich zamierzeniach mają doprowadzić do założonych celów, nie doprowadzają do oczekiwanego stanu i należy powtórzyć tę czynność.

4. Przypadek czwarty również uwzględnia bardzo charakterystyczną cechę metody GERT. Jest nią sprzężenie zwrotne, czyli sytuacja cofnięcia się do zdarzenia oddalonego o więcej niż jedną czynność.



Rysunek 9. Przypadek czwarty redukcji fragmentu grafu aktywności w metodzie GERT
Źródło: Opracowanie własne

Dla fragmentu grafu zawierającego sprzężenie zwrotne jest możliwość przeprowadzenia redukcji, w wyniku której otrzymany układ będzie układem bez sprzężenia zwrotnego i samosprzężenia. W przypadku takiej redukcji, jeżeli wyjściowy układ czynności zostanie zredukowany do jednego łuku zastępczego $(1,3)$, wierzchołek 2 po redukcji znika. Wtedy prawdopodobieństwo realizacji łuku wynosi

$$p_r = \frac{p_i \cdot p_j}{1 - p_i \cdot p_k}, \tag{4.8}$$

a czas trwania czynności odpowiadającej łukowi jest określony wzorem

$$t_r = t_i + t_j + \frac{p_i \cdot p_k \cdot (t_i + t_k)}{1 - p_i \cdot p_k}. \tag{4.9}$$

Wynik analizy przeprowadzonej przy użyciu GERT jest znacznie bogatszy niż przy użyciu technik CPM lub PERT, jednak wymaga znacznie więcej informacji oraz analizy ryzyka. Wymaga również większych zdolności obliczeniowych, zwłaszcza w przypadku dużych i złożonych sieci.

5. Harmonogramowanie projektu budowlanego

Harmonogramowanie projektu budowlanego to proces planowania i zarządzania kolejnością oraz terminami prac dotyczących budowy, remontu lub projektu związanego z infrastrukturą. Dobrze opracowany harmonogram projektu budowlanego jest kluczowy dla skutecznej realizacji projektu, ponieważ pomaga zaplanować, monitorować i kontrolować postęp prac oraz zasoby.

Projekty budowlane to skomplikowane i czasochłonne przedsięwzięcia [10]. Całkowity rozwój projektu składa się zwykle z kilku faz wymagających zróżnicowanego zakresu specjalistycznych usług. Od momentu wstępnego planowania do ukończenia projektu typowe zadanie przechodzi przez kolejne i odrębne etapy, które wymagają wkładu różnych obszarów. Podczas samego procesu budowy nawet skromna konstrukcja wymaga wielu umiejętności, materiałów i różnych operacji. Proces montażu musi przebiegać zgodnie z założeniami technologicznymi, a jego opis stanowi skomplikowany wzorzec indywidualnych wymagań czasowych. Do pewnego stopnia każdy projekt budowlany jest unikalny – żadne dwa zadania nie są dokładnie takie same, a każda konstrukcja jest dostosowana do otoczenia tak, aby spełniała swoją funkcję. Proces budowy podlega wpływowi bardzo zmiennych i czasami nieprzewidywalnych czynników. Zespół budowlany zmienia się w zależności od zlecenia. Wszystkie złożoności związane z różnymi placami budowy, takie jak: warunki gruntowe, topografia powierzchni, pogoda, transport, dostawy materiałów, media i usługi, lokalni podwykonawcy, warunki pracy i dostępne technologie, są nieodłączną częścią procesu budowlanego. W konsekwencji projekty budowlane charakteryzują się złożonością i różnorodnością oraz niestandardowym charakterem ich realizacji.

Projekt budowlany przebiega w określonym porządku. Gdy właściciel zidentyfikuje potrzebę budowy nowego obiektu, musi zdefiniować wymagania i określić ograniczenia czasowe i budżetowe. Kontrola czasu realizacji projektu rozpoczyna się już na etapie projektowania. Ma ona na celu zminimalizowanie czasu trwania budowy zgodnie z jakością projektu i całkowitym kosztem. Sprawdzane są terminy dostaw materiałów i sprzętu. Jeśli w grę wchodzą długie okresy dostaw, projekt jest zmieniany lub zamówienia są inicjowane, gdy tylko projekt jest na wystarczająco zaawansowanym etapie, aby umożliwić sporządzenie szczegółowych specyfikacji zakupów. Wybierane są metody budowy, których charakterystyka kosztowa jest korzystna i dla których, w razie potrzeby, dostępna będzie odpowiednia siła robocza i sprzęt budowlany. Wstępny harmonogram projektu jest zazwyczaj przygotowywany wraz z postępem prac.

Zdecydowana większość projektów budowlanych boryka się z przekroczeniami czasowymi. Czas opóźnień może różnić się znacząco w zależności od wielu czynników, w tym skomplikowania projektu, jego wielkości, rodzaju prac budowlanych,

dostępności zasobów, warunków atmosferycznych, dostaw materiałów, przepisów regulacyjnych i zarządzania projektem. Opóźnienia w projektach budowlanych mogą wynikać z wielu przyczyn i mają różne skutki. W celu utrzymania projektów w ramach harmonogramu i ramach czasowych, w większości firm budowlanych stosowane są różne techniki zarządzania projektami. Jednak większość takich technik jest skomplikowana i wymaga specjalistycznej wiedzy, a poziom komplikacji związany z tymi technikami często zniechęca firmy do ich powszechnego stosowania.

Metody CPM (*critical path method*) oraz PERT (*program evaluation and review technique*) to dwie szeroko stosowane techniki grafowe służące do planowania i harmonogramowania działań budowlanych jako skuteczne narzędzia do zarządzania projektami budowlanymi. Większość kierowników projektów wykorzystuje te techniki jako standardowe narzędzia do planowania różnych działań na etapie budowy. Jednak zarówno PERT, jak i CPM mają wadę polegającą na tym, że graf projektu musi być acykliczny, co oznacza, że nie mogą występować w nim cykle ani pętle. Chociaż zapętlenie działań eliminuje rozbieżności na wczesnym etapie projektu, jest uważane za błąd.

Natomiast metoda GERT to metoda, która pozwala na warunkowe i probabilistyczne traktowanie logicznych relacji między działaniami projektu po losowo określonej sekwencji obserwacji. Algorytm GERT umożliwia tworzenie pętli między zadaniami i uwzględnia węzły probabilistyczne, a tym samym bierze pod uwagę stopień niepewności w projektach. Warunkiem poprawnej konstrukcji harmonogramu projektu budowlanego i analizy grafu aktywności jest ustalenie programu działania, czyli określenie tego, co, gdzie i kiedy oraz w jakiej kolejności ma być wykonane oraz ustalenie czasu trwania poszczególnych działań i czasu wykonania całego projektu.

Na podstawie literatury [1,4,15] oraz po konsultacjach z ekspertem budowlanym, w poniższej tabeli sporządzono zestawienie wszystkich czynności składających się na projekt budowlany, określono kolejność działań oraz przybliżony czas ich realizacji.

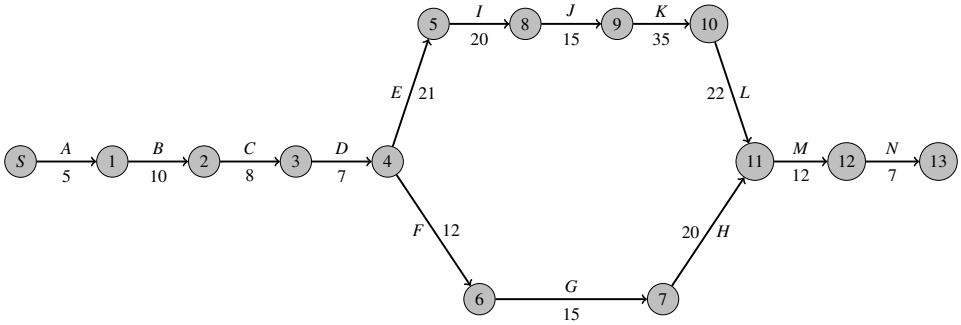
Tabela 1. Harmonogram działań projektu budowlanego

Ozn.	**Nazwa działania**	**Czas trwania (dni)**
A	Prace przygotowawcze (plac budowy)	5
B	Prace ziemne	10
C	Wykonanie elementów zbrojenia	8
D	Wykonanie fundamentów	7
E	Wykonanie instalacji wodno-kanalizacyjnej (przyłącza)	21

Ozn.	Nazwa działania	Czas trwania (dni)
F	Postawienie ścian zewnętrznych i ścianek działowych	12
G	Prace konstrukcyjne: schody wewnętrzne i stropy	15
H	Prace na dachu: pokrycie dachu, izolacja dachu	20
I	Prace sanitarne i instalacje elektryczne	20
J	Stolarka wewnętrzna: drzwi i okna	15
K	Prace tynkarskie i wylewki posadzek	35
L	Prace malarskie	22
M	Prace końcowe: czyszczenie, dekoracja	12
N	Zamknięcie projektu i kontrola końcowa	7

Źródło: Opracowanie własne

Zakładając, że wszystkie czynności będą wykonywane w wyznaczonych czasach trwania jedna po drugiej, czas trwania projektu wyniesie 209 dni. Po sporządzeniu tabeli 1 oraz ustaleniu, jakie czynności poprzedzają daną czynność, jakie następują po danej czynności oraz jakie czynności mogą przebiegać równolegle, możemy narysować graf aktywności przedstawiony na rysunku 10. Krawędziom zostały przypisane oznaczenia literowe czynności w projekcie budowlanym oraz czasy trwania odpowiadające poszczególnym z nich.



Rysunek 10. Graf projektu budowlanego
Źródło: Opracowanie własne

W następnej części zostanie przedstawiona analiza grafu aktywności projektu budowlanego metodami CPM, PERT oraz GERT.

6. Analiza grafu aktywności metodami CPM, PERT i GERT

Analiza grafu aktywności projektu budowlanego składa się z:

1. określenia możliwie najkrótszego czasu wykonania całego projektu, czyli określenia długości drogi krytycznej,
2. wyznaczenia przebiegu drogi krytycznej, idącej od wierzchołka początkowego do wierzchołka końcowego grafu aktywności i czynności krytycznych leżących na tej drodze,
3. określenia najwcześniejszych możliwych i najpóźniejszych dopuszczalnych czasów trwania poszczególnych czynności wchodzących w skład grafu.

6.1. Critical path method

W metodzie CPM całkowity czas realizacji projektu odpowiada trwaniu najdłuższego ciągu czynności, czyli zsumowanym czasom najdłuższej drogi od zdarzenia początkowego do zdarzenia końcowego. Po przygotowaniu grafu aktywności projektu budowlanego przeprowadzamy analizę drogi krytycznej. Wierzchołki grafu są odpowiednio numerowane, a pod tymi numerami znajdują się: najwcześniejszy możliwy i najpóźniejszy dopuszczalny czas, w którym może zajść dana czynność.

Analiza drogi krytycznej w grafie aktywności składa się z dwóch kroków. Najpierw wykonujemy krok w przód, w którym otrzymujemy najwcześniejszy możliwy czas rozpoczęcia zdarzeń. Drugi krok, czyli krok w tył, umożliwia wyznaczenie najpóźniejszego dopuszczalnego czasu zajścia poszczególnych zdarzeń w grafie. Różnica pomiędzy tymi czasami określana jest dla każdego zdarzenia jako całkowity przepływ. Zdarzenia, dla których całkowity przepływ jest równy zeru są to zdarzenia krytyczne, a ścieżka w grafie od zdarzenia początkowego, poprzez zdarzenia krytyczne, do zdarzenia końcowego jest poszukiwaną ścieżką krytyczną. Czas wykonania czynności zgodnie z tą ścieżką określa możliwy najkrótszy termin zakończenia projektu.

W tym projekcie w analizie drogi krytycznej zakładamy, że czas rozpoczęcia realizacji projektu następuje w dniu zerowym i do tego terminu odnoszą się wszystkie inne wyznaczone terminy zdarzeń i czynności. Są one oczywiście umowne.

Przy wyżej opisanym harmonogramie projektu budowlanego analiza wygląda nastepująco:

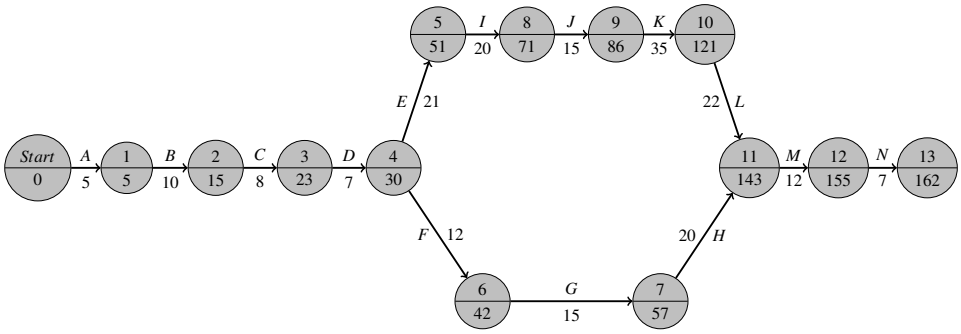
1. Obliczamy czas trwania projektu, rozpoczynając od zdarzenia początkowego (krok w przód) i korzystając ze wzoru (2.1).
   $EF_{(0,1)} = ES_{(0)} + t_{(0,1)} = 0 + 5 = 5$
   $EF_{(1,2)} = ES_{(1)} + t_{(1,2)} = 5 + 10 = 15$
   itd.

Postępując w ten sposób w odniesieniu do ostatniej czynności, wyznaczamy najwcześniejszy możliwy czas wykonania wszystkich zdarzeń. Szczególnym przypadkiem jest czynność M, która musi być zrealizowana po zakończeniu czynności H oraz L. Wtedy jako najwcześniejszy możliwy czas wykonania wybieramy większą wartość obliczoną na podstawie wzoru (2.1), czyli

$$\max\{EF_{(10,11)} = ES_{(10)} + t_{(10,11)}, EF_{(7,11)} = ES_{(7)} + t_{(7,11)}\} =$$

$$\max\{EF_{(10,11)} = 121 + 22 = 143, EF_{(7,11)} = 57 + 20 = 77\} = 143.$$

Wyniki analizy sieci zależności w pierwszym kroku (krok w przód) przedstawia rysunek 11.



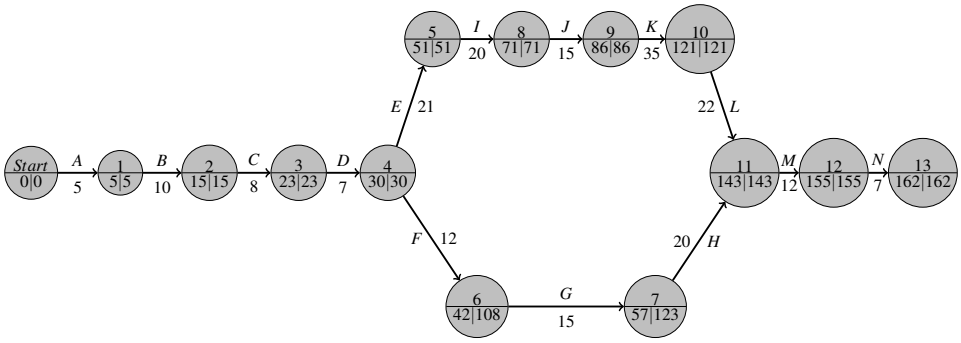
Rysunek 11. Analiza grafu metodą CPM – krok w przód
Źródło: Opracowanie własne

2. Obliczamy czas trwania projektu, rozpoczynając od zdarzenia końcowego (krok w tył) i korzystając ze wzoru (2.2).
   $LS_{(13,12)} = LF_{(13)} - t_{(13,12)} = 162 - 7 = 155$
   $LS_{(12,11)} = LF_{(12)} - t_{(12,11)} = 155 - 12 = 143$
   itd.
   W tym przypadku podobnie, postępując w ten sposób w odniesieniu do ostatniej czynności, wyznaczamy najpóźniejszy dopuszczalny czas wykonania wszystkich zdarzeń. Szczególnym przypadkiem jest czynność D, po której następują czynności E oraz F. W tym przypadku jako najpóźniejszy dopuszczalny czas wykonania wybiera się mniejszą spośród wartości obliczonych na podstawie wzoru (2.2), czyli

$$\min\{LS_{(5,4)} = LF_{(5)} - t_{(5,4)}, LS_{(6,4)} = LF_{(6)} - t_{(6,4)}\} =$$

$$\min\{LS_{(5,4)} = 51 - 21 = 30, LS_{(6,4)} = 108 - 12 = 96\} = 30.$$

Wyniki analizy grafu aktywności krok w tył przedstawia rysunek 12.



Rysunek 12. Analiza grafu metodą CPM – krok w tył
Źródło: Opracowanie własne

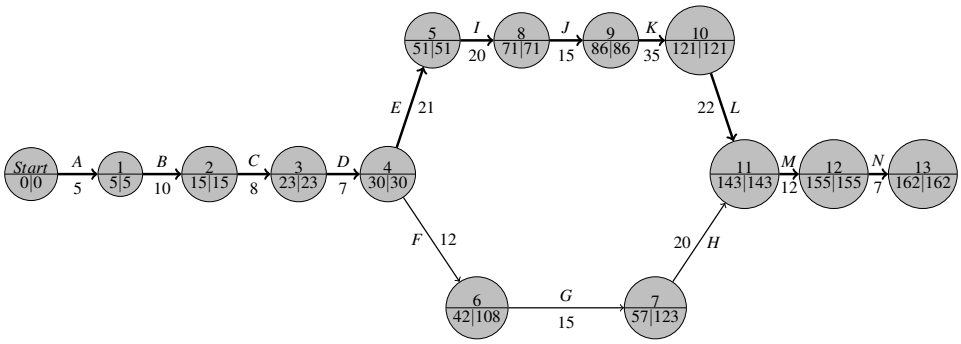
3. Obliczamy całkowite przepływy i wolne przepływy, korzystając ze wzorów (2.3) oraz (2.4).
$TF_{(0,1)} = LF_{(0,1)} - ES_{(0,1)} - t_{(0,1)} = 5 - 0 - 5 = 0$
$FF_{(0,1)} = EF_{(0,1)} - ES_{(0,1)} - t_{(0,1)} = 5 - 0 - 5 = 0$
$TF_{(1,2)} = LF_{(1,2)} - ES_{(1,2)} - t_{(1,2)} = 15 - 5 - 10 = 0$
$FF_{(1,2)} = EF_{(1,2)} - ES_{(1,2)} - t_{(1,2)} = 15 - 5 - 10 = 0$
itd.
W ten sposób wyznaczamy TF oraz FF dla wszystkich czynności w projekcie. Poprzednio obliczone wartości ES, EF, LS oraz LF, jak i wartości TF oraz FF są przedstawione w tabeli 2.

Tabela 2. Wyniki analizy projektu metodą CPM

Ozn.	Dni	ES	EF	LS	LF	FF	TF
A	5	0	5	0	5	0	0
B	10	5	15	5	15	0	0
C	8	15	23	15	23	0	0
D	7	23	30	23	30	0	0
E	21	30	51	30	51	0	0
F	12	30	42	30	108	0	66
G	15	42	57	108	123	0	66
H	20	57	143	123	143	66	66
I	20	51	71	51	71	0	0
J	15	71	86	71	86	0	0

Ozn.	Dni	ES	EF	LS	LF	FF	TF
K	35	86	121	86	121	0	0
L	22	121	143	121	143	0	0
M	12	143	155	143	155	0	0
N	7	155	162	155	162	0	0

Źródło: Opracowanie własne



Rysunek 13. Przebieg ścieżki krytycznej
Źródło: Opracowanie własne

Wniosek 6.1. Jak możemy zauważyć na rysunku 13, przebieg ścieżki krytycznej to: $A-B-C-D-E-I-J-K-L-M-N$. Jest to najdłuższa droga w grafie, która zarazem określa możliwie najkrótszy termin realizacji całego projektu.

Sumując czas trwania czynności na ścieżce krytycznej, uzyskujemy informację, że ukończenie projektu według metody CPM wynosi 162 dni.

Możemy również wnioskować, że wydłużenie którejkolwiek czynności krytycznej, na przykład o jeden dzień, powoduje opóźnienie realizacji całego projektu o jeden dzień. Natomiast każde skrócenie czasu czynności na drodze krytycznej o kilka dni powoduje skrócenie czasu realizacji projektu o tych kilka dni.

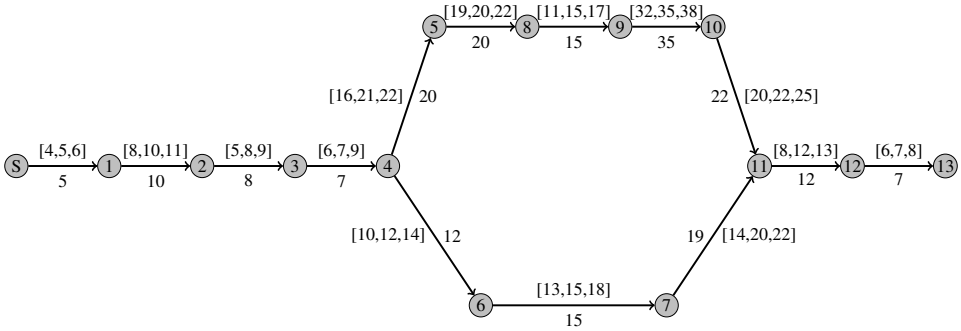
6.2. Program evaluation and review technique

W metodzie PERT czasy trwania czynności mają charakter probabilistyczny. Są one zmiennymi losowymi, a rozkład prawdopodobieństwa ich występowania podlega rozkładowi beta. Metoda ta różni się od metody CPM posiadaniem trzech oszacowań czasów realizacji projektu:

1. czasu optymistycznego – czasu trwania czynności w najbardziej sprzyjających warunkach,

2. czasu pesymistycznego – czasu trwania czynności w najmniej sprzyjających warunkach,
3. czasu najbardziej prawdopodobnego – czasu, który najczęściej występuje przy wielokrotnym powtarzaniu czynności.

Czynności w grafie opisane są na nim czterema parametrami, co przedstawia rysunek 14. Nad krawędziami są podane odpowiednio wartości czasów optymistycznych, najbardziej prawdopodobnych i pesymistycznych, a poniżej krawędzi wartości czasów oczekiwanych T_e obliczonych na podstawie wzoru (3.1).



Rysunek 14. Analiza grafu metodą PERT z wyszczególnieniem charakterystyk czasowych
Źródło: Opracowanie własne

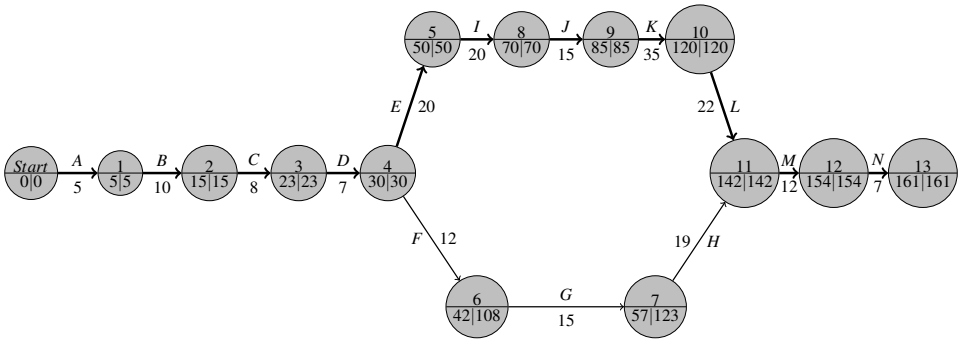
Na podstawie wartości czasu optymistycznego oraz pesymistycznego obliczamy wariancję V_{T_e} daną wzorem (3.2), a następnie wyznaczamy ścieżkę krytyczną, podobnie jak w metodzie CPM, korzystając ze wzorów (2.1), (2.2), (2.3), ale na podstawie wartości czasu T_e. Obliczone wartości przedstawia tabela 3.

Tabela 3. Wyniki analizy projektu metodą PERT

Ozn.	t_o	t_m	t_p	T_e	V	ES	EF	LS	LF	TF
A	4	5	6	5	0.11	0	5	0	5	0
B	8	10	11	10	0.25	5	15	5	15	0
C	5	8	9	8	0.44	15	23	15	23	0
D	6	7	9	7	0.25	23	30	23	30	0
E	16	21	22	20	1	30	50	30	50	0
F	10	12	14	12	0.44	30	42	30	108	66
G	13	15	18	15	0.69	42	57	108	123	66
H	14	20	22	19	1.77	57	142	123	142	66
I	19	20	22	20	0.25	50	70	50	70	0
J	11	15	17	15	1	70	85	70	85	0

Ozn.	t_o	t_m	t_p	T_e	V	ES	EF	LS	LF	TF
K	32	35	38	35	1	85	120	85	120	0
L	20	22	25	22	0.69	120	142	120	142	0
M	8	12	13	12	0.69	142	154	142	154	0
N	6	7	8	7	0.11	154	161	154	161	0

Źródło: Opracowanie własne



Rysunek 15. Analiza grafu metodą PERT
Źródło: Opracowanie własne

Wniosek 6.2. Ścieżka krytyczna to: $A-B-C-D-E-I-J-K-L-M-N$. Sumując czas trwania czynności na ścieżce krytycznej, uzyskujemy informację, że ukończenie projektu według metody PERT wynosi 161 dni.

Po wyznaczeniu ścieżki krytycznej zgodnie z koncepcją PERT obliczamy odchylenie standardowe ze wzoru (3.3). Najpierw należy zsumować wariancje na ścieżce krytycznej.

$$\begin{aligned} V_{T_e} &= V(A)+V(B)+V(C)+V(D)+V(E)+V(I) \\ &\quad +V(J)+V(K)+V(L)+V(M)+V(N) \\ &= 0.11+0.25+0.44+0.25+1+0.25+1+1+0.69+0.69+0.11 \\ &= 5.79 \\ \sigma_{T_e} &= \sqrt{5.79} \end{aligned}$$

Znając oczekiwany termin wykonania projektu wyznaczony dzięki ścieżce krytycznej oraz jego wariancję, możemy obliczyć prawdopodobieństwo zakończenia projektu w zadanym terminie T_s, wiedząc, że T_s wynosi 162 dni (obliczone w metodzie CPM), $\sum T_e$ na ścieżce krytycznej wynosi 161 dni, zaś σ_{T_e} wynosi $\sqrt{5.79}$.

Ze wzoru (3.5) wynika, że

$$Z = \frac{162 - 161}{\sqrt{5.79}} \approx 0.4156 \approx 0.42.$$

Dla obliczonego współczynnika Z z tablic dystrybuanty rozkładu normalnego standaryzowanego odczytujemy prawdopodobieństwo dotrzymania narzuconego z góry terminu. Dla wartości $Z = 0.42$ wynosi ono 0.6628.

Jeżeli wartość prawdopodobieństwa dotrzymania planowanego czasu zakończenia projektu utrzymuje się w granicach od 0.25 do 0.6, to możemy uznać, że wykonanie projektu w terminie jest realne. Jeżeli jednak wartość prawdopodobieństwa jest mniejsza niż 0.25, to istnieje bardzo mała szansa, że projekt zostanie zakończony w planowanej liczbie dni. Przy wartości prawdopodobieństwa większej niż 0.6 przyjmujemy, że w harmonogramie projektu występuje nadmiar zasobów siły roboczej, maszyn lub urządzeń, więc istnieje możliwość, że projekt zostanie wykonany szybciej niż zaplanowano.

Wniosek 6.3. Z przeprowadzonej analizy wynika, że prawdopodobieństwo ukończenia projektu w czasie 162 dni obliczone metodą PERT wynosi 66.28%. Wartość dystrybuanty rozkładu normalnego jest większa od 0.6, więc możemy uznać, że projekt może zostać wykonany szybciej niż w czasie 162 dni np. ze względu na nadmiar zasobów siły roboczej, maszyn lub urządzeń.

6.3. Graphical evaluation and review technique

W grafie opisanym metodą GERT istnieją dwa główne rodzaje zdarzeń: deterministyczne i probabilistyczne. Należy zauważyć, że wszystkie łuki (czynności) wychodzące z węzłów odpowiadających zdarzeniom, odpowiadają prawdopodobieństwom realizacji p. W związku z tym konieczne jest zastosowanie różnych rodzajów węzłów, które mogą mieć różne typy wejść i wyjść.

Wiadomo, że w trakcie wykonywania projektów budowlanych mogą wystąpić różne nieprzewidziane sytuacje, które mogą wymagać dodatkowego czasu na poprawę niektórych aspektów. Przechodząc do opisu poszczególnych zdarzeń w grafie, zakładamy, że inwestor zatrudnił jedną ekipę budowlaną. Po pracach przygotowawczych oraz wykonaniu prac ziemnych, które realizują się z prawdopodobieństem 1, czyli są pewne, pracownicy budowlani przygotowują się do wykonania elementów zbrojenia, a następnie do wykonania fundamentów. Jednak zanim rozpoczną się prace przy fundamentach, konieczna może okazać się poprawa lub wykonanie dodatkowego zbrojenia. Jest bardzo małe prawdopodobieństwo wystąpienia takiej sytuacji (tylko 0.001) w stosunku do pozostałych poprawek, które mogą mieć miejsce w projekcie, a czas trwania takich poprawek to maksymalnie 2 dni. W zależności od tego, czy

taka sytuacja będzie miała miejsce, czy też nie, kolejna czynność zostanie wykonana z prawdopodobieństwem 0.999.

Po wykonaniu fundamentów, zgodnie z tabelą 1, czynności takie jak wykonywanie przyłączy oraz stawianie ścian zewnętrznych i ścianek działowych, mogą być realizowane jednocześnie. Prawdopodobieństwo ich realizacji wynosi 1. Również prace konstrukcyjne są wykonywane z takim samym prawdopodobieństwem. Jednak po ich przeprowadzeniu, w momencie gdy należy rozpocząć prace na dachu związane z pokryciem i izolacją, może okazać się, że termoizolacje lub zamówione materiały na pokrycie dachowe są nieodpowiednie i trzeba będzie je wymienić. W niektórych przypadkach może nastąpić zmiana koncepcji pokrycia dachowego, które według eksperta będzie lepszym rozwiązaniem. W takiej sytuacji nie można zacząć czynności H, dopóki ekipa budowlana nie otrzyma odpowiednich materiałów do prac na dachu. Zamówienie oraz dostawa takich materiałów może potrwać nawet do 10 dni, ponieważ producent nie dostarczy tego typu materiałów z dnia na dzień. Jednak taki przypadek ma bardzo małe prawdopodobieństwo wystąpienia, tylko 0.001, ponieważ dach jest bardzo ważną częścią budowy. Jeśli taka sytuacja będzie miała miejsce, to może ona znacznie wydłużyć prace budowlane.

Przechodząc do prac sanitarnych i instalacji elektrycznych, ekipa budowlana może napotkać kolejne przeszkody. Po wykonaniu prac może okazać się, że punkty świetlne są nieprawidłowo rozmieszczone lub występują przeciążenia na poszczególnych obwodach. Błędy podczas samej instalacji, takie jak nieprawidłowe połączenia czy zwarcia, mogą prowadzić do awarii systemu elektrycznego, a nawet pożaru. W takim przypadku poprawa tych elementów zajmuje średnio 4 dni, a prawdopodobieństwo wystąpienia takich awarii wynosi 0.2. W zależności od tego, czy taka sytuacja będzie miała miejsce, czy też nie, kolejna czynność, czyli wykonanie stolarki wewnętrznej, zostanie wykonana z prawdopodobieństwem 0.8.

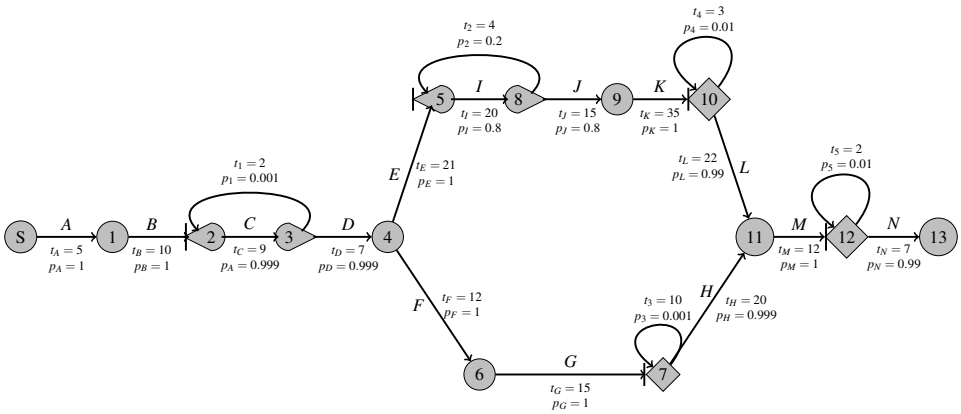
Następnie ekipa budowlana może przejść do prac tynkarskich oraz wylewania posadzek. Są to czynności wykonywane z prawdopodobieństwem 1.

Zgodnie z tabelą 1 kolejnym etapem projektu budowlanego są prace malarskie. Jednak przed ich rozpoczęciem może okazać się, że inwestor zmienił zdanie co do koloru ścian lub zamówiony kolor farby okazał się nieodpowiedni, więc należy zakupić nowe farby. Ekipa budowlana nie może przejść do wykonywania czynności L, dopóki nie zakupi nowych farb. Taka sytuacja może mieć miejsce z prawdopodobieństwem 0.01. Czas oczekiwania na nowe farby wynosi 3 dni, ponieważ są one zazwyczaj dostępne w wielu marketach budowlanych i wystarczy je zakupić oraz dostarczyć na budowę.

Następnie ekipa przechodzi do prac końcowych, które zwykle przebiegają bez żadnych problemów. Na koniec, aby zamknąć projekt, przeprowadzana jest kontrola. Jednak zanim nastąpi taka kontrola i sprawdzenie dokumentacji projektowej, może

się okazać, że dokumenty nie są kompletne, np. brakuje pewnych certyfikatów. W takich okolicznościach należy uzupełnić dokumentację, co zwykle zajmuje nie więcej niż 2 dni. Sytuacja ta może wydarzyć się z prawdopodobieństwem 0.01.

Graf uwzględniający wszystkie problemy i niespodziewane sytuacje w trakcie realizacji projektu budowlanego został przedstawiony na rysunku 16.



Rysunek 16. Analiza grafu metodą GERT
Źródło: Opracowanie własne

Wejście do węzła 2 jest wejściem typu „albo", ponieważ realizacja tego węzła może nastąpić w wyniku zakończenia jednej z dwóch czynności wzajemnie się wykluczających: albo w wyniku zakończenia prac ziemnych (*B*), albo na skutek poprawy wykonanego już zbrojenia lub wykonania dodatkowych elementów zbrojenia. Wyjście węzła 3 jest probabilistyczne, ponieważ w wyniku jego wystąpienia możliwe jest rozpoczęcie prac związanych z fundamentami lub też pojawia się konieczność dokonania poprawek w wykonanym zbrojeniu i powrót do węzła drugiego.

Podobnie wygląda opis zdarzeń przy węzłach 5 i 8. Wejście do węzła 5 jest również wejściem typu „albo", ponieważ realizacja tego węzła może nastąpić w wyniku zakończenia jednej z dwóch czynności wzajemnie się wykluczających: albo w wyniku zakończenia prac przy instalacji wodno-kanalizacyjnej (*E*), albo na skutek poprawy pewnych elementów w pracach sanitarnych oraz instalacjach elektrycznych. Wyjście węzła 8 jest probabilistyczne, ponieważ w wyniku jego wystąpienia możliwe jest rozpoczęcie prac związanych ze stolarką wewnętrzną lub też pojawi się konieczność dokonania poprawek w wykonanych pracach sanitarnych i instalacjach elektrycznych.

Węzeł 7 ma również wejście typu „albo", ponieważ realizacja tego węzła może nastąpić w wyniku zakończenia jednej z dwóch czynności wzajemnie się wykluczających: albo w wyniku zakończenia prac konstrukcyjnych (*G*), albo na skutek

zamówienia nowych materiałów, które są potrzebne do prac na dachu. Natomiast wyjście z węzła 7 ma charakter probabilistyczny, ponieważ w wyniku jego wystąpienia możliwe jest rozpoczęcie prac na dachu lub też pojawi się konieczność zamówienia nowych materiałów. Podobnie interpretujemy zdarzenia przy węzłach 10 i 12.

Węzeł 11 ma wejście typu „i", ponieważ pomimo tego, że wchodzą do niego dwie czynności wzajemnie się wykluczające: H oraz L, to muszą się one obie zakończyć, tzn. muszą zakończyć się prace na dachu i muszą zakończyć się prace malarskie, aby ekipa budowlana mogła przejść do prac końcowych związanych z czyszczeniem i dekoracją wnętrza.

W przypadku metody GERT, w celu przeprowadzenia analizy grafu aktywności, dokonujemy wielu kolejnych uproszczeń jego struktury. Prowadzi to do prostych połączeń czynności. Kolejne redukcje polegają na stopniowym zastępowaniu sprzężeń zwrotnych i pętli łukami zastępczymi. Po przeprowadzeniu odpowiedniej liczby etapów redukcji otrzymujemy graf zredukowany do prostszej postaci, który możemy analizować znanymi już metodami CPM lub PERT.

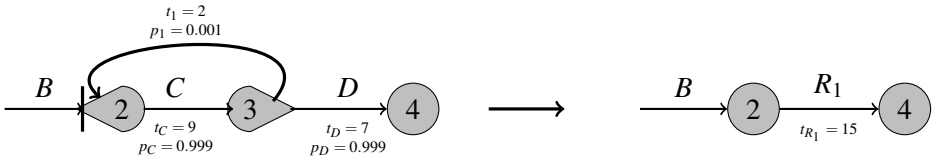
Upraszczanie grafu aktywności przedstawionego na rysunku 16 odbywa się w następujących krokach:

1. Redukcja sprzężenia zwrotnego pomiędzy węzłami 2 i 3.
   Przy wykorzystaniu wzorów (4.8) i (4.9), łuk zastępujący sprzężenie zwrotne 2–3 charakteryzuje się następującymi parametrami

   $$p_{R_1} = \frac{p_C \cdot p_D}{1 - p_C \cdot p_1} = \frac{0.999 \cdot 0.999}{1 - 0.999 \cdot 0.001} = 0.998999,$$

   $$t_{R_1} = t_C + t_D + \frac{p_C \cdot p_1 \cdot (t_C + t_1)}{1 - (p_C \cdot p_1)} = 8 + 7 + \frac{0.999 \cdot 0.001 \cdot (8+2)}{1 - 0.999 \cdot 0.001} = 15 + 0.0099999999 \approx 15.$$

   Po dokonanej redukcji należy sprawdzić poprawność wejść i wyjść węzłów, aby zachować ich prawidłową strukturę. Po uproszczeniu sprzężenia zwrotnego, a także wynikającej z tego likwidacji węzła 3, zmieni się struktura logiczna wejścia węzła 2. Ma on teraz charakter deterministyczny, co możemy zauważyć na rysunku 17.



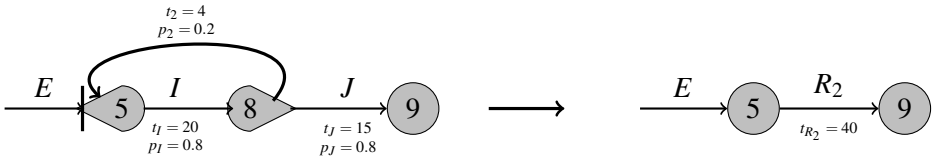
Rysunek 17. Redukcja pierwszego fragmentu grafu aktywności
Źródło: Opracowanie własne

2. Redukcja sprzężenia zwrotnego pomiędzy węzłami 5 i 8.
   Przy wykorzystaniu wzorów (4.8) i (4.9), łuk zastępujący sprzężenie zwrotne 5–8 charakteryzuje się następującymi parametrami

   $$p_{R_2} = \frac{p_I \cdot p_J}{1 - p_I \cdot p_2} = \frac{0.8 \cdot 0.8}{1 - 0.8 \cdot 0.2} = 0.76,$$

   $$t_{R_2} = t_I + t_J + \frac{p_I \cdot p_2 \cdot (t_I + t_2)}{1 - (p_I \cdot p_2)} = 20 + 15 + \frac{0.8 \cdot 0.2 \cdot (20+4)}{1 - 0.8 \cdot 0.2} = 35 + 4.57 = 39.57 \approx 40.$$

   Podobnie jak po pierwszej redukcji, należy sprawdzić poprawność wejść i wyjść węzłów, aby zachować ich prawidłową strukturę. W tym przypadku po uproszczeniu sprzężenia zwrotnego i likwidacji węzła 8, zmieni się struktura logiczna wejścia węzła 5. Ma on teraz charakter deterministyczny, co możemy zauważyć na rysunku 18.



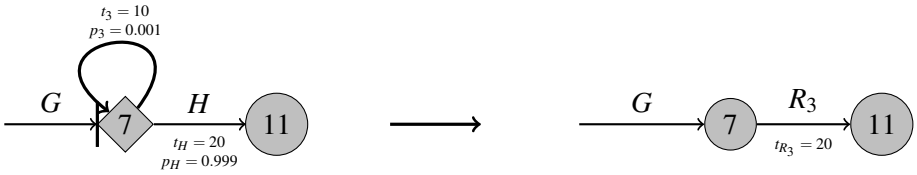
Rysunek 18. Redukcja drugiego fragmentu grafu aktywności
Źródło: Opracowanie własne

3. Redukcja pętli przy węźle 7.
   Przy wykorzystaniu wzorów (4.6) i (4.7), łuk przy węźle 7 charakteryzuje się następującymi parametrami

   $$p_{R_3} = \frac{p_H}{1 - p_3} = \frac{0.999}{1 - 0.001} = 1,$$

   $$t_{R_3} = t_H + \frac{p_3 \cdot t_3}{1 - p_3} = 20 + \frac{0.001 \cdot 10}{1 - 0.001} = 20 + 0.01001 = 20.01001 \approx 20.$$

   Tak jak w przypadku redukcji sprzężeń zwrotnych, tutaj również należy sprawdzić poprawność wejść i wyjść węzłów, aby zachować ich prawidłową strukturę. W wyniku redukcji pętli zmieni się struktura logiczna wejścia i wyjścia węzła 7. Mają one obecnie charakter deterministyczny, co możemy zauważyć na rysunku 19.



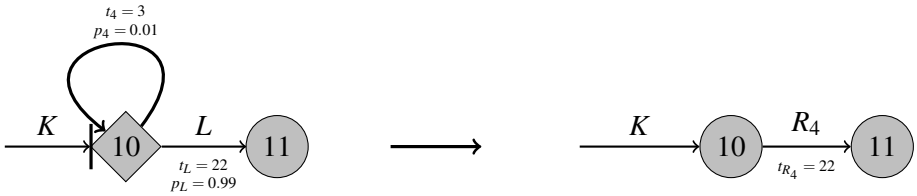
Rysunek 19. Redukcja trzeciego fragmentu grafu aktywności
Źródło: Opracowanie własne

4. Redukcja pętli przy węźle 10.
   Przy wykorzystaniu wzorów (4.6) i (4.7), łuk przy węźle 10 charakteryzuje się następującymi parametrami

   $p_{R_4} = \frac{p_L}{1-p_4} = \frac{0.99}{1-0.01} = 1,$

   $t_{R_4} = t_L + \frac{p_4 \cdot t_4}{1-p_4} = 22 + \frac{0.01 \cdot 3}{1-0.01} = 22 + 0.0303 \approx 22.0303 \approx 22.$

   Należy oczywiście sprawdzić, czy zredukowana część grafu zachowuje poprawną strukturę wejść i wyjść węzłów. W wyniku redukcji pętli zmieni się struktura logiczna wejścia i wyjścia węzła 10. Mają one obecnie charakter deterministyczny, co możemy zauważyć na rysunku 20.



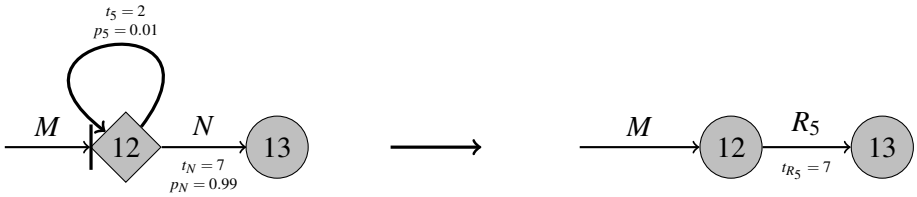
Rysunek 20. Redukcja czwartego fragmentu grafu aktywności
Źródło: Opracowanie własne

5. Redukcja pętli przy węźle 12.
   Przy wykorzystaniu wzorów (4.6) i (4.7), łuk przy węźle 12 charakteryzuje się następującymi parametrami

   $p_{R_5} = \frac{p_N}{1-p_5} = \frac{0.99}{1-0.01} = 1,$

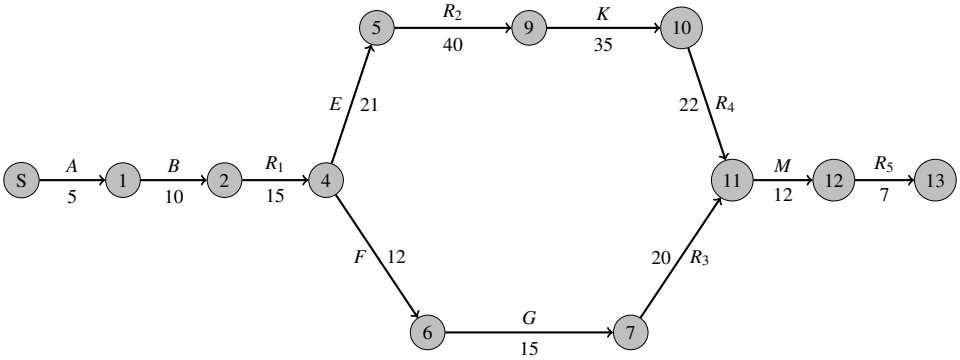
   $t_{R_5} = t_N + \frac{p_5 \cdot t_5}{1-p_5} = 7 + \frac{0.01 \cdot 2}{1-0.01} = 7 + 0.0202 \approx 7.0202 \approx 7.$

   Tak jak w przypadku dwóch poprzednich redukcji pętli, tu również należy sprawdzić poprawność wejść i wyjść węzłów, aby zachować ich prawidłową strukturę. W wyniku redukcji pętli zmieni się struktura logiczna wejścia i wyjścia węzła 12. Mają one obecnie charakter deterministyczny, co możemy zauważyć na rysunku 21.



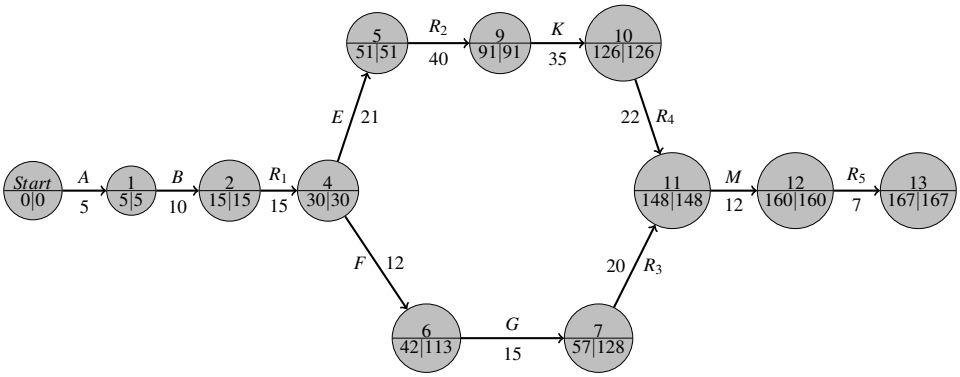
Rysunek 21. Redukcja piątego fragmentu grafu aktywności
Źródło: Opracowanie własne

Po dokonaniu wszystkich redukcji otrzymujemy graf o charakterze deterministycznym. Aby obliczyć liczbę dni trwania projektu, wystarczy zastosować metodę CPM na grafie zredukowanym, który przedstawiono na rysunku 22.



Rysunek 22. Zredukowany graf projektu budowlanego
Źródło: Opracowanie własne

Zgodnie z metodą CPM analizę drogi krytycznej w zredukowanym grafie aktywności zaczynamy od obliczenia czasu trwania projektu krokiem w przód oraz krokiem w tył. Wyniki tych obliczeń przedstawia graf z rysunku 23. Następnie, aby wyznaczyć ścieżkę krytyczną oraz całkowity czas trwania projektu, obliczamy całkowite przepływy i wolne przepływy. Tworzymy tabelę podobną do tabeli 2, ale dla grafu zredukowanego. Wartości ES, EF, LS, LF, FF oraz TF dla grafu z rysunku 23 przedstawia tabela 4.



Rysunek 23. Analiza grafu zredukowanego metodą CPM
Źródło: Opracowanie własne

Tabela 4. Wyniki analizy projektu metodą CPM dla grafu zredukowanego

Ozn.	Dni	ES	EF	LS	LF	FF	TF
A	5	0	5	0	5	0	0
B	10	5	15	5	15	0	0
R_1	15	15	30	15	30	0	0
E	21	30	51	30	51	0	0
F	12	30	42	30	113	0	71
G	15	42	57	113	128	0	71
R_3	20	57	148	128	148	71	71
R_2	40	51	91	51	91	0	0
K	35	91	126	91	126	0	0
R_4	22	126	148	126	148	0	0
M	12	1438	160	148	160	0	0
R_5	7	160	167	160	167	0	0

Źródło: Opracowanie własne

Wniosek 6.4. Ścieżka krytyczna $A-B-R_1-E-R_2-K-R_4-M-R_5$ określa możliwie najkrótszy termin realizacji całego projektu. Sumując czasy trwania czynności na ścieżce krytycznej, wnioskujemy, że ukończenie projektu według metody GERT wynosi 167 dni.

7. Porównanie wyników

Na podstawie osiągniętych wyników możemy stwierdzić, że zastosowanie metod CPM, PERT oraz GERT w harmonogramowaniu projektu budowlanego ma bardzo znaczący wpływ na określenie czasu realizacji projektu. Z tabeli 1 wynika, że jeżeli wszystkie czynności byłyby wykonywane w odpowiednim czasie trwania, jedna po drugiej, to czas trwania całego projektu wyniósłby 209 dni. Z obliczeń metodą CPM uzyskano, że czas wymagany do ukończenia projektu jako całości wynosi 162 dni. Jest to oszczędność czasu o 47 dni. Całkowity czas na ukończenie projektu wynika z sekwencji 11 działań na ścieżce krytycznej. Działania te wymagają większej uwagi ze strony ekipy budowlanej, ponieważ jeśli jedno lub więcej działań na ścieżce krytycznej zostanie opóźnione, spowoduje to opóźnienie zakończenia projektu w stosunku do wcześniej ustalonego harmonogramu. Dalsza analiza harmonogramu projektu budowlanego przy użyciu metody PERT daje prawdopodobieństwo ukończenia projektu w ciągu 162 dni na poziomie 66.28%. Oznacza to, że prawdopodobieństwo ukończenia projektu budowy domu w terminie jest wysokie. Zastosowanie metody GERT pozwala lepiej zrozumieć niepewność projektu

i poziom jego ryzyka, a także pozwala ekipie budowlanej uchwycić możliwość przebiegu projektu podczas różnych nieprzewidzianych sytuacji. W porównaniu z innymi technikami, GERT jest rzadko stosowana w złożonych projektach. Jednak podejście GERT rozwiązuje większość ograniczeń związanych z technikami CPM i PERT. Z uzyskanych wyników można wnioskować, że za pomoc metody GERT można najskuteczniej zaplanować oczekiwaną liczbę dni ukończenia projektu budowlanego tak, aby przewidzieć najbardziej prawdopodobną wartość czasu jego trwania, uwzględniając przy tym stopień niepewności w projekcie.

Literatura

[1] Ba'its H.A., Puspita I.A., Bay A.F., *Combination of Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) for Project Schedule Development*, School of Industrial and System Engineering, Telkom University, Bandung, 40257, Indonesia, 2020.

[2] Cook D.L., *Program Evaluation amnd Review Technique: Applications in Education*, School of Education, The Ohio State University, Cooperative Research, Monograph no.17, Ohio, 1966.

[3] Deo N., *Graph Theory with Applications to Engineering and Computer Science*, Dover Publications, Inc. Mineola, New York, 2016.

[4] Hendradewa A.P., *Schedule Risk Analysis by Different Phases of Construction Project Using CPM-PERT and Monte-Carlo Simulation*, Industrial Engineering Department, Faculty of Industrial Technology, Yogyakarta, Indonesia, 2018.

[5] Levy F.K., Thompson G.L., Wiest J.D., *The ABCs of the Critical Path Method*, Harvard Business Review, 1963.

[6] Liu M., *Program Evaluation and Review Technique (PERT) in Construction Risk Analysis*, Department of Civil Engineering, Shandong University, 17923 Jingshi, China, 2013.

[7] Pritsker A.A.B., *GERT: Graphical Evaluation and Review Technique*, National Aeronautics and Space Administration, Memorandum RM-4973 NASA, The RAND Corporation Santa Monica, California, 1966.

[8] Ramani T., Kannan R., *Scheduling of Industrialized Construction Project Using Graphical Evaluation and Review Technique (GERT)*, Proceedings of Second International Conference on Advances in Industrial Engineering Applications (ICAIEA2014), Anna University, 2014.

[9] Santiago J., Magallon D., *Critical Path Method*, CEE 320, VDC Seminar, 2009.

[10] Sears S.K., Sears G.A., Clough R.H., *Construction Project Management*, A Practical Guide to Field Construction Management, 5th ed., Published by John Wiley and Sons, Inc., Hoboken, New Jersey, 2008.

[11] Trocki M., Wyrozębski P., *Planowanie przebiegu projektów*, Oficyna Wydawnicza SGH, Warszawa, 2015.

[12] Wyrozębski P., Wyrozębska A., *Challenges of Project Planning in the Probabilistic Approach Using PERT, GERT and Monte Carlo*, „Journal of Management and Marketing", 2013, vol.1(1), p. 1–8.

[13] Yu L., Zuo M., *An Estimating Method for IT Project Expected Duration Oriented to GERT*, School of Information, Renmin University of China, Beijing 100872, P.R. China, 2007.

[14] Zawiślak S., Rysiński J., *Graph-Based Modelling in Engineering*, Faculty of Mechanical Engineering and Computer Science, University of Bielsko-Biala, Springer International Publishing, Switzerland, p. 165–173, 2017.

[15] *Etapy budowy domu krok po kroku*, https://www.velux.pl/inspiracje/jak-wybrac-projekt-domu/etapy-budowy-domu-krok-po-kroku, [dostęp: 09.10.2023r.].

Dagmara Dudek [1], **Anna Kuczmaszewska** [2]

Wprowadzenie do przestrzeni podliniowej wartości oczekiwanej

Streszczenie

W rozdziale przedstawiono najważniejsze definicje związane z przestrzenią podliniowej wartości oczekiwanej oraz kilka przykładów zastosowania tego nieklasycznego modelu probabilistycznego. Wprowadzono najistotniejsze, często wykorzystywane w dowodach, lematy oraz twierdzenia w wersji dostosowanej do przestrzeni z niepewnością. Na końcu wprowadzono model regresji liniowej w przestrzeni podliniowej wartości oczekiwanej i rozpatrzono dwa przypadki rozkładu błędu modelu.

Słowa kluczowe: *podliniowa wartość oczekiwana, pojemność, regresja liniowa*

Abstract

The chapter presents the most important definitions related to the sublinear expecatation space and several examples of the application of this non-classical probabilistic model. The most important lemmas and theorems, often used in proofs, were introduced in a version adapted to the space with uncertainty. Finally, a linear regression model was introduced in the sublinear expecatation space and two cases of the model error distribution were considered.

Keywords: *sublinear expected value, capacity, linear regression*

Wstęp

W badaniach statystycznych od dawna wykorzystywane są klasyczne modele probabilistyczne. Okazuje się, że dobrze znane modele prawdopodobieństwa mogą być niewystarczające w warunkach niepewności prawdopodobieństwa. Aby dobrze modelować zjawiska z niepewnością, konieczne jest znalezienie nowych narzędzi. Pomocne okazują się podaddytywne miary prawdopodobieństwa i podliniowe wartości oczekiwane. Umożliwiają one konstruowanie nowych modeli probabilistycznych, które pomagają np. w finansach podczas podejmowania decyzji dotyczących ryzyka. Wprowadzenie nowych narzędzi wymaga ponownego zdefiniowania podstawowych, dobrze znanych pojeć z klasycznej teorii prawdopodobieństwa, jak np. niezależność zmiennych losowych czy zmienne losowe o takim samym rozkładzie. Jest to stosunkowo nowe zagadnienie, jednak w ostatnich latach wzbudzające coraz większe zainteresowanie. W literaturze można znaleźć wiele pozycji, w których

[1] Katedra Matematyki Stosowanej, Politechnika Lubelska
[2] Katedra Matematyki Stosowanej, Politechnika Lubelska

zaprezentowano i udowodniono podstawowe lematy i twierdzenia w nowej formie dostosowanej do przestrzeni podliniowej wartości oczekiwanej.

W tym rozdziale zostaną przedstawione podstawowe definicje związane z podliniową wartością oczekiwaną oraz pojęciem pojemności, które może być identyfikowane jako odpowiednik prawdopodobieństwa z klasycznego modelu probabilistycznego. Pokazane zostaną proste przykłady wykorzystania przestrzeni z niepewnością prawdopodobieństwa. Następnie podane zostaną najważniejsze lematy i twierdzenia w wersji dostosowanej do rozważanej przestrzeni oraz przedstawiony zostanie model regresji liniowej w przestrzeni podliniowej wartości oczekiwanej.

1. Podstawowe definicje

W tej części zostaną przedstawione podstawowe definicje charakteryzujące przestrzeń podliniowej wartości oczekiwanej.

Na początku zdefiniujemy na nowo przestrzeń zmiennych losowych. W tym celu ustalamy następujące oznaczenia.

– $\mathscr{C}_{l.Lip}(\mathbb{R}^n)$ oznaczać będzie w dalszych rozważaniach przestrzeń liniową funkcji φ spełniających lokalny warunek Lipschitza

$$|\varphi(x)-\varphi(y)| \leqslant C(1+|x|^m+|y|^m)|x-y| \quad \text{dla} \quad x,y\in\mathbb{R}^n,$$

gdzie $C>0$ i $m\in\mathbb{N}$ zależą od φ.

– $(\Omega,\mathscr{F})$ będzie ustaloną przestrzenią mierzalną.

W artykule tym przestrzeń $\mathscr{H}$ zdefiniujemy jako liniową przestrzeń funkcji rzeczywistych określonych na $(\Omega,\mathscr{F})$ i spełniających warunek, że jeśli $X_1,...,X_n\in\mathscr{H}$, to $\varphi(X_1,...,X_n)\in\mathscr{H}$, gdzie $\varphi\in\mathscr{C}_{l.Lip}(\mathbb{R}^n)$.

Zauważmy, że

– $\forall_{c\in\mathbb{R}} \quad c\in\mathscr{H}$,
– $X\in\mathscr{H} \implies |X|\in\mathscr{H}$.

Definicja 1.1. [3] Podliniową wartością oczekiwaną $\mathbb{E}$ nazywamy funkcjonał $\mathbb{E}\colon \mathscr{H}\mapsto\mathbb{R}$ spełniający warunki:

(i) monotoniczność

$$\mathbb{E}[X]\leqslant\mathbb{E}[Y] \quad \text{jeśli} \quad X\leqslant Y,$$

(ii) zachowanie stałej

$$\mathbb{E}[c]=c \quad \text{dla} \quad c\in\mathbb{R},$$

(iii) podaddytywność

$$\forall_{X,Y\in\mathscr{H}} \quad \mathbb{E}[X+Y]\leqslant\mathbb{E}[X]+\mathbb{E}[Y],$$

(iv) nieujemna jednorodność

$$\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X] \quad \text{dla} \quad \lambda \geqslant 0.$$

Trójkę $(\Omega, \mathscr{H}, \mathbb{E})$ nazywamy przestrzenią podliniowej wartości oczekiwanej.

Uwaga 1.1. Jeśli spełnione są warunki (i) oraz (ii), to $\mathbb{E}$ nazywamy nieliniową wartością oczekiwaną, a trójkę $(\Omega, \mathscr{H}, \mathbb{E})$ przestrzenią nieliniowej wartości oczekiwanej.

Uwaga 1.2. Jeśli nierówność we własności (iii) stanie się równością, to $\mathbb{E}$ jest liniową wartością oczekiwaną, tzn. $\mathbb{E}$ jest liniowym funkcjonałem spełniającym własności (i) oraz (ii).

Uwaga 1.3. Własność (iv) równoważnie możemy zapisać jako

$$\mathbb{E}[\lambda X] = \lambda^{+}\mathbb{E}[X] + \lambda^{-}\mathbb{E}[-X] \quad \text{dla} \quad \lambda \in \mathbb{R}. \tag{1.1}$$

Dowód. Powyższa równość wynika z następujących faktów

$$\lambda^{+} = \begin{cases} \lambda, & \text{gdy } \lambda \geqslant 0 \\ 0, & \text{gdy } \lambda < 0 \end{cases}$$

oraz

$$\lambda^{-} = \begin{cases} -\lambda, & \text{gdy } \lambda \leqslant 0 \\ 0, & \text{gdy } \lambda > 0 \end{cases}.$$

Przekształcając stronę prawą mamy

$$\lambda^{+}\mathbb{E}[X] + \lambda^{-}\mathbb{E}[-X]$$

$$= \begin{cases} \lambda \mathbb{E}[X] + 0 \cdot \mathbb{E}[-X] = \mathbb{E}[\lambda X], & \text{gdy } \lambda \geqslant 0 \\ 0 \cdot \mathbb{E}[X] + (-\lambda)\mathbb{E}[-X] = \mathbb{E}[(-\lambda)\cdot(-X)] = \mathbb{E}[\lambda X], & \text{gdy } \lambda < 0 \end{cases}.$$

□

Dla podliniowej wartości oczekiwanej $\mathbb{E}$ definiujemy sprzężoną wartość oczekiwaną $\mathscr{E}$ jako

$$\mathscr{E}[X] := -\mathbb{E}[-X], \quad \text{dla każdego} \quad X \in \mathscr{H}.$$

Łatwo zauważyć, że dla każdego $X \in \mathscr{H}$ zachodzi, że $\mathscr{E}[X] \leqslant \mathbb{E}[X]$. W literaturze $\mathbb{E}[X]$ oraz $\mathscr{E}[X]$ są również nazywane, odpowiednio, górną wartością oczekiwaną i dolną wartością oczekiwaną.

Własności (iii) oraz (iv) nazywane są podliniowością. Wynika z niej kolejna własność podliniowej wartości oczekiwanej.

(v) Wypukłość

$$\mathbb{E}[\alpha X + (1-\alpha)Y] \leqslant \alpha\mathbb{E}[X] + (1-\alpha)\mathbb{E}[Y] \text{ dla } \alpha \in [0,1].$$

Jeśli podliniowa wartość oczekiwana $\mathbb{E}$ spełnia warunek wypukłości, to nazywamy ją wypukłą podliniową wartością oczekiwaną.

Warunki (ii) oraz (iii) implikują kolejną własność.

(vi) Własność *cash translatability*

$$\mathbb{E}[X+c] = \mathbb{E}[X] + c \quad \text{dla} \quad c \in \mathbb{R}.$$

Dowód. Równość ta wynika bezpośrednio z następującego oszacowania

$$\mathbb{E}[X] + c = \mathbb{E}[X] - \mathbb{E}[-c] \leqslant \mathbb{E}[X+c] \leqslant \mathbb{E}[X] + \mathbb{E}[c] = \mathbb{E}[X] + c.$$

□

Teraz wprowadzimy kilka kolejnych definicji istotnych z punktu widzenia przestrzeni $(\Omega, \mathscr{H}, \mathbb{E})$. Należą do nich: dominowanie jednej wartości oczekiwanej nad drugą, rozkład zmiennej losowej, pojęcie jednakowych rozkładów, pojęcie niezależności zmiennych losowych i różne rodzaje ich zależności.

Definicja 1.2. [3] Niech $\mathbb{E}_1$ oraz $\mathbb{E}_2$ będą dwoma podliniowymi wartościami oczekiwanymi zdefiniowanymi na $(\Omega, \mathscr{H})$. Mówimy, że $\mathbb{E}_2$ dominuje nad $\mathbb{E}_1$, jeśli

$$\mathbb{E}_1[X] - \mathbb{E}_1[Y] \leqslant \mathbb{E}_2[X-Y] \quad \text{dla} \quad X, Y \in \mathscr{H}. \tag{1.2}$$

Definicja 1.3. [3] Niech $X = (X_1, ..., X_n)$ będzie danym n-wymiarowym wektorem losowym określonym na przestrzeni podliniowych wartości oczekiwanych $(\Omega, \mathscr{H}, \mathbb{E})$. Rozkładem zmiennej X względem $\mathbb{E}$ nazywamy funkcjonał określony na przestrzeni $\mathscr{C}_{l.Lip}(\mathbb{R}^n)$ w następujący sposób

$$\mathbb{F}_X[\varphi] := \mathbb{E}[\varphi(X)] \colon \varphi \in \mathscr{C}_{l.Lip}(\mathbb{R}^n) \mapsto \mathbb{R}.$$

Uwaga 1.4. Trójka $(\mathbb{R}^n, \mathscr{C}_{l.Lip}(\mathbb{R}^n), \mathbb{F}_X)$ jest przestrzenią podliniowych wartości oczekiwanych. $\mathbb{F}_X$ jest również podliniową wartością oczekiwaną.

Podobnie jak w przypadku klasycznej przestrzeni probabilistycznej, z rozkładem prawdopodobieństwa zmiennej losowej $X \in \mathscr{H}$ związane są jej parametry. Odpowiednikami wartości oczekiwanej i wariancji zmiennej losowej są w nowej przestrzeni dolna i górna wartość oczekiwana oraz dolna i górna wariancja

$$\overline{\mu} := \mathbb{E}[X], \quad \underline{\mu} := -\mathbb{E}[-X],$$

$$\overline{\sigma}^2 := \mathbb{E}[X^2], \quad \underline{\sigma}^2 := -\mathbb{E}[-X^2].$$

Przedział $[\underline{\mu}, \overline{\mu}]$ opisuje niepewność średniej zmiennej X, natomiast przedział $[\underline{\sigma}^2, \overline{\sigma}^2]$ opisuje niepewność wariancji zmiennej losowej X.

Definicja 1.4. [3] Niech X_1 oraz X_2 będą n-wymiarowymi wektorami losowymi zdefiniowanymi na przestrzeni podliniowych wartości oczekiwanych, odpowiednio $(\Omega_1, \mathscr{H}_1, \mathbb{E}_1)$ oraz $(\Omega_2, \mathscr{H}_2, \mathbb{E}_2)$. Mówimy, że wektory X_1 oraz X_2 mają jednakowe rozkłady (co oznaczamy jako $X_1 \overset{d}{=} X_2$), jeśli

$$\mathbb{E}_1[\varphi(X_1)] = \mathbb{E}_2[\varphi(X_2)] \quad \text{dla wszystkich} \quad \varphi \in \mathscr{C}_{b.Lip}(\mathbb{R}^n),$$

gdzie $\mathscr{C}_{b.Lip}(\mathbb{R}^n)$ jest przestrzenią ciągłych i ogarniczonych funkcji spełniających warunek Lipschitza.

Definicja 1.5. (Niezależność [3]) Niech dana będzie przestrzeń podliniowej wartości oczekiwanej $(\Omega, \mathscr{H}, \mathbb{E})$. Wektor losowy $Y = (Y_1, ..., Y_n)$, $Y_i \in \mathscr{H}$ nazywa się niezależnym od innego wektora losowego $X = (X_1, ..., X_m)$, $X_i \in \mathscr{H}$, jeśli dla dowolnej funkcji $\varphi \in \mathscr{C}_{b.Lip}(\mathbb{R}^{m+n})$ mamy

$$\mathbb{E}[\varphi(X,Y)] = \mathbb{E}[\mathbb{E}[\varphi(x,Y)]_{x=X}]. \tag{1.3}$$

Należy zauważyć, że w przypadku podliniowej wartości oczekiwanej, jeśli wektor Y jest niezależny od wektora X, to nie implikuje to, w przypadku ogólnym, że wektor X jest niezależny od wektora Y. Obrazuje to poniższy przykład.

Przykład 1.1. Rozważmy sytuację, gdzie $\mathbb{E}$ jest podliniową wartością oczekiwaną oraz X, $Y \in \mathscr{H}$ są wektorami losowymi o jednakowych rozkładach, przy czym $\mathbb{E}[X] = \mathbb{E}[-X] = 0$ oraz $\overline{\sigma}^2 = \mathbb{E}[Y^2] > \underline{\sigma}^2 = -\mathbb{E}[-Y^2]$. Dodatkowo zakładamy, że $\mathbb{E}[|X|] > 0$, stąd $\mathbb{E}[X^+] = \frac{1}{2}\mathbb{E}[|X| + X] = \frac{1}{2}\mathbb{E}[|X|] > 0$. Ponadto dowolną liczbę x możemy zapisać jako $x = x^+ - x^-$. Rozpatrzmy przypadek, gdy Y jest niezależny od X. Wtedy

$$xY^2 = \begin{cases} x^+Y^2, & \text{gdy } x > 0 \\ 0, & \text{gdy } x = 0 \\ -x^-Y^2, & \text{gdy } x < 0 \end{cases}$$

oraz

$$\mathbb{E}[xY^2] = \begin{cases} x^+\mathbb{E}[Y^2], & \text{gdy } x > 0 \\ 0, & \text{gdy } x = 0 \\ x^-\mathbb{E}[-Y^2], & \text{gdy } x < 0 \end{cases}$$

$$= x^+\mathbb{E}[Y^2] + x^-\mathbb{E}[-Y^2] = x^+\overline{\sigma}^2 - x^-\underline{\sigma}^2.$$

Mamy zatem

$$\mathbb{E}[XY^2] = \mathbb{E}[\mathbb{E}[xY^2]]_{x=X} = \mathbb{E}[x^+\overline{\sigma}^2 - x^-\underline{\sigma}^2]_{x^+=X^+,x^-=X^-}$$
$$= \mathbb{E}[X^+\overline{\sigma}^2 - X^-\underline{\sigma}^2] \geqslant \mathbb{E}[X^+\overline{\sigma}^2] - \mathbb{E}[X^-\underline{\sigma}^2]$$
$$= \overline{\sigma}^2\mathbb{E}[X^+] - \underline{\sigma}^2\mathbb{E}[X^-] = (\overline{\sigma}^2 - \underline{\sigma}^2)\mathbb{E}[X^+] > 0.$$

Gdy założymy, że X jest niezależny od Y mamy

$$\mathbb{E}[XY^2] = \mathbb{E}[\mathbb{E}[Xy^2]]_{y=Y} = \mathbb{E}[y^2\mathbb{E}[X]]_{y=Y} = \mathbb{E}[y^2 \cdot 0]_{y=Y} = 0,$$

co wyklucza implikację, że z niezależności Y od X wynika niezależność X od Y.

Definicja 1.6. (Ujemna zależność [7]) Niech dana będzie przestrzeń podliniowej wartości oczekiwanej $(\Omega, \mathscr{H}, \mathbb{E})$. Wektor losowy $Y = (Y_1, ..., Y_n)$, $Y_i \in \mathscr{H}$ nazywamy ujemnie zależnym względem innego wektora losowego $X = (X_1, ..., X_m)$, $X_i \in \mathscr{H}$, jeśli dla każdej pary funkcji $\varphi_1 \in \mathscr{C}_{l.Lip}(\mathbb{R}^m)$ and $\varphi_2 \in \mathscr{C}_{l.Lip}(\mathbb{R}^n)$, niemalejących lub nierosnących współrzędnościowo i takich, że

$$\varphi_1(X) \geqslant 0, \quad \mathbb{E}[\varphi_2(Y)] \geqslant 0,$$

$$\mathbb{E}[|\varphi_1(X)\varphi_2(Y)|] < \infty, \quad \mathbb{E}[|\varphi_1(X)|] < \infty, \quad \mathbb{E}[|\varphi_2(Y)|] < \infty$$

spełniony jest warunek

$$\mathbb{E}[\varphi_1(X)\varphi_2(Y)] \leqslant \mathbb{E}[\varphi_1(X)]\mathbb{E}[\varphi_2(Y)].$$

Definicja 1.7. (Zmienne losowe akceptowalne w szerszym sensie (ang. *widely acceptable*) [4]) Niech $\{Y_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych w przestrzeni podliniowej wartości oczekiwanej $(\Omega, \mathscr{H}, \mathbb{E})$. Ciąg $\{Y_n, n \geqslant 1\}$ nazywamy akceptowalnym w szerszym sensie, jeśli istnieje taki ciąg liczb dodatnich $\{g(n), n \geqslant 1\}$, że dla każdego $t \geqslant 0$ i dla każdego $n \in \mathbb{N}$ spełniony jest warunek

$$\mathbb{E}e^{\sum_{i=1}^n tY_i} \leqslant g(n)\prod_{i=1}^n \mathbb{E}e^{tY_i}.$$

Definicja 1.8. [3] Mówimy, że ciąg n-wymiarowych wektorów losowych $\{\eta_i\}_{i=1}^\infty$ zdefiniowany na przestrzeni podliniowych wartości oczekiwanych $(\Omega, \mathscr{H}, \mathbb{E})$ jest zbieżny według rozkładu, jeśli dla każdej funkcji $\varphi \in \mathscr{C}_{b.Lip}(\mathbb{R}^n)$, ciąg $\{\mathbb{E}[\varphi(\eta_i)]\}_{i=1}^\infty$ jest zbieżny.

Kolejnym ważnym pojęciem dla przestrzeni podliniowej wartości oczekiwanej jest pojemność, która może być traktowana jako odpowiednik prawdopodobieństwa w klasycznej przestrzeni $(\Omega, \mathscr{F}, \mathscr{P})$.

Definicja 1.9. [6] Niech dane będzie σ-ciało $\mathscr{G} \subset \mathscr{F}$. Funkcję $V\colon \mathscr{G} \to [0,1]$ nazywamy pojemnością, jeśli spełnione są następujące warunki

$$V(\emptyset) = 0, \quad V(\Omega) = 1$$

oraz

$$V(A) \leqslant V(B), \quad \text{dla dowolnych} \quad A \subset B, \quad A, B \in \mathscr{G}.$$

Funkcję tę nazywamy funkcją podaddytywną, jeśli spełnia warunek $V(A \cup B) \leqslant V(A) + V(B)$ dla każdego $A, B \in \mathscr{G}$.

Dla ustalonej przestrzeni $(\Omega, \mathscr{H}, \mathbb{E})$, używając pojęcia górnej wartości oczekiwanej, możemy zdefiniować parę $(\mathbb{V}, \mathscr{V})$ odpowiadających sobie pojemności:

$$\mathbb{V}(A) := \inf\{\mathbb{E}[\xi]\colon I_A \leqslant \xi,\ \xi \in \mathscr{H}\}, \quad \mathscr{V}(A) := 1 - \mathbb{V}(A^c), \quad \forall A \in \mathscr{F},$$

gdzie A^c jest dopełnieniem zbioru A.

Zauważmy, że górna pojemność $\mathbb{V}$ jest podaddytywna w przeciwieństwie do pojemności dolnej $\mathscr{V}$ i dolnej wartości oczekiwanej $\mathscr{E}$.

Lemat 1.1. Niech para $(\mathbb{V}, \mathscr{V})$ będzie odpowiednio górną i dolną pojemnością oraz niech para $(\mathbb{E}, \mathscr{E})$ będzie odpowiednio górną i dolną wartością oczekiwaną. Wtedy dla dowolnych $A, B \in \mathscr{G}$ oraz dowolnych $X, Y \in \mathscr{H}$ mamy

$$\mathscr{V}(A \cup B) \leqslant \mathscr{V}(A) + \mathbb{V}(B) \tag{1.4}$$

oraz

$$\mathscr{E}[X + Y] \leqslant \mathscr{E}[X] + \mathbb{E}[Y]. \tag{1.5}$$

Dowód. Zauważmy, że

$$\mathbb{V}(A^c \cap B^c) = \mathbb{V}((A \cup B)^c) = 1 - \mathscr{V}(A \cup B) \tag{1.6}$$

oraz

$$\mathbb{V}(A^c) \leqslant \mathbb{V}[(A^c - B) \cup B] \leqslant \mathbb{V}(A^c - B) + \mathbb{V}(B). \tag{1.7}$$

Z (1.7) dostajemy

$$\mathbb{V}(A^c \cap B^c) = \mathbb{V}(A^c - B) \geqslant \mathbb{V}(A^c) - \mathbb{V}(B). \tag{1.8}$$

Na podstawie (1.6), (1.8) oraz definicji dolnej pojemności otrzymujemy

$$1-\mathscr{V}(A\cup B)\geqslant 1-\mathscr{V}(A)-\mathbb{V}(B).$$

Odpowiednio przegrupowując nierówność, dostajemy (1.4).
Teraz zauważmy, że na podstawie (1.2) mamy

$$\mathbb{E}[-X-Y]\geqslant \mathbb{E}[-X]-\mathbb{E}[Y].$$

Na podstawie definicji dolnej podliniowej wartości oczekiwanej zachodzi, że

$$\mathbb{E}[-X]=-\mathscr{E}[X]$$

oraz

$$\mathbb{E}[-X-Y]=\mathbb{E}[-(X+Y)]=-\mathscr{E}[X+Y].$$

Stąd mamy

$$-\mathscr{E}(X+Y)\geqslant -\mathscr{E}(X)-\mathbb{E}(Y),$$

co jest równoważne z $\mathscr{E}[X+Y]\leqslant \mathscr{E}[X]+\mathbb{E}[Y]$. □

Teraz wprowadźmy kolejną definicję związaną z pojemnością.

Definicja 1.10. Parę pojemności $(\mathbb{V},\mathscr{V})$ nazywamy generowaną przez podliniową wartość oczekiwaną $\mathbb{E}$, jeśli

$$\mathbb{V}(A):=\mathbb{E}[I_A],\quad \mathscr{V}(A):=\mathscr{E}[I_A],\quad \text{dla}\quad I_A\in\mathscr{H}.$$

Zauważmy, że zachodzą poniższe warunki

$$\mathbb{E}[f]\leqslant \mathbb{V}(A)\leqslant \mathbb{E}[g],\quad \mathscr{E}[f]\leqslant \mathscr{V}(A)\leqslant \mathscr{E}[g],\quad \text{jeśli}\quad f\leqslant I_A\leqslant g,\quad f,g\in\mathscr{H}.$$

Dodatkowo, wprowadźmy pojęcie całki Choqueta

$$C_V[X]=\int_0^\infty V(X\geqslant t)dt+\int_{-\infty}^0 [V(X\geqslant t)-1]dt,$$

gdzie V może być zastąpione odpowiednio przez $\mathbb{V}$ lub $\mathscr{V}$.

W rozważaniach dotyczących przestrzeni podliniowej wartości oczekiwanej ważne jest zdefiniowanie takich własności jak przeliczalna podaddytywność i ciagłość w odniesieniu do podliniowej wartości oczekiwanej i pojemności.

Definicja 1.11. [6]

(I) Mówimy, że podliniowa wartość oczekiwana $\mathbb{E}\colon \mathscr{H}\to\mathbb{R}$ jest przeliczalnie podaddytywna, jeśli spełniony jest następujący warunek

(a) przeliczalna podaddytywność

$$\mathbb{E}[X] \leqslant \sum_{n=1}^{\infty} \mathbb{E}[X_n],$$

jeśli $X \leqslant \sum_{n=1}^{\infty} X_n$, $X, X_n \in \mathscr{H}$ oraz $X \geqslant 0$, $X_n \geqslant 0$, dla $n = 1, 2, \ldots$

(II) Mówimy, że podliniowa wartość oczekiwana $\mathbb{E}$ jest ciągła, jeśli spełnione są warunki

(a) ciągłość od dołu

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X],$$

jeśli $0 \leqslant X_n \uparrow X$, dla wszystkich $X_n, X \in \mathscr{H}$,

(b) ciągłość od góry

$$\mathbb{E}[X_n] \downarrow \mathbb{E}[X],$$

jeśli $0 \leqslant X_n \downarrow X$, dla wszystkich $X_n, X \in \mathscr{H}$.

(III) Mówimy, że funkcja $V : \mathscr{F} \to [0,1]$ jest przeliczalnie podaddytywna, jeśli

$$V\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} V(A_n), \quad \forall A_n \in \mathscr{F}.$$

(IV) Mówimy, że pojemność $V : \mathscr{F} \to [0,1]$ jest ciągła, jeśli spełnione są warunki

(a) ciągłość od dołu

$$V(A_n) \uparrow V(A),$$

jeśli $A_n \uparrow A$, dla wszystkich $A_n, A \in \mathscr{F}$,

(b) ciągłość od góry

$$V(A_n) \downarrow V(A),$$

jeśli $A_n \downarrow A$, dla wszystkich $A_n, A \in \mathscr{F}$.

Jednym z problemów, na jakie można natrafić w przestrzeni podliniowej wartości oczekiwanej jest brak możliwości zastosowania tradycyjnego indykatora, ze względu na jego nieciągłość. Oznacza to, że nie zawsze $I_A \in \mathscr{H}$, dlatego Zhong i Wu (2017) w pracy [8] wprowadzili propozycję funkcji, która może, w niektórych przypadkach, pełnić rolę podobną do indykatora w rozważanej przestrzeni.

Dla $0 < \mu < 1$, niech $g(x)$ będzie nierosnącą funkcją taką, że $g(x) \in \mathscr{C}_{l.Lip}(\mathbb{R})$, $0 \leqslant g(x) \leqslant 1$ dla każdego x oraz $g(x) = 1$ jeśli $|x| \leqslant \mu$, $g(x) = 0$ jeśli $|x| > 1$. Wtedy spełnione są warunki

$$I(|x| \leqslant \mu) \leqslant g(x) \leqslant I(|x| \leqslant 1), \quad I(|x| \geqslant 1) \leqslant 1 - g(x) \leqslant I(|x| \geqslant \mu).$$

1.1. Reprezentacja podliniowej wartości oczekiwanej

Podliniowa wartość oczekiwana może być zdefiniowana jako supremum liniowych wartości oczekiwanych.

Twierdzenie 1.1. *[3] Niech $\mathbb{E}$ będzie funkcjonałem zdefiniowanym na liniowej przestrzeni $\mathscr{H}$ i spełniającym warunki podaddytywności oraz nieujemnej jednorodności. Wtedy istnieje rodzina liniowych funkcjonałów $E_\theta : \mathscr{H} \mapsto \mathbb{R}$, indeksowana w zbiorze Θ, taka że*

$$\mathbb{E}[X] = \sup_{\theta\in\Theta} E_\theta[X] \quad dla \quad X \in \mathscr{H}. \tag{1.9}$$

Powyższe twierdzenie jest bardzo przydatne w rozważaniach dotyczących modelowania zjawisk zachodzących w warunkach niepewności.

Twierdzenie 1.2. *[3] (Twierdzenie Daniella-Stone'a) Załóżmy, że $(\Omega, \mathscr{H}, \mathbb{E})$ jest przestrzenią podliniowej wartości oczekiwanej, w której spełniony jest warunek*

$$\mathbb{E}[X_i] \to 0, \quad gdy \quad i \to \infty,$$

dla dowolnego ciągu zmiennych losowych $\{X_i\}_{i=1}^{\infty}$ z przestrzeni $\mathscr{H}$, spełniającego warunek, że $X_i(\omega) \downarrow 0$ dla każdego $\omega \in \Omega$. Wtedy istnieje rodzina miar prawdopodobieństwa $\{P_\theta\}_{\theta\in\Theta}$ zdefiniowana na przestrzeni mierzalnej $(\Omega, \sigma(\mathscr{H}))$ taka, że

$$\mathbb{E}[X] = \sup_{\theta\in\Theta} \int_\Omega X(\omega) dP_\theta, \quad dla\ dowolnego \quad X \in \mathscr{H},$$

gdzie $\sigma(\mathscr{H})$ jest najmniejszym σ-ciałem generowanym przez $\mathscr{H}$.

Podobne twierdzenie o reprezentacji można uzyskać dla rozkładu w przestrzeni $(\Omega, \mathscr{H}, \mathbb{E})$.

Twierdzenie 1.3. *Niech $(\Omega, \mathscr{H}, \mathbb{E})$ będzie przestrzenią podliniowej wartości oczekiwanej oraz niech $\mathbb{F}_X[\varphi] = \mathbb{E}[\varphi(X)]$ będzie podliniowym rozkładem wektora losowego $X \in \mathscr{H}^d$. Wtedy istnieje rodzina miar prawdopodobieństwa $\{F_\theta\}_{\theta\in\Theta}$ zdefiniowana na $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ taka, że*

$$\mathbb{F}_X[\varphi] = \sup_{\theta\in\Theta} \int_{\mathbb{R}^d} \varphi(x) F_\theta(dx), \quad dla\ dowolnej \quad \varphi \in \mathscr{C}_{l.Lip}(\mathbb{R}^d).$$

$\mathbb{F}_X[\cdot]$ charakteryzuje niepewność rozkładu zmiennej X.

1.2. Przykłady

W tym paragrafie podamy kilka przykładów, które pokażą użyteczność wybranych nowych pojęć.

Przykład 1.2. Urna zawiera B białych, C czarnych i Z zielonych kul. Gracz wybiera losowo jedną kulę z urny. Gracz nie zna dokładnej liczby kul poszczególnych kolorów, wie jedynie, że $B+C+Z=100$ oraz $B=C\in[20,25]$. Niech ξ będzie zmienną losową zdefiniowaną następująco

$$\xi=\begin{cases}1, & \text{jeśli wybrana została biała kula}\\ 0, & \text{jeśli wybrana została zielona kula}\\ -1, & \text{jeśli wybrana została czarna kula}\end{cases}.$$

Zdefiniujmy wygraną/stratę gracza jako zmienną losową $X=\varphi(\xi)$, gdzie $\varphi(\cdot)$ jest ustaloną funkcją spełniającą warunek Lipschitza. Funkcja $\varphi(\cdot)$ opisuje tę wygraną/stratę w trakcie gry.

W tabeli 1 podano rozkład zmiennej losowej ξ.

Tabela 1. Rozkład zmiennej losowej ξ

-1	0	1
p	$1-2p$	p

Jest to rozkład z niepewnością prawdopodobieństwa: $p\in\left[\frac{\underline{\mu}}{100},\frac{\overline{\mu}}{100}\right]=[0.2,0.25]$. Niepewność prawdopodobieństwa wynika z faktu, że nie jest znana dokładna liczba kul żadnego z kolorów.

Niech $\mathscr{P}$ oznacza zbiór wszystkich rozkładów zmiennej losowej ξ zależnych od wartości $p\in[0.2,0.25]$. Stąd wartość oczekiwaną wygranej/straty gracza, uwzględniającą niepewność rozkładu zmiennej ξ, obliczamy w następujący sposób

$$\mathbb{E}[\varphi(\xi)]:=\sup_{P\in\mathscr{P}}E_P[\varphi(\xi)]=\sup_{p\in[\frac{\underline{\mu}}{100},\frac{\overline{\mu}}{100}]}\Big[p\varphi(1)+(1-2p)\varphi(0)+p\varphi(-1)\Big].$$

Przykład 1.3. (Podliniowa wersja ciągu zmiennych losowych o rozkładzie Bernoulliego) W urnie jest c_1 czarnych kul oraz b_1 białych kul. Gracz wie, że c_1 i b_1 spełniają warunek $c_1+b_1=100$ oraz, że $c_1\in[\underline{\mu},\overline{\mu}]$, gdzie $\underline{\mu}$ oraz $\overline{\mu}$ są parą ustalonych liczb takich, że $0<\underline{\mu}<\overline{\mu}<100$. Gracz wybiera losowo jedną kulę z urny. Jeśli wylosowaną kulą jest kula czarna, to dostaje złotówkę, a jeśli wylosowaną kulą jest kula biała, to traci złotówkę. Zysk z takiej gry jest losową liczbą i może być zapisany jako

$$\xi_1(\omega)=\mathbf{1}(\omega)_{\{\text{czarna kula}\}}-\mathbf{1}(\omega)_{\{\text{biała kula}\}}.$$

Po każdej i-tej próbie gracz notuje wynik ξ_i i przystępuje do następnej $(i+1)$-szej próby, losując jedną kulę spośród 100, z zastrzeżeniem, że w $(i+1)$-szym losowaniu liczba kul czarnych zmieniła się i wynosi c_{i+1}, ale nadal zachodzi warunek, że $c_{i+1} \in [\underline{\mu}, \overline{\mu}]$. W ten sposób tworzony jest ciąg zmiennych losowych $\{\xi_i\}_{i=1}^{\infty}$.

Jeśli w chwili i sprzedajemy kontrakt $\varphi(\xi_i)$ w oparciu o i-ty wynik ξ_i, to, zakładając najgorszy scenariusz, wartość oczekiwana będzie wynosić

$$\mathbb{E}[\varphi(\xi_i)] = \mathbb{E}[\varphi(\xi_1)] = \sup_{p \in [\frac{\underline{\mu}}{100}, \frac{\overline{\mu}}{100}]} [p\varphi(1) + (1-p)\varphi(-1)], \quad i = 1, 2, \ldots$$

Zauważmy, że $\mathbb{E}[\varphi(\xi_i)] = \mathbb{E}[\varphi(\xi_j)]$ dla dowolnych $i, j = 1, 2, ...$ Oznacza to, że mamy do czynienia z ciągiem zmiennych losowych o jednakowym rozkładzie. Łatwo można również sprawdzić, że ξ_{i+1} jest niezależne od $(\xi_1, ..., \xi_i)$.

Jeśli rozpatrzymy funkcję straty $X(\omega) = \varphi(\xi_1, ..., \xi_i)$, wtedy wartość oczekiwana straty będzie wynosić

$$\mathbb{E}[X] = \mathbb{E}[\varphi(\xi_1, ..., \xi_i, \xi_{i+1})] = \mathbb{E}[\mathbb{E}[\varphi(x_1, ..., x_i, \xi_{i+1})]_{\{\xi_j = x_j, 1 \leqslant j \leqslant i\}}],$$

co wypełnia warunek (1.3). Jest to typowy ciąg zmiennych losowych o rozkładzie Bernoulliego przy założeniu niepewności prawdopodobieństwa.

Przykład 1.4. Rozważmy teraz inny ciąg zmiennych losowych o rozkładzie Bernoulliego, w którym w i-tej próbie mamy: c_i czarnych kul, b_i białych kul oraz z_i zielonych kul. Dodatkowo niech zachodzi, że $c_i = b_i$, $c_i + b_i + z_i = 100$ oraz $c_i \in [\underline{\mu}, \overline{\mu}]$, gdzie $0 < \underline{\mu} < \overline{\mu} < 50$ są znane i ustalone. Gra jest powtarzana podobie jak w poprzednim przykładzie, tzn. za wylosowanie kuli czarnej dostajemy złotówkę, za wylosowanie kuli białej tracimy złotówkę, a za wylosowanie kuli zielonej nic nie tracimy i nie zyskujemy. Zysk z takiej $i-$tej gry może być opisany jako

$$\xi_i(\omega) = \mathbf{1}(\omega)_{\{\text{czarna kula}\}} - \mathbf{1}(\omega)_{\{\text{biała kula}\}}.$$

Dostajemy wtedy ciąg zmiennych losowych o jednakowym rozkładzie, gdzie zachodzi:

$$\mathbb{E}[\xi_i] = \sup_{p \in [\frac{\underline{\mu}}{100}, \frac{\overline{\mu}}{100}]} [p \cdot 1 + p \cdot (-1) + (1-2p) \cdot 0] = 0$$

oraz

$$-\mathbb{E}[-\xi_i] = -\sup_{p \in [\frac{\underline{\mu}}{100}, \frac{\overline{\mu}}{100}]} [p \cdot (-1) + p \cdot 1 + (1-2p) \cdot 0] = 0.$$

Ten rodzaj ciągu Bernoulliego przy założeniu o niepewności nazywamy ciągiem symetrycznym.

2. Znane lematy i twierdzenia w przestrzeni podliniowej wartości oczekiwanej

W tej części pracy przedstawimy wybrane, ważne lematy i twierdzenia, które są odpowiednikami znanych narzędzi z klasycznej przestrzeni probabilistycznej $(\Omega, \mathscr{F}, P)$ i są użyteczne w wielu rozważaniach teoretycznych i zastosowaniach praktycznych.

Lemat 2.1. Niech $(\Omega, \mathscr{H}, \mathbb{E})$ będzie przestrzenią podliniowej wartości oczekiwanej, X i Y będą dwoma zmiennymi losowymi. Załóżmy ponadto, że spełniony jest warunek

$$\mathbb{E}[Y] = -\mathbb{E}[-Y].$$

Wtedy mamy

$$\mathbb{E}[X + \alpha Y] = \mathbb{E}[X] + \alpha \mathbb{E}[Y] \quad \text{dla} \quad \alpha \in \mathbb{R}. \tag{2.1}$$

W szczególności, jeśli $\mathbb{E}[Y] = \mathbb{E}[-Y] = 0$, to $\mathbb{E}[X + \alpha Y] = \mathbb{E}[X]$.

Dowód. Z równości (1.1) łatwo zauważyć, że

$$\mathbb{E}[\alpha Y] = \alpha^+ \mathbb{E}[Y] + \alpha^- \mathbb{E}[-Y] = \alpha^+ \mathbb{E}[Y] - \alpha^- \mathbb{E}[Y] = \alpha \mathbb{E}[Y] \text{ dla } \alpha \in \mathbb{R}.$$

Stąd mamy

$$\mathbb{E}[X + \alpha Y] \leqslant \mathbb{E}[X] + \mathbb{E}[\alpha Y] = \mathbb{E}[X] + \alpha \mathbb{E}[Y] = \mathbb{E}[X] - \mathbb{E}[-\alpha Y] \leqslant \mathbb{E}[X + \alpha Y].$$

□

Kolejny lemat jest uogólnieniem Lematu 2.1.

Lemat 2.2. Niech spełnione będą założenia Lematu 2.1. Niech $\hat{\mathbb{E}}$ będzie nieliniową wartością oczekiwaną określoną na $(\Omega, \mathscr{H})$ zdominowaną przez podliniową wartość oczekiwaną $\mathbb{E}$ w sensie Definicji 1.2. Jeśli $\mathbb{E}[Y] = -\mathbb{E}[-Y]$, to zachodzi

$$\hat{\mathbb{E}}[\alpha Y] = \alpha \hat{\mathbb{E}}[Y] = \alpha \mathbb{E}[Y] \quad \text{dla} \quad \alpha \in \mathbb{R} \tag{2.2}$$

oraz

$$\hat{\mathbb{E}}[X + \alpha Y] = \hat{\mathbb{E}}[X] + \alpha \hat{\mathbb{E}}[Y], \;\; X \in \mathscr{H}, \;\; \alpha \in \mathbb{R}. \tag{2.3}$$

W szczególności

$$\hat{\mathbb{E}}[X + c] = \hat{\mathbb{E}}[X] + c, \quad \text{dla} \quad c \in \mathbb{R}. \tag{2.4}$$

Dowód. Zauważmy, że

$$-\hat{\mathbb{E}}[Y] = \hat{\mathbb{E}}[0] - \hat{\mathbb{E}}[Y] \leqslant \hat{\mathbb{E}}[0-Y] \leqslant \mathbb{E}[0-Y] \leqslant \mathbb{E}[-Y] = -\mathbb{E}[Y] \leqslant -\hat{\mathbb{E}}[Y]$$

oraz

$$\mathbb{E}[Y] = -\mathbb{E}[-Y] \leqslant -\hat{\mathbb{E}}[-Y] = \hat{\mathbb{E}}[0] - \hat{\mathbb{E}}[-Y] \leqslant \hat{\mathbb{E}}[0+Y] \leqslant \mathbb{E}[0+Y] \leqslant \mathbb{E}[Y].$$

Z powyższych zależności wynika, że

$$\hat{\mathbb{E}}[Y] = \mathbb{E}[Y] = -\hat{\mathbb{E}}[-Y]. \tag{2.5}$$

Zauważmy, że dla $\alpha \in \mathbb{R}$ zachodzi

$$\mathbb{E}[\alpha Y] = -\mathbb{E}[-\alpha Y], \quad \text{gdy} \quad \mathbb{E}[Y] = -\mathbb{E}[-Y].$$

Wtedy mamy

$$\hat{\mathbb{E}}[\alpha Y] = \mathbb{E}[\alpha Y] = \alpha \mathbb{E}[Y] = \alpha \hat{\mathbb{E}}[Y],$$

co daje nam (2.2). Z warunku dominacji dostajemy

$$\hat{\mathbb{E}}[X+\alpha Y] - \hat{\mathbb{E}}[X] \leqslant \mathbb{E}[\alpha Y],$$

$$\hat{\mathbb{E}}[X] - \hat{\mathbb{E}}[X+\alpha Y] \leqslant \mathbb{E}[-\alpha Y].$$

W końcu otrzymujemy, że

$$\hat{\mathbb{E}}[X] + \alpha\hat{\mathbb{E}}[Y] = \hat{\mathbb{E}}[X] - \mathbb{E}[-\alpha Y] \leqslant \hat{\mathbb{E}}[X+\alpha Y] \leqslant \hat{\mathbb{E}}[X] + \mathbb{E}[\alpha Y] = \hat{\mathbb{E}}[X] + \alpha\hat{\mathbb{E}}[Y],$$

co daje nam (2.3).

□

Następny lemat podaje odpowiedniki znanych z klasycznej przestrzeni probabilistycznej nierówności momentowych.

Lemat 2.3. Dla dowolnych X oraz $Y \in \mathscr{H}$ mamy, że

$$\mathbb{E}[|X+Y|^r] \leqslant \max\{1, 2^{r-1}\}(\mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r]), \quad \text{dla} \quad r > 0; \tag{2.6}$$

$$\mathbb{E}[|XY|] \leqslant (\mathbb{E}[|X|^p])^{\frac{1}{p}} \cdot (\mathbb{E}[|Y|^q])^{\frac{1}{q}}, \quad \text{dla} \quad 1 < p,q < \infty, \quad \frac{1}{p} + \frac{1}{q} = 1; \tag{2.7}$$

$$(\mathbb{E}[|X+Y|^p])^{\frac{1}{p}} \leqslant (\mathbb{E}[|X|^p])^{\frac{1}{p}} + (\mathbb{E}[|Y|^p])^{\frac{1}{p}}, \quad \text{dla} \quad p > 1. \tag{2.8}$$

Nierówność (2.7) nazywamy nierównością Höldera.

W szczególnym przypadku, dla $1 \leqslant p < p'$, otrzymujemy nierówność Lapunowa

$$(\mathbb{E}[|X|^p])^{\frac{1}{p}} \leqslant (\mathbb{E}[|X|^{p'}])^{\frac{1}{p'}}. \tag{2.9}$$

Dowód. W dowodzie wykorzystamy dwie znane nierówności.

Lemat 2.4. [3] Dla $r > 0$, $1 < p, q < \infty$ oraz $\dfrac{1}{p} + \dfrac{1}{q} = 1$, zachodzi

$$|a+b|^r \leqslant \max\{1, 2^{r-1}(|a|^r + |b|^r)\} \quad \text{dla} \quad a, b \in \mathbb{R}; \tag{2.10}$$

$$|ab| \leqslant \frac{|a|^p}{p} + \frac{|b|^q}{q}. \tag{2.11}$$

Nierówność (2.11) nazywamy nierównością Younga. Nierówność (2.6) wynika bezpośrednio z nierówności (2.10).

Teraz udowodnimy nierówność (2.7). Rozpatrzmy przypadek $\mathbb{E}[|X|^p] \cdot \mathbb{E}[|Y|^q] > 0$. Korzystając z nierówności (2.11), otrzymujemy

$$\mathbb{E}\left[\left|\frac{X}{(\mathbb{E}[|X|^p])^{\frac{1}{p}}} \cdot \frac{Y}{(\mathbb{E}[|Y|^q])^{\frac{1}{q}}}\right|\right] \leqslant \mathbb{E}\left[\frac{1}{p}\frac{|X|^p}{\mathbb{E}[|X|^p]} + \frac{1}{q}\frac{|Y|^q}{\mathbb{E}[|Y|^q]}\right]$$

$$\leqslant \mathbb{E}\left[\frac{1}{p}\frac{|X|^p}{\mathbb{E}[|X|^p]}\right] + \mathbb{E}\left[\frac{1}{q}\frac{|Y|^q}{\mathbb{E}[|Y|^q]}\right] = \frac{1}{p}\frac{1}{\mathbb{E}[|X|^p]}\mathbb{E}[|X|^p] + \frac{1}{q}\frac{1}{\mathbb{E}[|Y|^q]}\mathbb{E}[|Y|^q]$$

$$= \frac{1}{p} + \frac{1}{q} = 1.$$

Stąd ostatecznie dostajemy

$$\mathbb{E}\left[\frac{|X|}{(\mathbb{E}[|X|^p])^{\frac{1}{p}}} \cdot \frac{|Y|}{(\mathbb{E}[|Y|^q])^{\frac{1}{q}}}\right] = \frac{1}{(\mathbb{E}[|X|^p])^{\frac{1}{p}} \cdot (\mathbb{E}[|Y|^q])^{\frac{1}{q}}}\mathbb{E}[|XY|] \leqslant 1,$$

co w konsekwencji daje nam (2.7).

W przypadku gdy $\mathbb{E}[|X|^p] \cdot \mathbb{E}[|Y|^q] = 0$, należy zastosować analogiczne rozważania, ale zamiast $\mathbb{E}[|X|^p]$ oraz $\mathbb{E}[|Y|^q]$, trzeba rozpatrzeć odpowiednio $\mathbb{E}[|X|^p] + \varepsilon_1$ oraz $\mathbb{E}[|Y|^q] + \varepsilon_2$. Dostajemy wtedy

$$\mathbb{E}\left[\left|\frac{X}{(\mathbb{E}[|X|^p] + \varepsilon_1)^{\frac{1}{p}}} \cdot \frac{Y}{(\mathbb{E}[|Y|^q] + \varepsilon_2)^{\frac{1}{q}}}\right|\right] \leqslant \mathbb{E}\left[\frac{1}{p}\frac{|X|^p}{\mathbb{E}[|X|^p] + \varepsilon_1} + \frac{1}{q}\frac{|Y|^q}{\mathbb{E}[|Y|^q] + \varepsilon_2}\right]$$

$$\leqslant \mathbb{E}\left[\frac{1}{p}\frac{|X|^p}{\mathbb{E}[|X|^p]+\varepsilon_1}\right]+\mathbb{E}\left[\frac{1}{q}\frac{|Y|^q}{\mathbb{E}[|Y|^q]+\varepsilon_2}\right]=\frac{1}{p}\frac{\mathbb{E}[|X|^p]}{\mathbb{E}[|X|^p]+\varepsilon_1}+\frac{1}{q}\frac{\mathbb{E}[|Y|^q]}{\mathbb{E}[|Y|^q]+\varepsilon_2}$$

$$\leqslant \frac{1}{p}+\frac{1}{q}=1.$$

Stąd dostajemy

$$\mathbb{E}[|XY|] \leqslant (\mathbb{E}[|X|^p]+\varepsilon_1)^{\frac{1}{p}}(\mathbb{E}[|Y|^q]+\varepsilon_2)^{\frac{1}{q}}.$$

Przy założeniu, że $\varepsilon_1 \to 0$ oraz $\varepsilon_2 \to 0$ otrzymujemy (2.7).

Teraz udowodnimy nierówność (2.8). Rozpatrzymy jedynie przypadek, gdy $\mathbb{E}[|X+Y|^p] > 0$. Wykorzystując nierówność (2.7), otrzymujemy

$$\mathbb{E}[|X+Y|^p] = \mathbb{E}[|X+Y|\cdot|X+Y|^{p-1}]$$

$$\leqslant \mathbb{E}[|X|\cdot|X+Y|^{p-1}]+\mathbb{E}[|Y|\cdot|X+Y|^{p-1}]$$

$$\leqslant (\mathbb{E}[|X|^p])^{\frac{1}{p}}\cdot(\mathbb{E}[|X+Y|^{(p-1)q}])^{\frac{1}{q}}+(\mathbb{E}[|Y|^p])^{\frac{1}{p}}\cdot(\mathbb{E}[|X+Y|^{(p-1)q}])^{\frac{1}{q}}.$$

Z zależności $(p-1)q = p$ dostajemy (2.8).

Aby udowodnić nierówność (2.9), skorzystamy z nierówności (2.7) w następującej formie

$$\mathbb{E}(|ZY|) \leqslant (\mathbb{E}[|Z|^r])^{\frac{1}{r}}\cdot(\mathbb{E}[|Y|^q])^{\frac{1}{q}},$$

gdzie $1 < r,q < \infty$ oraz $\dfrac{1}{r}+\dfrac{1}{q}=1$. Niech $Z = |X|^p$ oraz niech $Y = 1$, wtedy

$$\mathbb{E}[|X|^p] \leqslant (\mathbb{E}[|X|^{pr}]^{\frac{1}{r}}.$$

Wprowadźmy oznaczenie $pr = p'$. Wtedy

$$\mathbb{E}[|X|^p] \leqslant (\mathbb{E}[|X|^{p'}])^{\frac{p}{p'}}$$

i ostatecznie dostajemy

$$(\mathbb{E}[|X|^p])^{\frac{1}{p}} \leqslant (\mathbb{E}[|X|^{p'}])^{\frac{1}{p'}}.$$

□

W dalszej części tego paragrafu zajmiemy się nowymi wersjami fundamentalnych z punktu widzenia teorii prawdopodobieństwa twierdzeń i nierówności, takich jak lemat Borella-Cantelliego, centralne twierdzenie graniczne, nierówności Czebyszewa, Kołmogorowa, Jensena, Rosenthala.

Twierdzenie 2.1. *(Nierówność Chebyszewa [1]) Niech $X \in \mathscr{H}$ będzie zmienną losową oraz niech $f \in \mathscr{C}_{l.Lip}(\mathbb{R})$ będzie niemalejącą funkcją o wartościach dodatnich. Wtedy dla dowolnego $\varepsilon \in \mathbb{R}$ mamy*

$$\mathbb{V}(X \geqslant \varepsilon) \leqslant \frac{\mathbb{E}[f(X)]}{f(\varepsilon)}, \qquad \mathscr{V}(X \geqslant \varepsilon) \leqslant \frac{\mathscr{E}[f(X)]}{f(\varepsilon)}.$$

Twierdzenie 2.2. *(Nierówność Kolmogorova) Niech $\{X_1,...,X_n\}$ będzie ciągiem zmiennych losowych na $(\Omega,\mathscr{H},\mathbb{E})$ z $\mathbb{E}[X_k] = 0$, $k = 1,...,n$. Załóżmy, że X_k jest niezależne od $(X_{k+1},...,X_n)$ dla każdego $k = 1,...,n-1$. Niech $S_k = X_1 + ... + X_k$ oraz $S_0 = 0$. Wtedy*

$$\mathbb{E}\left[\left(\max_{k\leqslant n} S_k\right)^2\right] \leqslant \sum_{k=1}^{n} \mathbb{E}[X_k^2].$$

Dowód. Niech

$$T_k = \max(X_k, X_k + X_{k+1}, ..., X_k + \cdots + X_n).$$

Wtedy $T_k, T_k^+ \in \mathscr{H}$ oraz prawdziwe są następujące zależności

$$T_k = X_k + T_{k+1}^+$$

oraz

$$T_k^2 = X_k^2 + 2X_k T_{k+1}^+ + (T_{k+1}^+)^2.$$

Stąd dostajemy

$$\mathbb{E}[T_k^2] \leqslant \mathbb{E}[X_k^2] + 2\mathbb{E}[X_k T_{k+1}^+] + \mathbb{E}[(T_{k+1}^+)^2].$$

Zauważmy ponadto, że $\mathbb{E}[X_k T_{k+1}^+] = 0$, co pozwala stwierdzić, że

$$\mathbb{E}[T_k^2] \leqslant \mathbb{E}[X_k^2] + \mathbb{E}[(T_{k+1}^+)^2] \leqslant \mathbb{E}[X_k^2] + \mathbb{E}[T_{k+1}^2].$$

Stąd mamy

$$\mathbb{E}[T_1^2] \leqslant \mathbb{E}[X_1^2] + \mathbb{E}[T_2^2] \leqslant \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + ... + \mathbb{E}[T_n^2] = \sum_{k=1}^{n} \mathbb{E}[X_k^2],$$

co daje nam tezę twierdzenia. □

Twierdzenie 2.3. *(Nierówność Jensena [5]) Niech X będzie zmienną losową z przestrzeni podliniowej wartości oczekiwanej $(\Omega,\mathscr{H},\mathbb{E})$. Niech $f(\cdot)$ będzie wypukłą*

funkcją określoną na $\mathbb{R}$. *Załóżmy, że* $\mathbb{E}[X]$ *oraz* $\mathbb{E}[f(X)]$ *są skończone. Wtedy*

$$\mathbb{E}[f(X)] \geqslant f(\mathbb{E}[X]).$$

Twierdzenie 2.4. *(Nierówność typu Rosenthala [7]) Niech* $\{X_1,...,X_n\}$ *będzie ciągiem zmiennych losowych określonych na przestrzeni* $(\Omega,\mathscr{H},\mathbb{E})$ *oraz niech* $S_k = X_1 + ... + X_k$, $S_0 = 0$.

(a) Załóżmy, że X_k *jest zmienną losową ujemnie zależną względem* $(X_{k+1},...,X_n)$ *dla każdego* $k = 1,...,n-1$ *oraz* $\mathbb{E}[X_k] \leqslant 0$, *gdzie* $k = 1,...,n$. *Wtedy*

$$\mathbb{E}\left[\left|\max_{k\leqslant n} S_k\right|^p\right] \leqslant 2^{2-p}\sum_{k=1}^{n}\mathbb{E}[|X_k|^p], \quad dla \quad 1 \leqslant p \leqslant 2$$

oraz

$$\mathbb{E}\left[\left|\max_{k\leqslant n} S_k\right|^p\right] \leqslant C_p n^{\frac{p}{2}-1}\sum_{k=1}^{n}\mathbb{E}[|X_k|^p], \quad dla \quad p \geqslant 2.$$

(b) Załóżmy, że X_k *jest zmienną losową niezależną względem* $(X_{k+1},...,X_n)$ *dla każdego* $k = 1,...,n-1$ *oraz* $\mathbb{E}[X_k] \leqslant 0$, *gdzie* $k = 1,...,n$. *Wtedy*

$$\mathbb{E}\left[\left|\max_{k\leqslant n} S_k\right|^p\right] \leqslant C_p\left\{\sum_{k=1}^{n}\mathbb{E}[|X_k|^p] + \left(\sum_{k=1}^{n}\mathbb{E}[|X_k|^2]\right)^{\frac{p}{2}}\right\}, \quad dla \quad p \geqslant 2.$$

(c) Załóżmy, że X_k *jest ujemnie zależna względem* $(X_{k+1},...,X_n)$ *dla każdego* $k = 1,...,n-1$ *lub* X_{k+1} *jest ujemnie zależna względem* $(X_1,...,X_k)$ *dla każdego* $k = 1,...,n-1$. *Wtedy*

$$\mathbb{E}\left[\max_{k\leqslant n}|S_k|^p\right] \leqslant C_p\left\{\sum_{k=1}^{n}\mathbb{E}[|X_k|^p] + \left(\sum_{k=1}^{n}\mathbb{E}[|X_k|^2]\right)^{\frac{p}{2}} + \left(\sum_{k=1}^{n}[(\mathscr{E}[X_k])^- + (\mathbb{E}[X_k])^+]\right)^p\right\}.$$

Lemat 2.5. (Lemat Borela-Cantelliego) Niech $\{A_n, n \geqslant 1\}$ będzie ciągiem zdarzeń z σ-ciała $\mathscr{F}$. Niech $(\mathbb{V},\mathscr{V})$ będzie parą pojemności generowaną przez podliniową wartość oczekiwaną $\mathbb{E}$.

(1) Jeśli $\sum_{n=1}^{\infty}\mathbb{V}(A_n) < \infty$, to

$$\mathbb{V}\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}A_i\right) = 0.$$

(2) Niech $\mathscr{V}$ będzie ciągła od od dołu oraz niech $\{A_n^c\}_{n=1}^{\infty}$ będą wzajemnie niezależne względem $\mathscr{V}$, tzn. dla dowolnego $n \in \mathbb{N}$ zachodzi

$$\mathscr{V}\left(\bigcap_{i=n}^{\infty} A_i^c\right) = \prod_{i=n}^{\infty} \mathscr{V}(A_i^c).$$

Jeśli

$$\sum_{n=1}^{\infty} \mathbb{V}(A_n) = \infty, \tag{2.12}$$

to

$$\mathbb{V}\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\right) = 1.$$

Dowód. Łatwo zauważyć, że z własności monotoniczności oraz skończonej podaddytywności wynika, że

$$0 \leqslant \mathbb{V}\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\right) \leqslant \mathbb{V}\left(\bigcup_{i=n}^{\infty} A_i\right) \leqslant \sum_{i=n}^{\infty} \mathbb{V}(A_i) \to 0, \quad \text{gdy } n \to \infty,$$

co kończy dowód pierwszej części lematu.

Teraz udowodnimy drugą część lematu. Z faktu, że $\mathscr{V}$ jest ciągła od dołu dostajemy, że $\mathbb{V}$ jest ciągła od góry i na podstawie założenia (2.12) mamy wtedy

$$0 \leqslant 1 - \mathbb{V}\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\right) = 1 - \lim_{n\to\infty} \mathbb{V}\left(\bigcup_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty}\left[1 - \mathbb{V}\left(\bigcup_{i=n}^{\infty} A_i\right)\right]$$

$$= \lim_{n\to\infty} \mathscr{V}\left(\bigcap_{i=n}^{\infty} A_i^c\right) = \lim_{n\to\infty} \prod_{i=n}^{\infty} \mathscr{V}(A_i^c) = \lim_{n\to\infty} \prod_{i=n}^{\infty} (1 - \mathbb{V}(A_i))$$

$$\leqslant \lim_{n\to\infty} \prod_{i=n}^{\infty} \exp(-\mathbb{V}(A_i)) = \lim_{n\to\infty} \exp\left(-\sum_{i=n}^{\infty} \mathbb{V}(A_i)\right) = 0.$$

Tym samym dowiedliśmy prawdziwości drugiej części lematu. □

Twierdzenie 2.5. *(Centralne twierdzenie graniczne [6]) Załóżmy, że $\{X_n, n \geqslant 1\}$ jest ciągiem niezależnych zmiennych losowych określonych na przestrzeni $(\Omega, \mathscr{H}, \mathbb{E})$ o takim samym rozkładzie. Niech*

$$\mathbb{E}[X_1] = \mathbb{E}[-X_1] = 0, \quad \lim_{c\to\infty} \mathbb{E}[(X_1^2 - c)^+] = 0 \quad \textit{oraz} \quad \overline{\sigma}^2 = \mathbb{E}[X_1^2], \quad \underline{\sigma}^2 = \mathscr{E}[X_1^2].$$

Wtedy dla dowolnej ciągłej funkcji φ spełniającej warunek $|\varphi(x)| \leqslant C(1+x^2)$, gdzie $C > 0$, mamy

$$\lim_{n\to\infty} \mathbb{E}\left[\varphi\left(\frac{S_n}{\sqrt{n}}\right)\right] = \mathbb{E}[\varphi(\xi)], \tag{2.13}$$

gdzie $\xi \sim N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, tzn. zmienna losowa ξ ma rozkład normalny ze średnią równą zero oraz posiada niepewność wariancji, opisaną za pomocą, odpowiednio, dolnej i górnej wariancji. Dodatkowo, dla $p > 2$ oraz $\mathbb{E}[|X_1|^p] < \infty$ równość (2.13) zachodzi dla dowolnej ciągłej funkcji φ spełniającej warunek $|\varphi(x)| \leqslant C(1+|x|^p)$.

3. Podliniowy model regresji

Regresja liniowa jest to metoda statystyczna pozwalająca ocenić wpływ wielu różnych cech na pewną inną cechę, która szczególnie nas interesuje w badaniu statystycznym. Jest to narzędzie bardzo często wykorzystywane w modelowaniu w klasycznym przypadku. Ważnym, wymagającym rozważenia problemem jest sprawdzenie, jak model regresji liniowej będzie wyglądał w przypadku, gdy modelujemy zjawisko zachodzące w warunkach niepewności.

Rozważmy następujący model regresji liniowej

$$Y = \beta'\mathbf{x} + \varepsilon, \tag{3.1}$$

gdzie Y jest zmienną zależną, $\mathbf{x} = (X_1, ..., X_p)'$ jest p-wymiarową zmienną towarzyszącą, posiadającą określony rozkład $F_{\mathbf{x}}(x)$, oraz $\beta = (\beta_1, ..., \beta_p)'$ jest p-wymiarowym wektorem nieznanych parametrów (zauważmy, że wyraz wolny modelu jest równy zero), a $\varepsilon = (\varepsilon_1, ..., \varepsilon_p)$ jest błędem losowym.

Losowość błędu zapewnia, że wartość oczekiwana tego błędu jest zerowa. Zakładamy również, że błąd ε jest niezależny od $\mathbf{x}$.

Zdefiniujemy teraz pojęcie podliniowej warunkowej wartości oczekiwanej $\mathbb{E}[\cdot|\mathbf{x}]$ pod warunkiem $\mathbf{x}$.

Definicja 3.1. Podliniową warunkową wartość oczekiwaną zmiennej losowej $Y \in \mathscr{H}$ pod warunkiem $\mathbf{x}$ definiujemy następująco

$$\mathbb{E}[Y|\mathbf{x}] = \sup_{f_{Y|\mathbf{x}} \in \mathscr{F}_{Y|\mathbf{x}}} E_{f_{Y|\mathbf{x}}}[Y|\mathbf{x}],$$

gdzie $\{E_{f_{Y|\mathbf{x}}} : f_{Y|\mathbf{x}} \in \mathscr{F}_{Y|\mathbf{x}}\}$ jest rodziną liniowych warunkowych wartości oczekiwanych.

Uwaga 3.1. Tak zdefiniowana podliniowa warunkowa wartość oczekiwana zachowuje własności: monotoniczności, zachowania stałej, podaddytywności oraz nieujemnej jednorodności.

Uwaga 3.2. Zauważmy, że na początku założono, że wektor towarzyszący **x** ma ustalony rozkład $F_{\mathbf{x}}$ i wyraz wolny w modelu (3.1) jest równy zero. Założenie o pewności rozkładu zmiennej **x** jest konieczne, aby wektor współczynników β mógł być jednoznacznie wyznaczony. W sytuacji, kiedy zarówno ε oraz **x** posiadają niepewność rozkładu, β nie może zostać wyznaczone jednoznacznie.

3.1. *G*-normalna regresja

W pierwszej kolejności rozpatrzmy model, w którym błąd ε ma rozkład *G*-normalny. W tym celu zdefiniujemy najpierw pojęcie rozkładu *G*-normalnego.

Definicja 3.2. (Rozkład *G*-normalny [3]) Mówimy, że d-wymiarowy wektor losowy $X = (X_1, ..., X_d)$ określony na przestrzeni podliniowej wartości oczekiwanej $(\Omega, \mathscr{H}, \mathbb{E})$ ma rozkład *G*-normalny jeśli

$$aX + b\overline{X} \stackrel{d}{=} \sqrt{a^2 + b^2}X \quad \text{dla} \quad a, b \geqslant 0,$$

gdzie $\overline{X}$ jest niezależną kopią X.

Uwaga 3.3. Zauważmy, że dla rozkładu *G*-normalnego mamy

$$\mathbb{E}[X+\overline{X}] = \mathbb{E}[\mathbb{E}[x+\overline{X}]]_{x=X} = \mathbb{E}[x+\mathbb{E}[\overline{X}]]_{x=X} = \mathbb{E}[X+\mathbb{E}[X]] = \mathbb{E}[X]+\mathbb{E}[X] = 2\mathbb{E}[X]$$

oraz

$$\mathbb{E}[X+\overline{X}] = \mathbb{E}[\sqrt{2}X] = \sqrt{2}\mathbb{E}[X].$$

Stąd dostajemy $\mathbb{E}[X] = 0$. Podobnie można pokazać, że $\mathbb{E}[-X] = 0$. Wynika z tego, że zmienna o rozkładzie *G*-normalnym nie ma niepewności średniej, ale posiada niepewność wariancji.

Jak wspomniano wyżej, ε ma rozkład *G*-normalny, tzn.

$$\varepsilon \sim \mathscr{N} = N(\{0\} \times [\underline{\sigma}^2, \overline{\sigma}^2]).$$

Z własności *cash translatability* wynika, że dla modelu regresji (3.1), w którym ε ma rozkład *G*-normalny, zachodzi warunek

$$\mathbb{E}[Y|\mathbf{x}] = \beta'\mathbf{x}. \tag{3.2}$$

Uwaga 3.4. [2]

(1) Jeśli ε ma rozkład G-normalny, to $\mathbb{E}[Y|\mathbf{x}]$ może być jednoznaczne wyznaczona jako $\beta'\mathbf{x}$.

(2) Jeśli $E[\mathbf{x}\mathbf{x}']$ jest dodatnio określoną macierzą, wtedy β może być jednoznacznie wyznaczona jako

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\}, \tag{3.3}$$

gdzie liniowa wartość oczekiwana E pochodzi z ustalonego rozkładu $F_{\mathbf{x}}(x)$.

Dowód. Zauważmy, że

$$\beta'\mathbf{x} = (\beta_1,\beta_2,\ldots,\beta_p)\cdot(X_1,X_2,\ldots,X_p)' = \beta_1\cdot X_1+\beta_2\cdot X_2+\cdots+\beta_p\cdot X_p$$

$$= X_1\cdot\beta_1+X_2\cdot\beta_2+\cdots+X_p\beta_p = (X_1,X_2,\ldots,X_p)\cdot(\beta_1,\beta_2,\ldots,\beta_p)' = \mathbf{x}'\beta.$$

Z (3.2) dostajemy, że

$$\mathbf{x}\mathbb{E}[Y|\mathbf{x}] = \mathbf{x}\mathbf{x}'\beta.$$

Stąd otrzymujemy, że

$$E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\} = E[\mathbf{x}\mathbf{x}']\beta.$$

Powyższa zależność daje nam (3.3). □

3.2. Regresja z podliniową wartością oczekiwaną

Teraz rozpatrzmy model, w którym błąd ε posiada niepewność średniej oraz niepewność wariancji. Ponownie z własności *cash translatability* dla podliniowej wartości oczekiwanej mamy

$$\mathbb{E}[Y|\mathbf{x}] = \beta'\mathbf{x}+\overline{\mu}. \tag{3.4}$$

Uwaga 3.5. [2]

(1) Jeśli $\underline{\mu} < \overline{\mu}$, to dla ustalonego $\mathbf{x}$ podliniowa wartość oczekiwana zmiennej Y ma zmienność rzędu $\overline{\mu}$ lub, innymi słowy, podliniowa wartość oczekiwana zmiennej Y może być wyznaczona jako $\mathbb{E}[Y|\mathbf{x}] = \beta'\mathbf{x}+\overline{\mu}$.

(2) Jeśli $E[\mathbf{x}\mathbf{x}']$ jest dodatnio określoną macierzą, to β może być jednoznacznie wyrażone jako

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\}-\overline{\mu}(E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}]. \tag{3.5}$$

W szczególności, dla $E[\mathbf{x}] = 0$ zachodzi warunek

$$\beta = (E[\mathbf{x}\mathbf{x}'])^{-1}E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\}.$$

Dowód. Z (3.4) mamy

$$\mathbf{x}\mathbb{E}[Y|\mathbf{x}] = \mathbf{x}\mathbf{x}'\beta + \overline{\mu}\mathbf{x}$$

i dalej dostajemy

$$E\{\mathbf{x}\mathbb{E}[Y|\mathbf{x}]\} = E[\mathbf{x}\mathbf{x}']\beta + \overline{\mu}E[\mathbf{x}].$$

Powyższa zależność daje nam (3.5). □

4. Podsumowanie

Podliniowa wartość oczekiwana jest nowym i oryginalnym zagadnieniem z zakresu teorii prawdopodobieństwa. Na podstawie niniejszego rozdziału można zauważyć, że wiele definicji, lematów oraz twierdzenień zostało dobrze zaadaptowanych do tej specjalnej przestrzeni. W literaturze można znaleźć liczne prace, w których przedstawione są modele statystyczne przystosowane do warunków podliniowej wartości oczekiwanej, jak np. opisana w tej pracy regresja liniowa. Okazuje się, że podliniowa wartość oczekiwana generuje wiele ciekawych własności, które różnią się od tradycyjnej liniowej wartości oczekiwanej. Często te własności stanowią większe wyzwanie, np. przy dowodzeniu twierdzeń. Przestrzeń podliniowej wartości oczekiwanej może znaleźć różne zastosowania, szczególnie w finansach podczas pomiaru ryzyka. Wiele problemów związanych z tą przestrzenią zostało już rozwiązanych, jednak istnieje jeszcze liczna grupa zagadnień, która wymaga zbadania.

Literatura

[1] Huang W., Wu P., *Strong Laws of Large Numbers for General Random Variables in Sublinear Expectation Spaces*, „Journal of Inequalities and Applications", 2019, 2019:143.

[2] Lin L., Shi Y., Wang X., Yang S., *Sublinear Expectation Linear Regression*, „Statistics Theory", 2013, https://doi.org/10.48550/arXiv.1304.3559.

[3] Peng S., *Nonlinear Expectations and Stochastic Calculus under Uncertainty with Robust CLT and G-Brownian Motion*, Germany, Probability Theory and Stochastic Modelling, 2019.

[4] Wang Y., Li Y., Gao Q., *On the Exponential Inequality for Acceptable Random Variables*, „Journal of Inequalities and Applications", 2011, 2011:40.

[5] Xu J.P., Zhang L.X., *Three Series Theorem for Independent Random Variables under Sublinear Expectations with Applications*, „Acta Mathematica Sinica, English Series", 2019, vol. 35, p. 172–184.

[6] Zhang L., *Exponential Inequalities Under the Sublinear Expectations with Applications to Laws of the Iterated Logarithm*, „Science China Mathematics", 2016, vol. 59(12), p. 2503–2526.

[7] Zhang L., *Rosenthal's Inequalities for Independent and Negatively Dependent Random Variables under Sublinear Expectations with Applications*, „Science China Mathematics", 2015, vol. 59(4), p. 751–768.

[8] Zhong H., Wu Q., *Complete Convergence and Complete Moment Convergence for Weighted Sums of Extended Negatively Dependent Random Variables under Sublinear Expectation*, „Journal of Inequalities and Applications", 2017, 2017:261.

Yaroslav Chabanyuk[1], **Uliana Khimka** [2]

Ciągła procedura aproksymacji stochastycznej w środowisku półmarkowowskim

Streszczenie

Dla ciągłej procedury aproksymacji stochastycznej w środowisku półmarkowowskim otrzymano warunki zbieżności do punktu równowagi dla układu uśrednianego względem rozkładu ergodycznego wbudowanego łańcucha Markowa. Ponadto zostały zbadane procedury zarówno dla schematu uśredniania jak i dla schematu aproksymacji dyfuzyjnej w środowisku półmarkowowskim, z wykorzystaniem własności asymptotycznej operatora kompensującego i rozwiązania problemu zaburzenia osobliwego dla tego operatora.

Słowa kluczowe: *ciągła procedura aproksymacji stochastycznej, środowisko półmarkowowskie, zbieżność*

Abstract

For the continuous stochastic approximation procedure in the semi-Markov environment, sufficient conditions were obtained to converge to the equilibrium point for the system averaged relative to the ergodic distribution of the Markov chain inserted. In addition, the procedures were examined in both the averaging scheme and the diffusion approximation scheme in the smelting environment has been developed, using the asymptotic properties of the compensation operator and solving the problem of a peculiar disorder for this operator.

Keywords: *stochastic approximation method, semi-Markov environment, convergence*

Wstęp

Głównym problemem dotyczącym procedury aproksymacji stochastycznej (PAS) jest zbieżność procedury przy wykorzystaniu najmniejszej ilości informacji o funkcji regresji oraz charakterystyk tej funkcji w punkcie obserwacji.

W 1951 r. w pracy H. Robbinsa i S. Monro [34] została zaproponowana dyskretna (rekurencyjna) procedura aproksymacji stochastycznej wyznaczania pierwiastka u_0 równania regresji

$$C(u) = 0.$$

W pracy zbadano zbieżność średniokwadratową.

[1] Katedra Matematyki Stosowanej, Politechnika Lubelska

[2] Wydział Matematyki Stosowanej i Informatyki, Lwowski Uniwersytet imienia Iwana Franki

Ciągły odpowiednik procedury Robbinsa–Monro był po raz pierwszy rozpatrzony przez M. Drimla i J. Nedoma [12], którzy dowiedli jej zbieżności przy dość ogólnych warunkach nałożonych na procesy losowe. Zbieżność ciągłego odpowiednika procedury Kiefera-Wolfowitza udowodnił D.T. Sakrison.

Ciągła PAS z addytywnym białym szumem została wprowadzona przez R.Z. Hasminskiego i była badana w pracach E.B. Newelsona, L.G. Gladyshewa, R.G. Gjachkowa, E.N. Pynskera, E. Kifera i innych autorów. W szczególności, w pracy L. Ljunga, G. Pfluga i H. Walka [31] zbadano modyfikację procedury Robbinsa–Monro i wskazano na szereg zastosowań do znajdowania minimum globalnego funkcji $C(u), u \in S, S \in R^d$, gdzie S jest obszarem ograniczonym, oraz minimum lokalnego dla $u \in S \bigcap U$, gdzie U jest otwartym otoczeniem punktu u_0.

W pracy E.B. Newelsona i R.Z. Hasminskiego [33] dowiedziono zbieżności procedur stochastycznych, wykorzystując własności funkcji Lapunowa. Umożliwiło to uproszczenie badań i zastosowanie PAS.

Z drugiej strony zauważmy, że PAS stanowi podstawę teoretyczną optymalizacji konstrukcji ciągu zbieżnego do punktu równowagi w warunkach skomplikowanych wpływów środowiska zewnętrznego na funkcję regresji albo, co na jedno wychodzi, wpływów na układ stochastyczny.

Jednym z najważniejszych zadań ogólnej teorii układów ewolucyjnych w środowisku stochastycznym jest znalezienie warunków stabilności takich układów. Po raz pierwszy stabilność układu dynamicznego, który spełnia zasady uśredniania, została zbadana przez M.M. Bogolubowa [7] (zobacz też U.O. Mytropolski, A.M. Samojlenko [32]). Badanie stabilności układów dynamicznych w środowisku stochastycznym rozpoczęli R. Griego i T. Hersh [14] od rozwoju teorii ewolucji losowych. Stabilność układu dynamicznego z zaburzeniami markowowskimi z warunkiem stabilności układu uśrednionego została zbadana w pracach A.W. Skorochoda i E.F. Tsarkowa, a także V.S. Korolyuka [18]. Problem stabilności w warunkach aproksymacji dyfuzyjnej układu dynamicznego z markowowskimi zaburzeniami po raz pierwszy został rozwiązany za pomocą martyngałowej charakteryzacji odpowiedniego procesu Markowa w artykule G.L. Blankenshipa i G.C. Papanicolaou [6].

Konstrukcja procesu półmarkowowskiego oraz badanie asymptotycznych własności procesów losowych z przełączeniami półmarkowowskimi są rozważane w [1–4]. Warto zaznaczyć, że w pracy [28], w której analizowano asymptotyczne własności procesu półmarkowowskiego z liniowo zaburzonym operatorem, zastosowano własności półgrupy z procesem Markowa. Ostatnie wyniki są opisane w [21]. Klasyfikacja problemu rozwiązywania pojedynczej perturbacji procesu losowego z przełączeniami półmarkowowskimi jest opisana w [25] i [22], wraz z użyciem operatora kompensującego [38]. Wraz z operatorem kompensującym [10] możemy otrzymać warunki do zbieżności losowej ewolucji z przełączeniem półmarkowowskim

do procesu dyfuzyjnego w schemacie uśrednienia [20]. Wyniki tych badań zostały użyte w różnych pracach [8–10, 15].

Dlatego zacznijmy badania od warunków stabilności takich układów z przełączeniami półmarkowowskimi.

1. Środowisko półmarkowowskie

Zastosowanie algorytmów fazowego uśrednienia procesów półmarkowowskich, a także algorytmów dyfuzyjnego przybliżenia stochastycznych ewolucji jest wykorzystywane przez metody martyngałowe we współczesnej teorii procesów losowych [21, 23–26, 29].

Dalej zakładamy, że wszystkie operatory są zdefiniowane w przestrzeni Banacha, wszystkich funkcji rzeczywistych z normą supremum

$$\|\varphi\| := \sup_{x\in X} |\varphi(x)|.$$

Definicja 1.1. Proces półmarkowowski $x(t), t \geqslant 0$, w standardowej przestrzeni stanów $(X, \mathscr{E})$ definiujemy przez półmarkowowskie jądro

$$Q(x,B,t) = P(x,B)G_x(t), \quad x \in X, \quad B \in X, \quad t \geqslant 0,$$

gdzie jądro stochastyczne $P(x,B)$ zadaje prawdopodobieństwa przełączeń

$$P(x,B) = P\{\, x_{n+1} \in B | x_n = x\},$$

oraz definiuje wbudowany łańcuch Markowa $x_n = x(\tau_n)$ w momentach odnowy

$$\tau_n = \sum_{k=1}^{n} \theta_k, n \geqslant 0, \quad \tau_0 = 0,$$

z przedziałami $\theta_{k+1} = \tau_{k+1} - \tau_k$ pomiędzy momentami odnowy. Ponadto θ_n są zdefiniowane przez funkcję dystrybuanty

$$G_x(t) = P\{\theta_{n+1} \leqslant t | x_n = x\} =: P\{\theta_x \leqslant t\}.$$

Proces półmarkowowski jest zdefiniowany wzorem

$$x(t) = x_{\nu(t)}, \quad t \geqslant 0,$$

gdzie proces obliczania $\nu(t)$ jest dany wzorem

$$\nu(t) := \max\{n : \tau_n \leqslant t\}, \quad t \geqslant 0.$$

Rozważamy proces półmarkowowski $x(t)$, $t \geqslant 0$, który jest regularny i jednostajnie ergodyczny z rozkładem stacjonarnym $\pi(B)$, $B \in \mathscr{E}$

$$\pi(dx) = \rho(dx)g(x)/m.$$

Tutaj $\rho(B)$, $B \in \mathscr{E}$ jest rozkładem stacjonarnym wbudowanego łańcucha Markowa.

Definicja 1.2. Proces towarzyszący Markowa $x_0(t)$, $t \geqslant 0$, jest zdefiniowany generatorem

$$\mathbf{Q}\varphi(x) = q(x)\int_X P(x,dy)[\varphi(y) - \varphi(x)], \tag{1.1}$$

gdzie

$$q(x) := \quad g^{-1}(x), \quad g(x) = E\theta_x = \int_0^\infty \bar{G}_x(t)dt, \quad \bar{G}_x(t) = 1 - G_x(t).$$

Generator $\mathbf{Q}$ jest normalnie rozwiązywalny z potencjałem R_0 [21].

2. Stabilność układów stochastycznych z przełączeniami półmarkowowskimi

W tym podrozdziale ustalono warunki wystarczające stabilności stochastycznej systemów w środowisku półmarkowowskim stosowanych w badaniu zbieżności i asymptotycznej normalności procedury aproksymacji stochastycznej, a także metody konstrukcji operatora kompensującego i jego rozkłady asymptotyczne w schemacie uśredniania i w schemacie przybliżenia dyfuzyjnego.

2.1. Stabilność układów stochastycznych w schemacie uśrednienia

Wyjściowy dynamiczny układ autonomiczny z przełączeniami półmarkowowskimi w schemacie uśrednienia z małym parametrem serii $\varepsilon > 0$ jest dany przez ewolucyjne równanie różniczkowe

$$du^\varepsilon(t)/dt = C(u^\varepsilon(t), x(t/\varepsilon)), u^\varepsilon(0) = u_0, \tag{2.1}$$

w przestrzeni euklidesowej $R^d : u^\varepsilon(t) = (u_k^\varepsilon(t); k = \overline{1,d})$.

Prędkości układu dynamicznego są określone funkcją wektorową $C(u,x) = (C_k(u,x), k = \overline{1,d})$ i spełniają warunki istnienia globalnych rozwiązań układów deterministycznych

$$\frac{du_x^{\varepsilon}(t)}{dt} = C(u_x^{\varepsilon}(t),x), x \in X, \tag{2.2}$$

zależących od stanu fazowego $x \in X$. Uśredniony układ dynamiczny jest wyznaczany przez rozwiązanie ewolucyjnego równania deterministycznego

$$d\widehat{u}(t)/dt = C(\widehat{u}(t)), \widehat{u}(t) = u_0. \tag{2.3}$$

Średnia prędkość jest dana przez równość

$$C(u) = \int_X \pi(dx) C(u,x). \tag{2.4}$$

Zadaniem jest zapewnienie stabilności uśrednionego systemu (2.3) w celu ustalenia dodatkowych warunków zapewniających stabilność pierwotnego układu stochastycznego (2.1) dla wystarczająco małych wartości parametru serii $\varepsilon > 0$.

Twierdzenie 2.1. *Załóżmy, że dla uśrednionego układu dynamicznego* (2.3)–(2.4) *istnieje funkcja Lapunowa $V(u), u \in R^d$, dla której są spełnione następujące warunki*

*C*1: *warunek stabilności wykładniczej*

$C(u)V'(u) \leqslant -c_0 V(u), c_0 > 0$;

dodatkowe warunki

*C*2: $|C(u,x)V'(u)| \leqslant c_1 V(u)$;

*C*3: $|C(u,x)[C(u,x)V'(u)]'| \leqslant c_2 V(u)$;

*C*4: *dystrybuanty $G_x(t), t \geqslant 0, x \in X$, czasów w stanach procesu półmarkowowskiego $x(t), t \geqslant 0$, (patrz 1.2.2) spełniają warunek Cramera, jednostajnie wzdłuż $x \in X$,*

$$\sup_{x \in X} \int_0^{\infty} e^{ht} \overline{G}_x(t) dt \leqslant H < +\infty, h > 0,$$

*C*5: *a także mają miejsce oszacowania $0 < \underline{m} \leqslant g(x) \leqslant \overline{m} < +\infty$.*

Wtedy dla wszystkich $\varepsilon \leqslant \varepsilon_0$, gdzie ε_0 jest wystarczająco małe, rozwiązanie równania ewolucji (2.1) *jest asymptotycznie stabilne z prawdopodobieństwem jeden*

$$P\{\lim_{t \to \infty} \|u^{\varepsilon}(t)\| = 0\} = 1 \tag{2.5}$$

dla wszystkich warunków początkowych $|u^{\varepsilon}(0)| \leqslant u^*$, gdzie u^* wystarczająco małe.

Uwaga 2.1. Przy losowych warunkach początkowych jest spełniony warunek $|Eu^{\varepsilon}(0)| \leqslant u^*$ dla wystarczająco małego u^*.

W celu przeprowadzenia dowodu twierdzenia 2.1 sformułujemy potrzebne definicje i udowodnimy serię lematów.

Definicja 2.1. Rozszerzony proces odnowy Markowa (RPOM) $(u_n^{\varepsilon}, x_n^{\varepsilon})$ jest definiowany następująco

$$u_n^{\varepsilon} = u^{\varepsilon}(\tau_n^{\varepsilon}), x_n^{\varepsilon} = x(\tau_n^{\varepsilon}), \tau_n^{\varepsilon} = \varepsilon\tau_n, n \geqslant 0. \tag{2.6}$$

Definicja 2.2. Operator kompensujący dla RPOM (2.6) ustala się na podstawie wzoru

$$\boldsymbol{L}^{\varepsilon}\varphi(u,x) := \varepsilon^{-1}q(x)E[\varphi(u_{n+1}^{\varepsilon}, x_{n+1}^{\varepsilon},) - \varphi(u,x)| \\ u_n^{\varepsilon} = u, x_n^{\varepsilon} = x, \tau_n^{\varepsilon} = t]. \tag{2.7}$$

Niech zbiór półgrup $\boldsymbol{C}_t(x), t \geqslant 0, x \in X$, bedzie generowany przez towarzyszący układ (2.2), tj. ustalany przez generator

$$\boldsymbol{C}(x)\varphi(u) = C(u,x)\varphi'(u). \tag{2.8}$$

Lemat 2.1. Operator kompensujący dla RPOM (2.6) ma reprezentację analityczną

$$\mathbf{L}^{\varepsilon}\varphi(u,x) = \varepsilon^{-1}q(x)[\int_0^{\infty} G_x(ds)\mathbf{C}_{\varepsilon s}(x)\int_X P(x,dy)\varphi(u,y) - \varphi(u,x)].$$

Dowód. Najpierw obliczmy warunkową wartość oczekiwaną, używając półgrup $\boldsymbol{C}_t(x)$. Półgrupę $\boldsymbol{C}_t(x)$ w przestrzeni Banacha $C(R^d)$ ciągłych funkcji o wartościach rzeczywistych $\varphi(u), u \in R^d$, z normą maksimum, określa się poprzez układ

$$\boldsymbol{C}_{t'}(x) := \varphi(u(t+t')).$$

Dla $t' = \theta_x^{\varepsilon} = \varepsilon\theta_x$, mamy

$$\boldsymbol{C}_{\theta_x^{\varepsilon}}(x)\varphi(u) = \varphi(u_{n+1}^{\varepsilon})$$

albo

$$\boldsymbol{C}_{\varepsilon\theta_x}(x)\varphi(u) = \varphi(u_{n+1}^{\varepsilon}).$$

Biorąc to drugie pod uwagę, mamy warunkową wartość oczekiwaną

$$E[\varphi(u_{n+1}^{\varepsilon})|u_n^{\varepsilon} = u, x_n^{\varepsilon} = x, \tau_n^{\varepsilon} = t] = E_{u,x,t}\boldsymbol{C}_{\varepsilon\theta_x}(x)\varphi(u).$$

Biorąc pod uwagę reprezentację jądra półmarkowowskiego $Q(x,dy,dt) = G_x(dt)P(x,dy)$, w końcu otrzymujemy

$$E[\varphi(u^\varepsilon_{n+1},x^\varepsilon_{n+1})|u^\varepsilon_n = u, x^\varepsilon_n = x, \tau^\varepsilon_n = t] =$$

$$= E_{u,x,t}\varphi(u^\varepsilon_{n+1},x^\varepsilon_{n+1}) = \int_0^\infty G_x(ds)\mathbf{C}_{\varepsilon s}(x)\int_X P(x,dy)\varphi(u,y).$$

Podstawiając ostatnią reprezentację w (2.7), otrzymujemy tezę lematu. □

Lemat 2.2. Operator kompensujący na funkcjach testowych $\varphi(u,x) \in C^2(R^d \times X)$ ma reprezentacje asymptotyczne

$$\boldsymbol{L}^\varepsilon\varphi(u,x) = \varepsilon^{-1}\mathbf{Q}\varphi(u,x) + \theta^\varepsilon_1(x)\varphi(u,x), \tag{2.9}$$

oraz

$$\boldsymbol{L}^\varepsilon\varphi(u,x) = \varepsilon^{-1}\mathbf{Q}\varphi(u,x) + \mathbf{C}(x)\mathbf{P}\varphi(u,x) + \varepsilon\theta(x)\varphi(u,x), \tag{2.10}$$

gdzie

$$\mathbf{Q}\varphi(\cdot,x) = q(x)[\mathbf{P}-\mathbf{I}]\varphi(\cdot,x), x \in X, \mathbf{P}\varphi(\cdot,x) = \int_X P(x,dy)\varphi(\cdot,y).$$

Operatory resztowe $\theta^\varepsilon(x)$ i θ^ε_1 mają reprezentacje

$$\theta(x)\varphi(u,x) = q(x)\mathbf{C}^2(x)\mathbf{C}^\varepsilon_2(x)\mathbf{P}\varphi(u,\cdot),$$

$$\theta^\varepsilon_1(x)\varphi(u,x) = q(x)\mathbf{C}(x)\mathbf{C}^\varepsilon_1(x)\mathbf{P}\varphi(u,\cdot), \tag{2.11}$$

gdzie

$$\mathbf{C}^\varepsilon_2(x)\varphi(u) := \int_0^\infty \overline{G}^{(2)}_x(s)\mathbf{C}_{\varepsilon s}(x)\varphi(u)ds,$$

$$\mathbf{C}^\varepsilon_1(x) = \int_0^\infty \overline{G}_x(s)\mathbf{C}_{\varepsilon s}(x)ds,$$

$$\overline{G}^{(2)}_x(s) := \int_s^\infty \overline{G}_x(t)dt.$$

Dowód. Dowód opiera się na poniższej reprezentacji półgrup

$$\boldsymbol{C}_{\varepsilon s}(x) = \boldsymbol{I} + \varepsilon s\boldsymbol{C}(x) + \varepsilon^2\boldsymbol{C}^2(x)\boldsymbol{T}_s^{\varepsilon}(x) = \boldsymbol{I} + \boldsymbol{C}(x)\int\limits_0^{\varepsilon^2 s}\boldsymbol{C}_v(x)dv,$$

gdzie

$$\boldsymbol{T}_s^{\varepsilon}(x) = \int\limits_0^s (s-v)\boldsymbol{C}_{\varepsilon v}(x)dv.$$

Należy pamiętać, że do obliczeń używamy reprezentacji operatorów resztowych

$$\boldsymbol{C}_2^{\varepsilon}(x) = \int\limits_0^{\infty} G_x(ds)\boldsymbol{C}_s^{\varepsilon}(x) = \int\limits_0^{\infty}\overline{G}_x^{(2)}(s)\boldsymbol{C}_{\varepsilon s}(x)ds,$$

$$\boldsymbol{C}_1^{\varepsilon}(x) = \int\limits_0^{\infty}\overline{G}_x(s)\boldsymbol{C}_{\varepsilon s}(x)ds.$$

Zauważmy, że

$$\int\limits_0^{\infty}\overline{G}_x^{(2)}(s)ds = g_2(x)/2, g_2(x) := E\theta_x^2 = \int\limits_0^{\infty} t^2 G_x(dt).$$

□

Rozważmy zaburzoną funkcję Lapunowa

$$V^{\varepsilon}(u,x) = V(u) + \varepsilon V_1(u,x), \tag{2.12}$$

gdzie $V(u), u \in R^d$ jest funkcją Lapunowa dla układu uśrednionego (2.3), która spełnia warunki $C1 - C3$ twierdzenia.

Zaburzenie $V_1(u,x), u \in R^d, x \in X$, jest określone przez rozwiązanie osobliwe problemu związanego z zakłóceniami dla operatora

$$\boldsymbol{L}_0^{\varepsilon}\varphi(u,x) = \varepsilon^{-1}\mathbf{Q}\varphi(u,x) + \boldsymbol{C}(x)\varphi(u,x). \tag{2.13}$$

Według [21] rozwiązanie problemu osobliwych zaburzeń dla operatora (2.13) jest określone przez relacje

$$\boldsymbol{L}_0^{\varepsilon}V^{\varepsilon}(u,x) = \boldsymbol{C}V(u) + \varepsilon\boldsymbol{C}(x)V_1(u,x),$$

$$\boldsymbol{C}V(u) = C(u)V'(u),$$
$$V_1(u,x) = R_0\widetilde{\boldsymbol{C}}(x)V(u), \tag{2.14}$$

gdzie $\widetilde{\boldsymbol{C}}(x) := \boldsymbol{C}(x) - \boldsymbol{C}$ i operator $\boldsymbol{R}_0 = \Pi - (\mathbf{Q}+\Pi)^{-1}$ jest potencjałem operatora $\mathbf{Q}$. Wiadomo, że pod warunkiem $C4$ twierdzenia operator $\boldsymbol{R}_0$ jest ograniczony [28].

Lemat 2.3. Rozwiązanie problemu osobliwych zaburzeń dla operatora kompensującego (3.2) na zaburzonej funkcji Lapunowa (3.4) jest określone przez następujące równości

$$\mathbf{L}^{\varepsilon}V^{\varepsilon}(u,x) = \mathbf{C}V(u) + \varepsilon\theta_0^{\varepsilon}(x)V(u). \tag{2.15}$$

Operator resztowy ma postać

$$\theta_0^{\varepsilon}(x)V(u) = q(x)[\mathbf{C}^2(x)\mathbf{C}_2^{\varepsilon}(x) + \mathbf{C}(x)\mathbf{C}_1^{\varepsilon}(x)\mathbf{P}R_0\widetilde{\mathbf{C}}(x)]V(u). \tag{2.16}$$

Dowód. Użyjmy dwóch reprezentacji generatora $\boldsymbol{L}^{\varepsilon}$ z lematu 2.2, wprowadzając zapis (3.2)

$$\boldsymbol{L}_{(0)}^{\varepsilon} := \varepsilon^{-1}\mathbf{Q} + \theta_1^{\varepsilon}(x),$$

a także dla (2.9)

$$\boldsymbol{L}_{(1)}^{\varepsilon} := \varepsilon^{-1}\mathbf{Q} + \boldsymbol{C}(x)\boldsymbol{P} + \varepsilon\theta^{\varepsilon}(x).$$

Dla zaburzonej funkcji Lapunowa (3.4) generator $\boldsymbol{L}^{\varepsilon}$ jest prezentowany w formie

$$\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(u,x) = \varepsilon\boldsymbol{L}_{(0)}^{\varepsilon}V_1(u,x) + \boldsymbol{L}_{(1)}^{\varepsilon}V(u).$$

Ponieważ

$$\varepsilon\boldsymbol{L}_{(0)}^{\varepsilon}V_1(u,x) = \mathbf{Q}V_1(u,x) + \varepsilon\theta_1^{\varepsilon}(x)V_1(u,x),$$

i

$$\boldsymbol{L}_{(1)}^{\varepsilon}V(u) = \varepsilon^{-1}\mathbf{Q}V(u) + \boldsymbol{C}(x)V(u) + \varepsilon\theta^{\varepsilon}(x)V(u),$$

to

$$\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(u,x) = \varepsilon^{-1}\mathbf{Q}V(u) + \mathbf{Q}V_1(u,x) + \boldsymbol{C}(x)V(u) + \varepsilon[\theta^{\varepsilon}(x)V(u) + \theta_1^{\varepsilon}(x)V_1(u,x)].$$

Biorąc pod uwagę reprezentację operatorów resztowych $\theta^{\varepsilon}(x)$ i θ_1^{ε}, a także zaburzenie $V_1(u,x)$ w postaci (2.14), otrzymujemy postać (2.16) składnika resztowego. □

Wniosek 2.1. Przy spełnieniu warunków $C1$–$C5$ twierdzenia prawdziwa jest nierówność

$$\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(u,x) \leqslant -cV(u). \tag{2.17}$$

Aby udowodnić nierówność (2.17), najpierw obliczamy operator resztowy (2.16), stosując warunki $C2$–$C3$ twierdzenia

$$|\theta_0^\varepsilon(x)V(u)| \leqslant c_0 V(u), \tag{2.18}$$

i szacujemy również zaburzenie $V_1(u,x)$

$$|V_1(u,x)| \leqslant c_1 V(u). \tag{2.19}$$

Zgodnie z warunkiem wykładniczej stabilności $C1$, pierwszy wyraz w (2.15) spełnia nierówność

$$\boldsymbol{C}V(u) = C(u)V'(u) \leqslant -c_0 V(u), c_0 > 0. \tag{2.20}$$

Korzystając z (2.18), (2.19) i (2.20), otrzymujemy (2.17) dla wszystkich $\varepsilon \leqslant \varepsilon_0$, gdzie ε_0 jest wystarczająco małe, dla pewnej wartości $c > 0$.

Wniosek 2.2. Warunek $C2$ twierdzenia zapewnia oszacowanie półgrupy $\boldsymbol{C}_t(x)$, generowanej przez operator

$$|\boldsymbol{C}_t(x)V(u)| \leqslant e^{bt}V(u). \tag{2.21}$$

Mamy również nierówności (patrz (2.18), (2.19))

$$0 \leqslant b_1 V^\varepsilon(u,x) \leqslant V(u) \leqslant b_2 V^\varepsilon(u,x), 0 < b_1 < b_2.$$

Lemat 2.4. RPOM (2.6) charakteryzuje się martyngałem

$$\mu_{n+1}^\varepsilon = \varphi(u_{n+1}^\varepsilon, x_{n+1}^\varepsilon, \tau_{n+1}^\varepsilon) - \sum_{k=0}^{n} \theta_{k+1} \boldsymbol{L}^\varepsilon \varphi(u_k^\varepsilon, x_k^\varepsilon, \tau_k^\varepsilon), n \geqslant 0. \tag{2.22}$$

Dowód. Własności martyngałowe ciągu $(\mu_{n+1}^\varepsilon, n \geqslant 0)$ wynikają z jednorodności operatora kompensującego

$$E[\varphi(u_{n+1}^\varepsilon, x_{n+1}^\varepsilon, \tau_{n+1}^\varepsilon) - \varphi(u_n^\varepsilon, x_n^\varepsilon, \tau_n^\varepsilon)|F_n^\varepsilon] =$$

$$= E[\varphi(u_1^\varepsilon, x_1^\varepsilon, \tau_1^\varepsilon) - \varphi(u_0^\varepsilon, x_0^\varepsilon, \tau_0^\varepsilon)|F_0^\varepsilon].$$

Tutaj σ-algebra $F_n^\varepsilon := \sigma\{u_k^\varepsilon, x_k^\varepsilon, \tau_k^\varepsilon; 0 \leqslant k \leqslant n\}, n \geqslant 0$, jest wygenerowana przez RPOM. □

Aby skrócić zapis, zauważmy, że

$$\Phi^\varepsilon(t) := \varphi(u^\varepsilon(\tau^\varepsilon(t)), x^\varepsilon(\tau^\varepsilon(t)), \tau^\varepsilon(t)),$$

$$\Phi_+^\varepsilon(t) := \varphi(u^\varepsilon(\tau_+^\varepsilon(t)), x^\varepsilon(\tau_+^\varepsilon(t)), \tau_+^\varepsilon(t)),$$

gdzie $\tau^\varepsilon(t) := \tau^\varepsilon_{\nu^\varepsilon(t)}$, $\tau_+^\varepsilon(t) := \tau^\varepsilon_{\nu_+^\varepsilon(t)}$, $\nu_+^\varepsilon(t) := \nu^\varepsilon(t)+1$, $\nu^\varepsilon(t) := \nu(\varepsilon t)$.

Zauważmy, że $\nu_+^\varepsilon(t)$ jest momentem odnowy Markowa algebry F_n^ε.

W przyszłości skorzystamy z właściwości martyngałowej procesu z czasem ciągłym

$$\zeta^\varepsilon(t) = \Phi_+^\varepsilon(t) - \int_0^{\tau_+^\varepsilon(t)} \boldsymbol{L}^\varepsilon \Phi^\varepsilon(s)ds. \tag{2.23}$$

Zauważmy, że

$$\zeta^\varepsilon(\tau_n^\varepsilon) = \mu_{n+1}^\varepsilon, n \geqslant 0,$$

oraz

$$\zeta^\varepsilon(t) = \zeta^\varepsilon(\tau_+^\varepsilon(t)), \tau^\varepsilon(t) < t < \tau_+^\varepsilon(t).$$

Kluczowe znaczenie dla udowodnienia stabilności ma następujący lemat.

Lemat 2.5. Przy dowolnej ustalonej wartości rzeczywistej parametru $c \in R$, proces

$$\zeta_c^\varepsilon(t) = e^{c\tau_+^\varepsilon(t)}\Phi_+^\varepsilon(t) - \int_0^{\tau_+^\varepsilon(t)} [ce^{cs}\Phi_+^\varepsilon(s) + e^{c\tau^\varepsilon(s)}\mathbf{L}^\varepsilon\Phi^\varepsilon(s)]ds \tag{2.24}$$

ma właściwość martyngałową

$$E[\zeta_c^\varepsilon(t) - \zeta_c^\varepsilon(s)|F_s^\varepsilon] = 0, 0 \leqslant s < t,$$

względem filtracji σ-algebry

$$F_s^\varepsilon := \{\sigma(u^\varepsilon(l)), x^\varepsilon(l), \tau^\varepsilon(l); 0 \leqslant l \leqslant s\}.$$

Dowód. Proces (2.24) można przedstawić w następującej równoważnej formie

$$\zeta_c^\varepsilon(t) = e^{c\tau_+^\varepsilon(t)}\zeta^\varepsilon(t) - \int_0^{\tau_+^\varepsilon(t)} ce^{cs}\zeta^\varepsilon(s)ds, \tag{2.25}$$

gdzie proces $\zeta^\varepsilon(t), t \geqslant 0$, wyraża się wzorem (2.23).

Należy zauważyć, że właściwość martyngałowa procesu (2.25) wynika z tego, że proces jest skokowy i przyjmuje wartości

$$\zeta_c^\varepsilon(\tau_n^\varepsilon) = e^{c\tau_{n+1}^\varepsilon}\mu_{n+1}^\varepsilon - \sum_{k=0}^{n}[e^{c\tau_{k+1}^\varepsilon} - e^{c\tau_k^\varepsilon}]\mu_{k+1}^\varepsilon, \tag{2.26}$$

albo inaczej

$$\zeta_c^{\varepsilon}(\tau_n^{\varepsilon}) = \mu_0 + \sum_{k=0}^{n} e^{c\tau_k^{\varepsilon}}[\mu_{k+1}^{\varepsilon} - \mu_k^{\varepsilon}], n \geqslant 0. \tag{2.27}$$

Aby udowodnić równoważność reprezentacji (2.24) i (2.25), obliczmy całkę podwójną

$$I^{\varepsilon}(t) = \int_0^{\tau_+^{\varepsilon}(t)} ce^{cs}ds \int_0^{\tau_+^{\varepsilon}(s)} \boldsymbol{L}^{\varepsilon}\Phi^{\varepsilon}(v)dv = \sum_{k=0}^{\nu^{\varepsilon}(t)} \int_{\tau_k^{\varepsilon}}^{\tau_{k+1}^{\varepsilon}} ce^{cs}ds \int_0^{\tau_{k+1}^{\varepsilon}} \boldsymbol{L}^{\varepsilon}\Phi^{\varepsilon}(v)dv =$$

$$\sum_{k=0}^{\nu^{\varepsilon}(t)} (e^{c\tau_{k+1}^{\varepsilon}} - e^{c\tau_k^{\varepsilon}}) \sum_{r=0}^{k} \varepsilon\theta_{r+1}\boldsymbol{L}^{\varepsilon}\Phi^{\varepsilon}(\tau_r^{\varepsilon}) = \sum_{r=0}^{\nu^{\varepsilon}(t)} \theta_{r+1}\boldsymbol{L}^{\varepsilon}\Phi^{\varepsilon}(\tau_r^{\varepsilon})(e^{c\tau_{r+1}^{\varepsilon}} - e^{c\tau_r^{\varepsilon}}).$$

Wreszcie mamy

$$I^{\varepsilon}(t) = e^{c\tau_+^{\varepsilon}(t)} \int_0^{\tau_+^{\varepsilon}(t)} \boldsymbol{L}^{\varepsilon}\Phi^{\varepsilon}(s)ds - \int_0^{\tau_{k+1}^{\varepsilon}} e^{c\tau^{\varepsilon}(s)}\boldsymbol{L}^{\varepsilon}\Phi^{\varepsilon}(s)ds.$$

Stąd otrzymujemy reprezentację (2.24) z (2.23) i (2.25). □

Rozważmy teraz proces

$$\eta^{\varepsilon}(t) = \int_0^{\tau_+^{\varepsilon}(t)} [e^{cs}cV_+^{\varepsilon}(s) + e^{c\tau^{\varepsilon}(t)}\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(s)]ds. \tag{2.28}$$

Lemat 2.6. Dla wystarczająco małych wartości parametru c, dla wszystkich $\varepsilon \leqslant \varepsilon_0$, gdzie ε_0 jest wystarczająco małe, zgodnie z określonymi warunkami twierdzenia, proces $\eta^{\varepsilon}(t), t \geqslant 0$, spełnia warunek

$$E[\eta^{\varepsilon}(t) - \eta^{\varepsilon}(s)|F_s^{\varepsilon}] \leqslant 0, 0 \leqslant s < t. \tag{2.29}$$

Dowód. Ponieważ proces (2.28) jest skokowy, wystarczy rozważyć warunkową wartość oczekiwaną w momentach odnowy Markowa

$$\eta_{n+1}^{\varepsilon} := \eta^{\varepsilon}(\tau_n^{\varepsilon}) = \int_0^{\tau_{n+1}^{\varepsilon}} [e^{cs}cV_+^{\varepsilon}(s) + e^{c\tau^{\varepsilon}(s)}\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(s)]ds, \tag{2.30}$$

albo inaczej

$$\eta_{n+1}^{\varepsilon} = \sum_{k=0}^{n} e^{c\tau_k^{\varepsilon}} \alpha_{k+1}^{\varepsilon}, \tag{2.31}$$

gdzie

$$\alpha_{k+1}^{\varepsilon} = \beta_{k+1}^{\varepsilon} + \varepsilon\gamma_{k+1}^{\varepsilon}, \tag{2.32}$$

$$\beta_{k+1}^{\varepsilon} = (e^{\varepsilon c\theta_{k+1}} - 1)V_{k+1}^{\varepsilon}, \gamma_{k+1}^{\varepsilon} = \varepsilon\theta_{k+1}\boldsymbol{L}^{\varepsilon}V_k^{\varepsilon}, V_n^{\varepsilon} := V^{\varepsilon}(u_n^{\varepsilon}, x_n^{\varepsilon}, \tau_n^{\varepsilon}).$$

Aby oszacować warunkową wartość oczekiwaną pierwszego wyrazu w (2.32), zastosujemy estymację półgrupy (patrz (2.21)) $\boldsymbol{C}_t(x)$ wygenerowanej przez generator (3.1)

$$|\boldsymbol{C}_t(x)V(u)| \leqslant e^{bt}V(u), b > 0.$$

Stąd mamy

$$\begin{aligned} E[\beta_{k+1}^{\varepsilon}|F_k^{\varepsilon}] &= E[(e^{\varepsilon c\theta_{k+1}} - 1)V_{k+1}^{\varepsilon}|F_k^{\varepsilon}] = \\ &= E[(e^{\varepsilon c\theta_{k+1}} - 1)\boldsymbol{C}_{\varepsilon}\theta_{k+1}(x_k)V_k^{\varepsilon}|F_k^{\varepsilon}] \leqslant \\ &\leqslant E[(e^{\varepsilon c\theta_{k+1}} - 1)e^{\varepsilon c\theta_{k+1}}|F_k^{\varepsilon}]V_k^{\varepsilon} = \\ &= E[(e^{\varepsilon(c+b)\theta_{k+1}} - e^{\varepsilon b\theta_{k+1}})|F_k^{\varepsilon}]V_k^{\varepsilon} = \\ &= \varepsilon[b(g_k^{\varepsilon}(b+c) - g_k^{\varepsilon}(b)) + cg_k^{\varepsilon}(b+c)]V_k^{\varepsilon}. \end{aligned}$$

Biorąc pod uwagę

$$g_k^{\varepsilon}(c) = \int_0^{\infty} \overline{G}_k(t)e^{\varepsilon ct}dt$$

i z warunku $C4$ twierdzenia, mamy oszacowanie

$$m_k \leqslant g_k^{\varepsilon}(c) \leqslant m_k(1+\delta_{\varepsilon}),$$

gdzie $\delta_{\varepsilon} \leqslant \delta_0$, δ_0 jest wystarczająco małe przy wystarczająco małym $\varepsilon \leqslant \varepsilon_0$.

Mamy więc wreszcie oszacowanie

$$E[\beta_{k+1}^{\varepsilon}|F_k^{\varepsilon}] \leqslant \varepsilon[b\delta_{\varepsilon} + cm_k(1+\delta_{\varepsilon})]V_k^{\varepsilon} < \varepsilon\delta_0 V_k^{\varepsilon},$$

gdzie $\varepsilon \leqslant \varepsilon_0$ dla wystarczająco małych ε_0 i δ_0.

Warunkowa wartość oczekiwana drugiego członu w (2.32) jest oszacowana przy wykorzystaniu nierówności (2.20) i warunków

$$E[\gamma_{k+1}^{\varepsilon}|F_k^{\varepsilon}] = m_k\boldsymbol{L}^{\varepsilon}V_k^{\varepsilon} \leqslant -c_0 m_k V_k^{\varepsilon} \leqslant -c_0\underline{m}V_k^{\varepsilon}.$$

W końcu mamy oszacowanie

$$E[\alpha_{k+1}^{\varepsilon}|F_k^{\varepsilon}] \leqslant -\varepsilon[c_0\underline{m} - \delta_0]V_k^{\varepsilon} \leqslant -\varepsilon c_0 V_k^{\varepsilon}, c_0 > 0,$$

gdzie $\delta_0 < c_0\underline{m}$. □

Dowód twierdzenia 2.1.

Dowód. Przede wszystkim z lematu 2.6 otrzymujemy

Wniosek 2.3. Ciąg

$$w_{n+1}^{\varepsilon} = e^{c\tau_{n+1}^{\varepsilon}} V_{n+1}^{\varepsilon}, V_{n+1}^{\varepsilon} := V^{\varepsilon}(u_{n+1}^{\varepsilon}, x_{n+1}^{\varepsilon}, \tau_{n+1}^{\varepsilon}),$$

jest supermartyngałem

$$E[w_{n+1}^{\varepsilon}|F_n^{\varepsilon}] \leqslant w_n^{\varepsilon}, n \geqslant 0. \tag{2.33}$$

Rzeczywiście, zgodnie z lematem 2.5, mamy (patrz (2.24) i (2.30)–(2.31))

$$w_{n+1}^{\varepsilon} = \zeta_c^{\varepsilon}(\tau_n^{\varepsilon}) + \zeta_{n+1}^{\varepsilon}.$$

Własność martyngałowa ciągu (2.27) i warunek (2.29) daje warunek (2.33)

$$E[w_{n+1}^{\varepsilon}|F_n^{\varepsilon}] = E[\zeta_c^{\varepsilon}(\tau_n^{\varepsilon}) + \zeta_{n+1}^{\varepsilon}|F_n^{\varepsilon}] =$$

$$= w_n^{\varepsilon} + E[\zeta_c^{\varepsilon}(\tau_n^{\varepsilon}) - \zeta_c^{\varepsilon}(\tau_{n+1}^{\varepsilon})|F_n^{\varepsilon}] + E[\zeta_{n+1}^{\varepsilon} - \zeta_n^{\varepsilon}|F_n^{\varepsilon}] \leqslant w_n^{\varepsilon}.$$

Regularność procesu odnowy Markowa oznacza, że

$$\tau_n^{\varepsilon} \to \infty,\ n \to \infty,$$

z prawdopodobieństwem 1. A więc prawdopodobieństwo zdarzenia losowego $A_{T,n}^{\varepsilon} = \{\tau_n^{\varepsilon} > T\}$ można oszacować jako

$$P\{A_{T,n}^{\varepsilon}\} \geqslant 1 - \Delta, n \geqslant N_T,$$

dla dowolnego $\Delta > 0$.

Wprowadźmy oznaczenie

$$a_n^{\varepsilon} := e^{c(\tau_n^{\varepsilon} - T)}, n \geqslant 0.$$

W zdarzeniu $A_{T,n}^{\varepsilon}$ mamy $a_n^{\varepsilon} \geqslant 1, n \geqslant 0$.

Teraz szacujemy prawdopodobieństwo

$$P\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta \bigcap A_{T,n}^\varepsilon\} \leqslant P\{e^{cT}\sup_{n\geqslant N_T} a_n^\varepsilon V(u_n^\varepsilon) > \delta \bigcap A_{T,n}^\varepsilon\} \leqslant$$

$$\leqslant P\{\sup_{n\geqslant N_T} e^{c\tau_n^\varepsilon} V(u_n^\varepsilon) > \delta \bigcap A_{T,n}^\varepsilon\}.$$

Biorąc pod uwagę (2.4), mamy

$$P\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta \bigcap A_{T,n}^\varepsilon\} \leqslant P\{\sup_{n\geqslant N_T} w_n^\varepsilon > \delta \bigcap A_{T,n}^\varepsilon\}.$$

Z własności supermartyngału (2.33) można uzyskać następujące oszacowanie dla prawdopodobieństwa

$$P\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta \bigcap A_{T,n}^\varepsilon\} \leqslant V(u)/\delta_2,$$

gdzie $\delta_2 = \delta b_2/b_1$.

Teraz mamy

$$P\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta\} = P\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta \bigcap A_{T,n}^\varepsilon\} +$$

$$+P\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta \bigcap \overline{A}_{T,n}^\varepsilon\} \leqslant V(u)/\delta_2 + \Delta \leqslant 2\Delta, \tag{2.34}$$

gdy $|u| \leqslant u_0$ jest wystarczająco małe.

Relacja

$$\{e^{cT}\sup_{n\geqslant N_T} V(u_n^\varepsilon) > \delta\} \subset \{\lim_{n\to\infty} V(u_n^\varepsilon) = 0\} = \{\lim_{n\to\infty} \|u_n^\varepsilon\| = 0\}$$

kończy dowód zbieżności

$$P\{\lim_{n\to\infty} \|u_n^\varepsilon\| = 0\} = 1. \tag{2.35}$$

Stwierdzenie (2.5) jest konsekwencją (2.34) i oszacowania (2.20) dla półgrupy generującej układ stochastyczny $u^\varepsilon(t)$. Rzeczywiście, z warunku $C5$ wynika następujące oszacowanie

$$P\{\max_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} \theta_n > \delta\} \leqslant C\Delta(\varepsilon), \tag{2.36}$$

gdzie $\Delta(\varepsilon) \to 0$, $\varepsilon \to 0$.

Rozważmy reprezentację

$$V_T^\varepsilon := \sup_{0\leqslant t\leqslant T} V(u^\varepsilon(t)) = \sup_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} [\sup_{0\leqslant s\leqslant \theta_n} |\boldsymbol{C}_{\varepsilon s}(x_{n-1}^\varepsilon)V_{n-1}^\varepsilon|].$$

Biorąc pod uwagę oszacowanie (2.21) dla półgrupy, mamy nierówność

$$V_T^\varepsilon \leqslant \sup_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} e^{\varepsilon c_1\theta_n} V_{n-1}^\varepsilon =: \overline{V}_T^\varepsilon. \tag{2.37}$$

Prawdziwe jest więc oszacowanie

$$P\{\sup_{0\leqslant t\leqslant T} V(u^\varepsilon(t)) > \delta\} \leqslant P\{\overline{V}_T^\varepsilon > \delta\}. \tag{2.38}$$

Aby oszacować prawą stronę (2.38), używamy równości

$$P\{\overline{V}_T^\varepsilon > \delta\} = P\{\overline{V}_T^\varepsilon > \delta \bigcap\nolimits_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} \max \theta_n > T/\varepsilon\} +$$

$$+P\{\overline{V}_T^\varepsilon > \delta \bigcap\nolimits_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} \max \theta_n \leqslant T/\varepsilon\}. \tag{2.39}$$

Pierwszy człon (2.39) ma oszacowanie z (2.37) i (2.34),

$$P\{\overline{V}_T^\varepsilon > \delta \bigcap\nolimits_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} \max \theta_n \leqslant T/\varepsilon\} \leqslant$$

$$\leqslant P\{e^{c_1 T} \sup_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} V_{n-1}^\varepsilon > \delta\} \leqslant 2\Delta.$$

Szacując drugi człon, bierzemy pod uwagę (2.36)

$$P\{\overline{V}_T^\varepsilon > \delta \bigcap\nolimits_{1\leqslant n\leqslant \nu^\varepsilon(T/\varepsilon)} \max \theta_n > T/\varepsilon\} \leqslant C\Delta(\varepsilon).$$

Z dwóch ostatnich oszacowań mamy (patrz (2.38) i (2.39))

$$P\{\sup_{1\leqslant t\leqslant T} V(u^\varepsilon(t)) > \delta\} \leqslant 2\Delta + C\Delta(\varepsilon),$$

gdzie ε jest wystarczająco małe.
Następnie twierdzenie modelowe 1.1 [25] prowadzi do tezy twierdzenia 2.1. □

2.2. Stabilność układów stochastycznych w schemacie aproksymacji dyfuzyjnej

Analiza stabilności układu dynamicznego w schemacie aproksymacji dyfuzyjnej z przełączeniami półmarkowowskimi [11] rozpatrywana jest w bardziej ogólnej formie i wdrażana za pomocą operatora kompensującego procesu półmarkowowskiego. Układ dynamiczny w ośrodku półmarkowowskim w warunkach aproksymacji dyfuzyjnej wynika z ewolucyjnego równania

$$\frac{du^{\varepsilon}(t)}{dt} = \varepsilon^{-1} C_1(u^{\varepsilon}(t), x(\frac{t}{\varepsilon^2})) + C_0(u^{\varepsilon}(t), x(\frac{t}{\varepsilon^2})), \tag{2.40}$$

$$u^{\varepsilon}(0) = u_0,$$

gdzie $\varepsilon > 0$ jest małym parametrem, a $u^{\varepsilon}(t) = (u_k^{\varepsilon}(t), k = \overline{1,d})$. Prędkości $C_k(u,x)$ = ($C_{ki}(u,x)$; $i = \overline{1,d}$), $k = 1,0, u \in R^d, x \in X$, spełniają warunki, które zapewniają istnienie rozwiązań globalnych układów deterministycznych dla każdego $\varepsilon > 0$

$$\frac{du_x^{\varepsilon}(t)}{dt} = \varepsilon^{-1} C_1(u_x^{\varepsilon}(t), x) + C_0(u_x^{\varepsilon}(t), x), x \in X. \tag{2.41}$$

Tutaj $x(t), t > 0$, jest procesem półmarkowowskim (PSM) w standardowej przestrzeni fazowej stanów $(X, \boldsymbol{X})$, generowanym przez proces odnowy Markowa $x_n, \tau_n, n \geqslant 0$.

Rozważana jest stabilność układu stochastycznego [11] w warunkach stabilności uśrednionego układu, które w stanie równowagi

$$\int_X \pi(dx) C_1(u,x) \equiv 0, u \in R^d,$$

wyznacza się poprzez rozwiązanie układu stochastycznego (2.40)

$$\begin{cases} du(t) = C_0(u(t))dt + d\zeta(t) \\ d\zeta(t) = a(u(t))dt + \sigma(u(t))dw(t). \end{cases} \tag{2.42}$$

Uśrednianie odbywa się według rozkładu stacjonarnego $\pi(dx)$

$$C_0(u) := \int_X \pi(dx) C_0(u,x).$$

Zdefiniowano proces dyfuzji $\zeta(t), t \geqslant 0$, za pomocą funkcji wektora przesunięcia

$$a(u) = a_1(u) + a_2(u),$$

gdzie

$$a_1(u) = \int\limits_X \pi(dx) C_1(u,x) R_0 C_1'(u,x), a_2(u) = \frac{1}{2} q \int\limits_X \rho(dx) \mu(x) C_1(u,x) C_1'(u,x), \tag{2.43}$$

oraz macierzy wariancji $\sigma(u)$, która jest określona jako

$$B(u) = \sigma(u)\sigma^*(u),$$

przy założeniu dodatniej określoności macierzy

$$B(u) = B_0(u) + B_1(u), \tag{2.44}$$

gdzie

$$B_0(u) = 2\int_X \pi(dx) C_1(u,x) R_0 C_1(u,x),$$

$$B_1(u) = q\int_X \rho(dx)\mu(x) C_1^2(u,x). \tag{2.45}$$

W (2.43) i (2.45)

$$\mu(x) = g_2(x) - 2g^2(x), g_2(x) = \int\limits_0^\infty \overline{G}_x^{(2)}(s)ds,$$

oraz $\bar{G}_x^{(2)}(t) := \int\limits_t^\infty \bar{G}_x(s)ds$.

Wniosek 2.4. Dla funkcji rozkładu wykładniczego o natężeniu $q(x) = g^{-1}(x)$ mamy $\mu(x) = 0$. Dlatego składniki $a_2(u)$ przesunięcia i wariancji $B_1(u)$ charakteryzują niemarkowowowskość przełączenia procesu.

Wniosek 2.5. Generator układu stochastycznego (2.42) jest zdefiniowany przez funkcje testowe $\varphi(u) \in C^2(R^d)$ w postaci

$$\boldsymbol{L}\varphi(u) = C(u)\varphi'(u) + \frac{1}{2} Sp[B(u)\varphi''(u)], \tag{2.46}$$

gdzie

$$C(u) = C_0(u) + a(u). \tag{2.47}$$

Zadanie polega na tym, że w warunkach zbieżności układu stochastycznego (2.40) do układu uśrednionego (2.42) przy $\varepsilon \to 0$ należy określić dodatkowe warunki zapewniające stabilność pierwotnego układu (2.40) dla wszystkich $\varepsilon \leqslant \varepsilon_0$

i wystarczająco małego ε_0. Stabilność układu (2.40) jest rozważana w warunkach wykładniczej stabilności układu uśrednionego (2.42)

$$\frac{d\widetilde{u}(t)}{dt} = C_0(\widetilde{u}(t)).$$

Twierdzenie 2.2. *Niech dla uśrednionego stochastycznego układu* (2.42)*, wyznaczonego przez generator* (2.46)*, istnieje funkcja Lapunowa* $V(u)$, $u \in R^d$ *, dla której spełniony jest warunek stabilności wykładniczej*

$C1$: $C_0(u)V'(u) \leqslant -c_0V(u)$, $c_0 > 0$,

oraz następujące dodatkowe warunki dla $k,r,l = 0,1$

$C2$: $| C_k(u,x)V'(u) | \leqslant c_1V(u)$, $c_1 > 0$,

$| C_k(u,x)R_0[C_r(u,x)V'(u)]' | \leqslant c_2V(u)$, $c_2 > 0$,

$| C_k(u,x)R_0[C_r(u,x)R_0[C_l(u,x)V'(u)]']' | \leqslant c_3V(u)$, $c_3 > 0$,

$C3$: *funkcje rozkładu* $G_x(t)$, $t \geqslant 0$, $x \in X$, *spełniają warunek Cramera jednostajnie według* $x \in X$: $\sup\limits_{x\in X} \int\limits_0^\infty e^{ht}\overline{G}_x(t)dt \leqslant H < +\infty$, $h > 0$, *i zachodzą oszacowania*

$$0 < \underline{m} \leqslant g(x) \leqslant \overline{m} \leqslant +\infty.$$

Wtedy dla wszystkich $\varepsilon \leqslant \varepsilon_0$*, gdzie* ε_0 *wystarczająco mały, rozwiązanie ewolucyjnego równania* (2.40)*, dla wszystkich warunków początkowych* $| u^\varepsilon(0) | \leqslant u^*$*, gdzie* u^* *jest dostatecznie małe, jest asymptotycznie stabilny z prawdopodobieństwem 1*

$$P\{\lim_{t\to\infty} \| u^\varepsilon(t) \| = 0\} = 1.$$

Niech zbiór półgrup $\boldsymbol{C}_t^\varepsilon(x)$, $t \geqslant 0, x \in X$, będzie generowany przez układ (2.41) i określony generatorem

$$\boldsymbol{C}^\varepsilon(x)\varphi(u) = C^\varepsilon(u,x)\varphi'(u), \tag{2.48}$$

gdzie

$$C^\varepsilon(u,x) = \varepsilon^{-1}C_1(u,x) + C_0(u,x).$$

Rozważmy także operator $\boldsymbol{C}_\varepsilon(x)$, który ma postać

$$\boldsymbol{C}_\varepsilon(x) := \varepsilon\boldsymbol{C}^\varepsilon(x) = \boldsymbol{C}_1(x) + \varepsilon\boldsymbol{C}_0(x), \tag{2.49}$$

i którego składniki są określone wzorami

$$\boldsymbol{C}_1(x)\varphi(u) := C_1(u,x)\varphi'(u), \quad \boldsymbol{C}_0(x)\varphi(u) = C_0(u,x)\varphi'(u).$$

Lemat 2.7. Operator kompensujący (2.9) dla RPOM (2.6) na funkcjach testowych $\varphi(u,x)$ wygląda następująco

$$\mathbf{L}^{\varepsilon}\varphi(u,x) = \varepsilon^{-2}q(x)[\int_0^{\infty} G_x(ds)\boldsymbol{C}^{\varepsilon}_{\varepsilon^2 s}(x)\int_X P(x,dy)\varphi(u,y) - \varphi(u,x)]. \qquad (2.50)$$

Dowód. Z tego, że

$$E\varphi(u_1^{\varepsilon},x_1^{\varepsilon}) = E\boldsymbol{C}^{\varepsilon}_{\theta_x}(x)\varphi(u,x_1^{\varepsilon}) = \int_0^{\infty} G_x(ds)\boldsymbol{C}^{\varepsilon}_{\varepsilon^2 s}(x)\int_X P(x,dy)\varphi(u,y)$$

i (2.9), mamy (2.50). □

Lemat 2.8. Operator kompensujący (2.50) na funkcjach testowych $\varphi(u,\cdot)\in C^3(R^d)$, posiada reprezentacje asymptotyczne

$$\boldsymbol{L}^{\varepsilon}\varphi(u,x) = \varepsilon^{-2}\mathbf{Q}\varphi(u,x) + \varepsilon^{-1}\theta_0^{\varepsilon}(x)\varphi(u,x) =$$

$$= \varepsilon^{-2}\mathbf{Q}\varphi(u,x) + \varepsilon^{-1}\mathbf{Q}_1(x)\varphi(u,x) + \theta_1^{\varepsilon}(x)\varphi(u,x) =$$

$$= \varepsilon^{-2}\mathbf{Q}\varphi(u,x) + \varepsilon^{-1}\mathbf{Q}_1(x)\varphi(u,x) + \mathbf{Q}_2(x)\varphi(u,x) + \varepsilon\theta_2^{\varepsilon}(x)\varphi(u,x), \qquad (2.51)$$

gdzie operator $\mathbf{Q}$ jest zdefiniowany w (1.1), operatory $\mathbf{Q}_1(x)$ i $\mathbf{Q}_2(x)$ wyznaczane są zależnościami

$$\mathbf{Q}_1(x)\varphi(u,x) = \mathbf{C}_1(x)\mathbf{P}\varphi(u,x),$$

$$\mathbf{Q}_2(x)\varphi(u,x) = [\,\mathbf{C}_0(x) + \mu_2(x)\mathbf{C}_1^2(x)]\mathbf{P}\varphi(u,x),$$

$$\mu_2(x) = \frac{g_2(x)}{2g(x)},$$

a pozostałe składniki reprezentacji (2.51) mają postaci

$$\theta_0^{\varepsilon}(x)\varphi(u,x) = q(x)\mathbf{C}_{\varepsilon}(x)\mathbf{G}_1^{\varepsilon}(x)\ \mathbf{P}\varphi(u,x), \qquad (2.52)$$

$$\theta_1^{\varepsilon}(x)\varphi(u,x) = q(x)[\mathbf{C}_{\varepsilon}^2(x)\mathbf{G}_2^{\varepsilon}(x) + \mathbf{C}_0(x)]\ \mathbf{P}\varphi(u,x), \qquad (2.53)$$

$$\theta_2^{\varepsilon}(x)\varphi(u,x) = q(x)[\mathbf{C}_{\varepsilon}^3(x)\ \mathbf{G}_3^{\varepsilon}(x) + \frac{m_2(x)}{2}\mathbf{C}_2^{\varepsilon}(x)]\mathbf{P}\varphi(u,x). \qquad (2.54)$$

Tutaj operatory $\mathbf{G}_k^{\varepsilon}(x)$, $k=\overline{1,3}$ są wyznaczane przez rekurencję

$$\mathbf{G}_k^{\varepsilon}(x) = \int_0^{\infty}\overline{G}_x^{(k)}(s)ds\mathbf{C}^{\varepsilon}_{\varepsilon^2 s}(x),$$

gdzie $\overline{G}_x^{(1)}(s) = \overline{G}_x(s)$, i $\mathbf{C}_2^{\varepsilon}(x) = \mathbf{C}_0(x)[2\mathbf{C}_1(x) + \varepsilon\mathbf{C}_0(x)]$.

Dowód. Najpierw skorzystajmy z oczywistej reprezentacji

$$\boldsymbol{L}^{\varepsilon} = \varepsilon^{-2}\mathbf{Q} + \varepsilon^{-2}\boldsymbol{L}_1^{\varepsilon}(x), \tag{2.55}$$

gdzie

$$\boldsymbol{L}_1^{\varepsilon}(x) := q(x)[\boldsymbol{G}_{\varepsilon}(x) - I]\boldsymbol{P}, \tag{2.56}$$

i po prawej stronie (2.56)

$$\boldsymbol{G}_{\varepsilon}(x) := \int_0^{\infty} G_x(ds)\boldsymbol{C}_{\varepsilon^2 s}^{\varepsilon}(x). \tag{2.57}$$

Następnie, całkując przez części i korzystając z równania dla półgrupy

$$d\boldsymbol{C}_{\varepsilon^2 s}^{\varepsilon}(x) = \varepsilon^2 \boldsymbol{C}^{\varepsilon}(x)\boldsymbol{C}_{\varepsilon^2 s}^{\varepsilon}(x)ds,$$

oraz biorąc pod uwagę warunek $C3$, otrzymujemy reprezentację

$$\boldsymbol{G}_{\varepsilon}(x) - I = \varepsilon^2\, \boldsymbol{C}^{\varepsilon}(x)\boldsymbol{G}_1^{\varepsilon}(x), \tag{2.58}$$

gdzie

$$\boldsymbol{G}_1^{\varepsilon}(x) = \int_0^{\infty} \overline{G}_x(s)ds C_{\varepsilon^2 s}^{\varepsilon}(x). \tag{2.59}$$

Podobnie z (2.59) mamy

$$\boldsymbol{G}_1^{\varepsilon}(x) = g(x)I + \varepsilon^2 \boldsymbol{C}^{\varepsilon}(x)\boldsymbol{G}_2^{\varepsilon}(x), \tag{2.60}$$

gdzie

$$\boldsymbol{G}_2^{\varepsilon}(x) := \int_0^{\infty} \overline{G}_x^{(2)}(s)ds C_{\varepsilon^2 s}^{\varepsilon}(x), \tag{2.61}$$

a także z (2.61)

$$\boldsymbol{G}_2^{\varepsilon}(x) = \frac{g_2(x)}{2} I + \varepsilon^2 \boldsymbol{C}^{\varepsilon}(x)\boldsymbol{G}_3^{\varepsilon}(x), \tag{2.62}$$

gdzie

$$\boldsymbol{G}_3^{\varepsilon}(x) := \int_0^{\infty} \overline{G}_x^{(3)}(s)ds C_{\varepsilon^2 s}^{\varepsilon}(x)$$

i

$$\overline{G}_x^{(3)}(s) := \int_s^{\infty} \overline{G}_x^{(2)}(s)ds.$$

Łącząc (2.59), (2.61) i (2.62), mamy

$$\boldsymbol{G}_\varepsilon(x) - I = \varepsilon^2 g(x)\boldsymbol{C}^\varepsilon(x) + \varepsilon^4 g_2(x)[\boldsymbol{C}^\varepsilon(x)]^2 + \varepsilon^6[\boldsymbol{C}^\varepsilon(x)]^3\boldsymbol{G}_3^\varepsilon(x). \tag{2.63}$$

Teraz, biorąc pod uwagę (2.48) i (2.49), w końcu z (2.63), otrzymujemy

$$\boldsymbol{G}_\varepsilon(x) - I = \varepsilon g(x)\boldsymbol{C}_1(x) + \varepsilon^2[g(x)\boldsymbol{C}_0(x) + g_2(x)\boldsymbol{C}_1^2(x)] + \varepsilon^3\theta_3^\varepsilon(x), \tag{2.64}$$

gdzie składnik resztowy ma reprezentację

$$\theta_3^\varepsilon(x) = g_2(x)(2\boldsymbol{C}_1(x)\boldsymbol{C}_0(x) + \varepsilon\boldsymbol{C}_0^2(x)) + \boldsymbol{C}_\varepsilon^3(x)\boldsymbol{G}_3^\varepsilon(x). \tag{2.65}$$

Łącząc (2.56), (2.57), (2.64) i (2.65), otrzymujemy wszystkie trzy asymptotyczne reprezentacje (2.51). □

Dowód twierdzenia 2.2 opiera się na zastosowaniu rozwiązania problemu osobliwych zaburzeń (patrz 1.2.1) dla operatora kompensującego podanego w ostatniej reprezentacji (2.51) lematu 2.8.

Przedstawmy zaburzoną funkcję Lapunowa jako

$$V^\varepsilon(u,x) = V(u) + \varepsilon V_1(u,x) + \varepsilon^2 V_2(u,x), \tag{2.66}$$

gdzie $V(u)$ jest funkcją Lapunowa dla dyfuzji granicznej (2.42).

Lemat 2.9. Na zaburzonej funkcji Lapunowa (2.66) operatorem kompensującym jest (2.50), co umożliwia następującą reprezentację

$$\mathbf{L}^\varepsilon V^\varepsilon(u,x) = \mathbf{L}V(u) + \varepsilon\theta_L^\varepsilon(x)V(u).$$

Tutaj $\mathbf{L}$ jest generatorem dyfuzji granicznej (2.42), a operator resztowy $\theta_L^\varepsilon(x)$ jest określony równością

$$\theta_L^\varepsilon(x)V(u) = \theta_2^\varepsilon(x)V(u) + \theta_1^\varepsilon(x)\mathbf{P}V_1(u,x) + \theta_0^\varepsilon(x)\mathbf{P}V_2(u,x). \tag{2.67}$$

Zaburzenia funkcji Lapunowa mają postać

$$V_1(u,x) = \mathbf{R}_0 Q_1(x)V(u), \tag{2.68}$$

$$V_2(u,x) = \mathbf{R}_0\tilde{\mathbf{L}}(x)V(u), \tag{2.69}$$

gdzie

$$\widetilde{\boldsymbol{L}}(x) := \boldsymbol{L}(x) - \boldsymbol{L} \tag{2.70}$$

i

$$\boldsymbol{L}(x) := \mathbf{Q}_2(x) + \mathbf{Q}_1(x)\mathbf{P}\mathbf{R}_0\mathbf{Q}_1(x).$$

Dowód. Rozważmy operator obcięty do (2.51), a mianowicie

$$\boldsymbol{L}_0^{\varepsilon} = \varepsilon^{-2}\mathbf{Q} + \varepsilon^{-1}\mathbf{Q}_1(x) + \mathbf{Q}_2(x).$$

Dla operatora $\boldsymbol{L}_0^{\varepsilon}$ na funkcjach (2.66) dostajemy rozkład

$$\boldsymbol{L}_0^{\varepsilon}V^{\varepsilon}(u,x) =$$

$$= \varepsilon^{-2}\mathbf{Q}V(u) + \varepsilon^{-1}[\mathbf{Q}V_1(u,x) + \mathbf{Q}_1(x)V(u)] +$$

$$+[\mathbf{Q}V_2(u,x) + \mathbf{Q}_1(x)V_1(u,x) + \mathbf{Q}_2(x)V(u)] + \varepsilon\theta_{L0}^{\varepsilon}(x)V(u).$$

Rozwiązanie problemu zaburzeń osobliwych dla takiego rozkładu daje reprezentację (2.68), (2.69) zaburzeń funkcji Lapunowa (2.66), a także operator graniczny $\boldsymbol{L}$, który jest obliczany według wzoru

$$\boldsymbol{L}\Pi = \Pi\mathbf{Q}_2(x)\Pi + \Pi\mathbf{Q}_1(x)\mathbf{R}_0\mathbf{Q}_1(x)\Pi.$$

Oznaczmy każdą z reprezentacji (2.51) przez

$$\boldsymbol{L}_{(0)}^{\varepsilon} = \varepsilon^{-2}\mathbf{Q} + \varepsilon^{-1}\theta_0^{\varepsilon}(x),$$

$$\boldsymbol{L}_{(1)}^{\varepsilon} = \varepsilon^{-2}\mathbf{Q} + \varepsilon^{-1}\mathbf{Q}_1(x) + \theta_1^{\varepsilon}(x),$$

$$\boldsymbol{L}_{(2)}^{\varepsilon} = \varepsilon^{-2}\mathbf{Q} + \varepsilon^{-1}\mathbf{Q}_1(x) + \mathbf{Q}_2(x) + \varepsilon\theta_2^{\varepsilon}(x).$$

Wtedy $\boldsymbol{L}^{\varepsilon}$ na zaburzonej funkcji Lapunowa $V^{\varepsilon}(u,x)$ ma rozwinięcie

$$\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(u,x) = \varepsilon^2\boldsymbol{L}_{(0)}^{\varepsilon}V_2(u,x) + \varepsilon\boldsymbol{L}_{(1)}^{\varepsilon}V_1(u,x) + \boldsymbol{L}_{(2)}^{\varepsilon}V(u).$$

Ponieważ

$$\varepsilon^2\boldsymbol{L}_{(0)}^{\varepsilon}V_2(u,x) = \mathbf{Q}V_2(u,x) + \varepsilon\theta_0^{\varepsilon}(x)V_2(u,x),$$

$$\varepsilon\boldsymbol{L}_{(1)}^{\varepsilon}V_1(u,x) = \varepsilon^{-1}\mathbf{Q}V_1(u,x) + \mathbf{Q}_1(x)V_1(u,x) + \varepsilon\theta_1^{\varepsilon}(x)V_1(u,x),$$

$$\boldsymbol{L}_{(2)}^{\varepsilon}V(u) = \varepsilon^{-2}\mathbf{Q}V(u) + \varepsilon^{-1}\mathbf{Q}_1(x)V(u) + \mathbf{Q}_2(x)V(u) + \varepsilon\theta_2^{\varepsilon}(x)V(u),$$

to

$$\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(u,x) =$$

$$= \varepsilon^{-2}\mathbf{Q}V(u) + \varepsilon^{-1}[\mathbf{Q}V_1(u,x) + \mathbf{Q}_1(x)V(u)] +$$
$$+ [\mathbf{Q}V_2(u,x) + \mathbf{Q}_1(x)V_1(u,x) + \mathbf{Q}_2(x)V(u)] + \varepsilon\theta_L^{\varepsilon}(x)V(u),$$

gdzie operator resztowy $\theta_L^{\varepsilon}(x)$ ma reprezentację (2.67). □

Wniosek 2.6. Operator graniczny określający średnią dyfuzji (2.42) wyraża się równością (2.46)

$$\boldsymbol{L}\varphi(u) = C(u)\varphi'(u) + \frac{1}{2}Sp[B(u)\varphi''(u)],$$

gdzie $C(u)$ i $B(u)$ są obliczane odpowiednio za pomocą wzorów (2.47) i (2.44).

Biorąc pod uwagę reprezentacje (2.68), (2.69) funkcji zaburzających $V_k(u,x)$, $k = 1,2$ i wyrażenie (2.67) dla operatora resztowego $\theta_L^{\varepsilon}(x)$, mamy następującą reprezentację operatora resztowego na funkcjach Lapunowa $V(u)$

$$\theta_L^{\varepsilon}(x)V(u) = [\theta_2^{\varepsilon}(x) + \theta_1^{\varepsilon}(x)\boldsymbol{P}R_0Q_1(x) + \theta_0^{\varepsilon}(x)\boldsymbol{P}R_0\tilde{\boldsymbol{L}}(x)]V(u).$$

Uwzględniając wyrażenia (2.52)–(2.54) dla operatorów resztowych $\theta_k^{\varepsilon}(x)$, $k = 0,1,2$, i reprezentacji (2.70) operatora $\widetilde{\boldsymbol{L}}(x)$, wnioskujemy, że w operatorze resztowym $\theta_L^{\varepsilon}(x)$ wykorzystuje się operatory różniczkowania względem zmiennej $u \in R^d$ co najwyżej trzeciego rzędu. Ponadto z warunków twierdzenia wynika, że operatory $\boldsymbol{G}_k^{\varepsilon}(x), k = 1,2,3$, i potencjał R_0 są ograniczone w przestrzeni funkcji $V(u) \in C^3(R^d)$. Ma więc miejsce następujący wniosek.

Wniosek 2.7. Przy założeniach z twierdzenia 2.2 prawdziwe jest następujące oszacowanie

$$|\theta_L^{\varepsilon}(x)V(u)| \leqslant c_L V(u).$$

Korzystając z warunku $C1$ twierdzenia i asymptotycznej reprezentacji z lematu 2.8, sformułujemy następujący wniosek.

Wniosek 2.8. Przy założeniach $C1$–$C3$ z twierdzenia 2.2 dla wszystkich $\varepsilon \leqslant \varepsilon_0$, gdzie ε_0 jest wystarczająco małe ($\varepsilon_0 \leqslant c/c_L$) otrzymujemy kluczową nierówność

$$\boldsymbol{L}^{\varepsilon}V^{\varepsilon}(u,x) \leqslant -cV(u), c > 0. \tag{2.71}$$

Dowód twierdzenia 2.2.

Zakończenie dowodu twierdzenia 2.2 realizowane jest zgodnie ze schematem dowodu twierdzenia 2.1. Reprezentacje (2.68), (2.69) funkcji perturbacyjnych $V_k(u,x)$,

$k = 1,2$ i warunek $C2$ twierdzenia dają po obustronnym oszacowaniu zaburzonej funkcji Lapunowa $V^\varepsilon(u,x)$ następujący ciąg nierówności

$$0 < (1-\varepsilon c)V(u) \leqslant V^\varepsilon(u,x) \leqslant (1+\varepsilon c)V(u). \tag{2.72}$$

Własność martyngałowa procesu $\eta^\varepsilon(t) := V^\varepsilon(u^\varepsilon(\tau^\varepsilon(t)), x(t/\varepsilon^2))$ (1.1.3), $\tau^\varepsilon(t) = \tau^\varepsilon_{v(t)}$, oraz kluczowa nierówność (2.71) charakteryzują ten proces $\eta^\varepsilon(t)$ jako supermartyngał całkowy. Zatem istnieje nieujemna granica v^ε z prawdopodobieństwem 1

$$P\{\lim_{t\to\infty} V^\varepsilon(u^\varepsilon(\tau^\varepsilon(t)), x(t/\varepsilon^2)) = v^\varepsilon\} = 1.$$

Jednocześnie zmienna losowa v^ε ma skończoną wartość oczekiwaną

$$EV^\varepsilon(u^\varepsilon(\tau^\varepsilon(t)), x(t/\varepsilon^2)) \leqslant V^\varepsilon(u,x) \leqslant (1+\varepsilon c)V(u).$$

Biorąc pod uwagę dodatkową własność funkcji Lapunowa

$$V(u) \to \infty, u \to \infty, \tag{2.73}$$

wnioskujemy, że $P\{v^\varepsilon < \infty\} = 1$. Znowu kluczowa nierówność (2.71) i oszacowanie (2.72) dają

$$P\{\lim_{t\to\infty} V(u^\varepsilon(\tau^\varepsilon(t)) = 0\} = 1,$$

tj.

$$P\{\lim_{n\to\infty} u^\varepsilon(\tau_n) = 0\} = 1.$$

Wreszcie dodatniość funkcji Lapunowa $V(u)$ >0, dla $u \neq 0$, własność (2.73) i regularność procesu półmarkowowskiego $x(t), t \geqslant 0$, prowadzą do tezy twierdzenia.

3. PAS w przestrzeni półmarkowowskiej

3.1. PAS w schemacie uśredniania

Ciągły PAS w środowisku półmarkowowskim w schemacie uśredniania jest określony równaniem ewolucji losowej

$$du^\varepsilon(t) = a(t)C(u^\varepsilon(t), x(t/\varepsilon))dt, u^\varepsilon(0) = u. \tag{3.1}$$

Funkcja regresji $C(u,x) = (C_k(u,x), k = \overline{1,d})$ jest funkcją wektorową spełniającą warunki istnienia rozwiązania globalnego układów towarzyszących

$$du_x(t)/dt = C(u_x(t), x), x \in X.$$

Zbieżność PAS (3.1) rozpatrywana jest w warunkach wykładniczej stabilności układu uśrednionego

$$du(t)dt = C(u(t)), C(u) = \int_X \pi(dx)C(u,x), u(0) = u, \tag{3.2}$$

który ma dokładnie jeden punkt równowagi $u_0 : C(u_0) = 0$. Bez zmniejszenia ogólności rozważań zakładamy dalej, że $u_0 = 0$. Tutaj $\pi(B)$, $B \in \boldsymbol{X}$, jest rozkładem stacjonarnym jednostajnie ergodycznym towarzyszącego procesu Markowa $x_0(t), t \geqslant 0$, zadanego przez generator

$$\mathbf{Q}\varphi(x) = q(x) \int_X P(x,dy)[\varphi(y) - \varphi(x)]$$

w mierzalnej przestrzeni fazowej $(X, \boldsymbol{X})$.

Twierdzenie 3.1. *Niech dana będzie funkcja Lapunowa $V(u), u \in R^d$, zapewniająca wykładniczą stabilność układu uśrednionego*

*C*1: $C(u)V'(u) \leqslant -c_0 V(u), c > 0$.

Ponadto, niech dla $\widetilde{C}(u,x) := C(u,x) - C(u)$ spełnione będą dodatkowe warunki

*C*2: $| \boldsymbol{R}_0 \widetilde{C}(u,x)V'(u) | \leqslant c_1(1 + V(u))$,

*C*3: $[C(u,x)\boldsymbol{R}_0[\widetilde{C}(u,x)V'(u)]] \leqslant c_2(1 + V(u))$,

*C*4: $| C(u,x)[C(u,x)[C(u,x)V'(u)]']' | \leqslant c_3(1 + V(u))$.

Niech dystrybuanty $G_x(t), x \in X$, spełniają warunek Cramera jednostajnie na $x \in X$

*C*5: $\sup\limits_{x \in X} \int\limits_0^\infty e^{ht} \overline{G}_x(t)dt \leqslant H < \infty, h > 0$.

Dodatkowo niech funkcja normalizująca $a(t)$ będzie monotonicznie malejąca, ograniczona i spełniająca warunki

*C*6: $\int\limits_0^\infty a(t)dt = \infty, \int\limits_0^\infty a^2(t)dt < \infty,$

$$[a(t) - a(t + \varepsilon s)] \leqslant \varepsilon s a^2(t), \forall (t,s) \geqslant 0.$$

Wtedy dla każdego dodatniego $\varepsilon \leqslant \varepsilon_0$, gdzie ε_0 jest wystarczająco mały, PAS (3.1) *zbiega dla $t \to \infty$ z prawdopodobieństwem 1 do punktu równowagi układu uśrednionego* (3.2)

$$P\{\lim_{t \to \infty} u^\varepsilon(t) = 0\} = 1.$$

Dowód. Rozważmy rodzinę niejednorodnych półgrup $\boldsymbol{C}_s^t(x), 0 \leqslant t \leqslant s$, $x \in X$, wygenerowanych przez PAS

$$du_x(t) = a(t)C(u_x(t), x), \ x \in X.$$

Oznacza to, że funkcje testowe $\varphi(u) \in B(R^d)$ przyjmują postać

$$\boldsymbol{C}_s^t(x)\varphi(u) = \varphi(u_x(s)), u_x(t) = u. \tag{3.3}$$

Operator generujący półgrupy (3.3) jest określony za pomocą wzorów

$$\boldsymbol{C}_t(x)\varphi(u) = a(t)\boldsymbol{C}(x)\varphi(u), \boldsymbol{C}(x)\varphi(u) := C(u,x)\varphi'(u).$$

Definicja 3.1. Operator kompensujący RPOM układu (3.1) jest dany zależnością

$$\mathbf{L}_t^{\varepsilon}\varphi(u,x,t) = \varepsilon^{-1}q(x)[\int_0^{\infty} G_x(ds)\mathbf{C}_{t+s\varepsilon}^t(x)\int_X P(x,dy)\varphi(u,y,t+\varepsilon s) -$$

$$-\varphi(u,x,t)]. \tag{3.4}$$

Uwaga 3.1. Operator kompensujący jest również definiowany przez warunkową wartość oczekiwaną

$$\mathbf{L}_t^{\varepsilon}\varphi(u,x,t) = \varepsilon^1 q(x)E[\varphi(u_{n+1}^{\varepsilon}, x_{n+1}^{\varepsilon}, \tau_{n+1}^{\varepsilon}) - \varphi(u,x,t) \mid u_n^{\varepsilon} = u, x_n^{\varepsilon} = x, \tau_n^{\varepsilon} = t].$$

Lemat 3.1. Operator kompensujący (3.4) na funkcjach $\varphi(u,x) \in C^2(R^d)$ dopuszcza reprezentacje asymptotyczne

$$\mathbf{L}_t^{\varepsilon}\varphi(u,x) = \varepsilon^{-1}\mathbf{Q}\varphi(u,x) + q(x)\mathbf{C}_t(x)\theta_1^{\varepsilon}(t,x)\mathbf{P}\varphi(u,x) \tag{3.5}$$

i

$$\mathbf{L}_t^{\varepsilon}\varphi(u,x) = \varepsilon^{-1}\mathbf{Q}\varphi(u,x) + \mathbf{C}_t(x)\mathbf{P}\varphi(u,x) + \varepsilon q(x)\mathbf{C}_t^2(x)\theta_2^{\varepsilon}(t,x)\mathbf{P}\varphi(u,x). \tag{3.6}$$

Tutaj

$$\theta_1^{\varepsilon}(t,x) = \int_0^{\infty} \overline{G}_x(s)\mathbf{C}_{t+s\varepsilon}^t(x)b_1(t,\varepsilon s)ds = m(x)I + \varepsilon\mathbf{C}_t(x)\theta_2^{\varepsilon}(t,x), \tag{3.7}$$

$$\theta_2^{\varepsilon}(t,x) = \int_0^{\infty} [\overline{G}_x^{(2)}(s)b_1(t,\varepsilon s) + \overline{G}_x(s)b_2(t,\varepsilon s)]\mathbf{C}_{t+s\varepsilon}^t(x)ds \tag{3.8}$$

i

$$b_1(t,\varepsilon s) := a(t+\varepsilon s)/a(t) \leqslant 1,$$

$$b_2(t,\varepsilon s) := [a(t+\varepsilon s) - a(t)]/a^2(t) = b_1(t,\varepsilon s)[1/a(t) - 1/a(t+\varepsilon s)].$$

Jak zawsze

$$\mathbf{P}\varphi(x) := \int_X P(x,dy)\varphi(y).$$

Dowód. Po pierwsze z (3.4) mamy oczywistą zależność

$$\boldsymbol{L}_t^{\varepsilon}\varphi(u,x) = \varepsilon^{-1}\mathbf{Q}\varphi(u,x) + \varepsilon^{-1}q(x)[\boldsymbol{G}_t^{\varepsilon}(x) - \boldsymbol{I}]\boldsymbol{P}\varphi(u,x), \tag{3.9}$$

gdzie

$$\boldsymbol{G}_t^{\varepsilon}(x) := \int_0^{\infty} G_x(ds)\boldsymbol{C}_{t+s\varepsilon}^t(x). \tag{3.10}$$

Teraz, korzystając z równości dla półgrupy $\boldsymbol{C}_s^t(x)$

$$d_s\boldsymbol{C}_{t+s\varepsilon}^t(x) = \varepsilon\boldsymbol{C}_{t+s\varepsilon}(x)\boldsymbol{C}_{t+s\varepsilon}^t(x)ds$$

i wzoru na całkowanie przez części dla (3.10), mamy

$$\boldsymbol{G}_t^{\varepsilon}(x) - \boldsymbol{I} = \varepsilon a(t)\boldsymbol{C}(x)\theta_1^{\varepsilon}(t,x), \tag{3.11}$$

$$\theta_1^{\varepsilon}(t,x) = \int_0^{\infty} \overline{G}_x(s)\boldsymbol{C}_{t+s\varepsilon}^t(x)b_1(t,\varepsilon s)ds. \tag{3.12}$$

Zatem $\theta_1^{\varepsilon}(t,x)$ jest obliczane według pierwszej reprezentacji w (3.7). Przeprowadzając podobne rozważania dla (3.12), mamy

$$\theta_1^{\varepsilon}(t,x) = g(x)\boldsymbol{I} + \varepsilon\boldsymbol{C}_t(x)\theta_2^{\varepsilon}(t,x), \tag{3.13}$$

gdzie $\theta_2^{\varepsilon}(t,x)$ oblicza się ze wzoru (3.8).

Podstawienia (3.11)–(3.13) do (3.9) dają asymptotyczne reprezentacje (3.5) i (3.6). □

Funkcja zaburzenia $V_1(u,x)$ jest wyznaczana poprzez rozwiązanie pojedynczego problemu zaburzeń.

Lemat 3.2. Operator kompensujący (3.4) na zaburzonej funkcji Lapunowa przyjmuje postać

$$\mathbf{L}_t^{\varepsilon}V^{\varepsilon}(u,x,t) = a(t)C(u)V'(u) + \varepsilon\theta_0^{\varepsilon}(t)V(u), \tag{3.14}$$

a funkcja zaburzenia $V_1(u,x)$ ma reprezentację

$$V_1(u,x) = \widetilde{C}_0(u,x)V'(u), \tag{3.15}$$

gdzie

$$\widetilde{C}_0(u,x) := \mathbf{R}_0[C(u,x) - C(u)].$$

W tym przypadku operator resztowy $\theta_0^\varepsilon(t)$ w (3.14) wyznacza zależność

$$\theta_0^\varepsilon(t) = a^2(t)q(x)\mathbf{C}(x)[\mathbf{C}(x)\theta_2^\varepsilon(t,x) + \widetilde{\mathbf{C}}_0(x)\theta_1^\varepsilon(t,x)], \tag{3.16}$$

gdzie

$$\widetilde{\mathbf{C}}_0(x)V(u) := \widetilde{C}_0(u,x)V'(u).$$

Dowód. Reprezentacja (3.14) jest bezpośrednim wnioskiem z rozwiązania pojedynczego problemu zaburzeń, a reprezentacje asymptotyczne (3.5) i (3.6) dają rozkład

$$\begin{aligned}
&\boldsymbol{L}_t^\varepsilon V^\varepsilon(u,x,t) = \boldsymbol{L}_t^\varepsilon[V(u) + \varepsilon a(t)V_1(u,x)] = \\
&= [\varepsilon^{-1}\mathbf{Q} + a(t)\boldsymbol{C}(x)\boldsymbol{P} + \varepsilon a^2(t)q(x)\boldsymbol{C}^2(x)\theta_2^\varepsilon(t,x)\boldsymbol{P}]V(u) + \\
&\quad + a(t)[\varepsilon^{-1}\mathbf{Q} + a(t)q(x)\boldsymbol{C}(x)\theta_1^\varepsilon(t,x)\boldsymbol{P}]\varepsilon V_1(u,x) = \\
&\quad = \varepsilon^{-1}\mathbf{Q}V(u) + [\mathbf{Q}V_1(u,x) + \boldsymbol{C}(x)V(u)] + \\
&\quad + \varepsilon a^2(t)[q(x)\boldsymbol{C}^2(x)\theta_2^\varepsilon(t,x)\boldsymbol{P}V(u) + q(x)\boldsymbol{C}(x)\theta_1^\varepsilon(t,x)\boldsymbol{P}V_1(u,x)].
\end{aligned} \tag{3.17}$$

Wówczas rozwiązanie pojedynczego problemu zaburzeń definiuje $V_1(u,x,t)$ w postaci (3.15) z równania

$$\mathbf{Q}V_1(u,x) + \boldsymbol{C}(x)V(u) = C(u)V'(u).$$

Podstawiając (3.15) do składnika resztowego rozwinięcia (3.1), otrzymujemy reprezentację operatora resztowego $\theta_0^\varepsilon(t)$ w postaci (3.16). □

RPOM (3.4) generuje funkcje testowe $\varphi(u,x,t), u \in R^d, x \in X, t \geqslant 0$ oraz martyngał z parametrem dyskretnym

$$\mu_{n+1}^\varepsilon = \varphi(u_{n+1}^\varepsilon, x_{n+1}^\varepsilon, \tau_{n+1}^\varepsilon) - \varphi(u_0, x_0, \tau_0) - \varepsilon\sum_{k=0}^{n}\theta_{k+1}\boldsymbol{L}_{\tau_k^\varepsilon}^\varepsilon\varphi(u_k^\varepsilon, x_k^\varepsilon, \tau_k^\varepsilon) \tag{3.18}$$

względem σ-algebry

$$F_{\tau_n^\varepsilon}^\varepsilon = \sigma\{u_s^\varepsilon, x_s^\varepsilon, \tau_s^\varepsilon, 0 \leqslant s \leqslant \tau_n^\varepsilon\}, n \geqslant 0.$$

Wprowadźmy skrócone oznaczenia

$$\varphi_n^\varepsilon = \varphi(u_n^\varepsilon, x_n^\varepsilon, \tau_n^\varepsilon), n \geqslant 0.$$

Następnie z (3.18) mamy

$$\mu^{\varepsilon}_{n+1} - \mu^{\varepsilon}_{n} = \varphi^{\varepsilon}_{n+1} - \varphi^{\varepsilon}_{n} - \varepsilon\theta_{n+1}\boldsymbol{L}^{\varepsilon}_{\tau^{\varepsilon}_{n}}\varphi^{\varepsilon}_{n}. \tag{3.19}$$

Zgodnie z definicją operatora kompensującego i z uwzględnieniem $E[\theta_{n+1} \mid F^{\varepsilon}_{\tau^{\varepsilon}_{n}}] = g(x^{\varepsilon}_{n}) = 1/q(x^{\varepsilon}_{n})$, obliczamy wartość oczekiwaną z (3.19)

$$E[\mu^{\varepsilon}_{n+1} - \mu^{\varepsilon}_{n} | F^{\varepsilon}_{n}] = E[\varphi^{\varepsilon}_{n+1} - \varphi^{\varepsilon}_{n} | F^{\varepsilon}_{n}] - E[\varphi^{\varepsilon}_{n+1} - \varphi^{\varepsilon}_{n} | F^{\varepsilon}_{n}] = 0.$$

Wprowadźmy proces z czasem ciągłym

$$w^{\varepsilon}(t) := u^{\varepsilon}(\tau^{\varepsilon}(t)), \quad x^{\varepsilon}_{t} := x(t/\varepsilon), t \geqslant 0. \tag{3.20}$$

Zgodnie z definicją procesu punktowego $\tau^{\varepsilon}(t) := \tau^{\varepsilon}_{\nu^{\varepsilon}(t)}$, $\nu^{\varepsilon}(t) = \nu(\varepsilon t)$, mamy

$$w^{\varepsilon}_{n} := w^{\varepsilon}(\tau^{\varepsilon}_{n}) = u^{\varepsilon}_{n} = u^{\varepsilon}(\tau^{\varepsilon}_{n}), n \geqslant 0. \tag{3.21}$$

□

Dlatego proces $w^{\varepsilon}(t)$ jest dyskretnym włożeniem w PAS (3.1).

Lemat 3.3. Proces (3.21) jest wyznaczany przez martyngałową własność procesu

$$\mu^{\varepsilon}(t) = V^{\varepsilon}(w^{\varepsilon}(t), x^{\varepsilon}_{t}, \tau^{\varepsilon}(t)) - \int\limits_{0}^{\tau^{\varepsilon}(t)} \mathbf{L}^{\varepsilon}_{\tau^{\varepsilon}(s)} V^{\varepsilon}(w^{\varepsilon}(s), x^{\varepsilon}_{s}, \tau^{\varepsilon}(s))ds.$$

Dowód. Użyjmy równości $\mu^{\varepsilon}(\tau^{\varepsilon}_{n}) = \mu^{\varepsilon}_{n}, n \geqslant 0$, gdzie μ^{ε}_{n} zdefiniowano w (3.18). Dla $s > t$ mamy

$$E[\mu^{\varepsilon}(s) - \mu^{\varepsilon}(t) \mid F^{\varepsilon}_{\tau^{\varepsilon}(t)}] = E[\sum_{k=0}^{\nu^{\varepsilon}(s)-1} [V^{\varepsilon}_{k+1} - V^{\varepsilon}_{k} - \varepsilon\theta_{k+1}\boldsymbol{L}^{\varepsilon}_{\tau^{\varepsilon}(t)+k}V^{\varepsilon}_{k}] \mid F^{\varepsilon}_{\tau^{\varepsilon}(t)}] =$$

$$= E[\sum_{k=0}^{\nu^{\varepsilon}(s)-1} [\mu^{\varepsilon}_{k+1} - \mu^{\varepsilon}_{k} \mid F^{\varepsilon}_{\tau^{\varepsilon}(t)}] = 0.$$

□

Następnie ustalamy kluczową nierówność.

Lemat 3.4. Operator kompensujący dla zaburzonej funkcji Lapunowa zgodnie z warunkami twierdzenia 3.1 można oszacować w następujący sposób

$$\mathbf{L}^{\varepsilon}_{t} V^{\varepsilon}(u,x,t) \leqslant -\delta a(t)V(u), \delta > 0. \tag{3.22}$$

Dowód. Najpierw szacujemy wpływ operatorów resztowych (3.7)–(3.8) na funkcję Lapunowa $V(u)$, korzystając z estymacji dla półgrup

$$\mid \boldsymbol{C}^{\varepsilon}_{t+\varepsilon s}(x)V(u) \mid \leqslant ce^{\int\limits_{t}^{t+\varepsilon s} a(v)dv} \mid \boldsymbol{C}_t(x)V(u) \mid,$$

i warunków Cramera $C5$. □

Biorąc pod uwagę monotoniczność funkcji normalizującej $a(t)$ i jej ograniczenie $a(t) \leqslant a$, a także reprezentację (3.7) dla operatora resztowego $\theta^{\varepsilon}_1(t,x)$, ustalamy oszacowanie

$$\mid \theta^{\varepsilon}_1(t,x)V(u) \mid \leqslant cV(u). \tag{3.23}$$

Podobne oszacownie zachodzi dla drugiego operatora resztowego (3.8)

$$\mid \theta^{\varepsilon}_2(t,x)V(u) \mid \leqslant c_2V(u). \tag{3.24}$$

W tym przypadku stosowany jest dodatkowy warunek $C3$. Główny warunek $C1$ twierdzenia 3.1 pozwala nam zakończyć szacowanie (3.22) dla każdego $\varepsilon \leqslant \varepsilon_0$, jeśli ε_0, jest wystarczająco małe. Mianowicie $c - \varepsilon_0 ac_0 \geqslant \delta, \varepsilon_0 \leqslant (c-\delta)/ac_0$.

Reprezentacja (3.15) funkcji zaburzenia $V_1(u,x)$ i warunek $C2$ twierdzenia 3.1 dają oszacowania

$$0 < (1 - \varepsilon a(t)c)V(u) \leqslant V^{\varepsilon}(u,x,t) \leqslant (1 + \varepsilon a(t)c)V(u).$$

Kluczowa nierówność (3.22) i oszacowania (3.23)–(3.24) umożliwiają uzupełnienie dowodu twierdzenia 3.1, wykorzystując wyniki Nevelsona-Hasminsky'ego i twierdzenie 1.7 [33].

Własność martyngałowa procesu $V^{\varepsilon}(u^{\varepsilon}(\tau^{\varepsilon}(t)), x^{\varepsilon}_t, \tau^{\varepsilon}(t))$ i kluczowa nierówność (3.22) pozwalają scharakteryzować ten proces jako supermartyngał całkowy. Zatem istnieje skończona granica z prawdopodobieństwem jeden

$$P\{\lim_{t\to\infty} V^{\varepsilon}(u^{\varepsilon}(\tau^{\varepsilon}(t)), x^{\varepsilon}_t, t) = v^{\varepsilon}\} = 1.$$

Jednocześnie zmienna losowa v^{ε} ma skończoną wartość oczekiwaną

$$EV^{\varepsilon}(u^{\varepsilon}(\tau^{\varepsilon}(t)), x^{\varepsilon}_t, t)V^{\varepsilon}(u,x,0) \leqslant (1 + \varepsilon a(t)c)V(u) \leqslant cV(u).$$

Biorąc pod uwagę dodatkową własność funkcji Lapunowa

$$V(u) \to \infty, \| u \| \to \infty,$$

wnioskujemy, że

$$P\{v^\varepsilon < \infty\} = 1.$$

Ponownie kluczową nierównością jest (3.22) i oszacowania (3.23)–(3.24)

$$E\int_0^\infty a(s)V(u^\varepsilon(\tau^\varepsilon(s))ds \leqslant cV(u).$$

Zatem rozważana całka jest zbieżna z prawdopodobieństwem 1.

$$P\{\int_0^\infty a(s)V(u^\varepsilon(\tau^\varepsilon(s))ds < \infty\} = 1.$$

Ponadto, zgodnie z warunkami twierdzenia 3.1

$$P\{\lim_{t\to\infty} V(u^\varepsilon(\tau^\varepsilon(t)) = 0\} = 1,$$

to znaczy

$$P\{\lim_{t\to\infty} u^\varepsilon(\tau_n) = 0\} = 1. \tag{3.25}$$

Wreszcie, dodatniość funkcji Lapunowa $V(u), u \neq 0$, własność (3.25) i regularność procesu półmarkowowskiego $x(t), t \geqslant 0$, dają tezę twierdzenia 3.1.

Dla PAS (3.26) z $\sigma(u) \neq 0$

$$du^\varepsilon(t) = a(t)[C(u^\varepsilon(t), x(t/\varepsilon)dt + \sigma(u^\varepsilon(t))dw(t)], u^\varepsilon(0) = u \tag{3.26}$$

zachodzi następujące twierdzenie.

Twierdzenie 3.2. *Niech będą spełnione warunki C*1*–C*3 *oraz C*5 *i C*6 *twierdzenia 3.1, a także dodatkowy warunek*

$$C4' :| \sigma^2(u)R_0[C(u,x)V'(u)]'' | \leqslant c_4[1+V(u)].$$

Następnie dla każdego dodatniego $\varepsilon \leqslant \varepsilon_0$, ε_0 *jest wystarczająco mały, PAS* (3.26) *zbiega dla* $t \to \infty$ *z prawdopodobieństwem 1 do punktu równowagi układu uśrednionego* (3.2)

$$P\{\lim_{t\to\infty} u^\varepsilon(t) = 0\} = 1.$$

Dowód. Dowód twierdzenia 3.2 przeprowadza się jak poprzednio, uwzględniając niezerowe zaburzenie w (3.26). □

3.2. PAS w schemacie aproksymacji dyfuzjnej

Ciągły PAS z przełączeniami półmarkowowskimi w schemacie aproksymacji dyfuzji jest określony równaniem ewolucji

$$du^{\varepsilon}(t) = a(t)C^{\varepsilon}(u^{\varepsilon}(t),x(t/\varepsilon^2))dt, u^{\varepsilon}(0) = u, \tag{3.27}$$

gdzie

$$C^{\varepsilon}(u,x) = C(u,x) + \varepsilon^{-1}C_0(u,x).$$

Funkcja regresji $C(u,x) = (C_k(u,x), k = \overline{1,d})$, $u \in R^d$, $x \in X$ i jego zaburzenie $C_0(u,x) = (C_{0k}(u,x), k = \overline{1,d})$, $u \in R^d$, $x \in X$, spełniają warunki istnienia rozwiązania globalnego systemów towarzyszących (patrz 1.2.3)

$$du_x(t) = C^{\varepsilon}(u_x(t),x)dt, x \in X.$$

Przyjmujemy dodatkowe założenia:

*P*1: Funkcja osobliwa perturbacyjna $C_0(u,x)$, $u \in R^d$, $x \in X$, spełnia *warunek równowagi*

$$\int_X \pi(dx)C_0(u,x) \equiv 0. \tag{3.28}$$

*P*2: Zbieżność PAS (3.27) rozpatrywana jest w warunkach wykładniczej stabilności uśrednionego układu dynamicznego

$$du(t)/dt = C(u(t)), C(u) := \int_X \pi(dx)C(u,x), u(0) = u, \tag{3.29}$$

który ma dokładnie jeden punkt równowagi $u_0 = 0$. Jednocześnie wykorzystywane są warunki zbieżności rozwiązania równania ewolucji

$$du_0^{\varepsilon}(t) = [C(u_0^{\varepsilon}(t),x(t/\varepsilon^2)) + \varepsilon^{-1}C_0(u_0^{\varepsilon}(t),x(t/\varepsilon^2))]dt, \tag{3.30}$$

do rozwiązania stochastycznego równania różniczkowego

$$du_0(t) = C(u_0(t))dt + d\zeta(t),$$

$$d\zeta(t) = b(u_0(t))dt + \frac{1}{2}SpB(u_0(t))dw_t.$$

Tutaj zdefiniowany jest proces dyfuzji $\zeta(t)$, $t \geqslant 0$, funkcja przesunięcia

$$b(u) = b_0(u) + b_{\mu}(u), \tag{3.31}$$

gdzie

$$b_0(u) := \int\limits_X \pi(dx) C_0(u,x) R_0 C_0'(u,x), b_\mu(u) := \int\limits_X \pi(dx)\mu(x) C_0(u,x) C_0'(u,x),$$

$$C_0'(u,x) := \frac{\partial C_0(u,x)}{\partial u} = \left\{ C_0^{kr}(u,x) := \frac{\partial C_0^k(u,x)}{\partial u_r}; k,r = \overline{1,d} \right\},$$

z macierzą dyfuzji

$$B(u) = B_0(u) + B_\mu(u), \tag{3.32}$$

$$B_0(u) := 2\int\limits_X \pi(dx) C_0(u,x) R_0 C_0(u,x), B_\mu(u) := \int\limits_X \pi(dx)\mu(x) C_0^T(u,x) C_0(u,x)$$

i

$$\mu(x) := (g_2(x) - 2g^2(x))/(g(x)),$$

gdzie $g_2(x)$ określa drugi moment zmiennej losowej określającej czas trwania w stanie $x \in X$

$$g_2(x) := \int\limits_0^\infty t^2 G_x(dt) = 2\int\limits_0^\infty \bar{G}_x^{(2)}(t)dt, \bar{G}_x^{(2)}(t) := \int\limits_t^\infty \bar{G}_x(s)ds.$$

Parametr $\mu(x)$ określa odchylenie od rozkładu wykładniczego, gdyż $\mu(x) = 0$ dla $G_x(t) = 1 - e^{-q(x)t}$, $t \geqslant 0$.

Procedura uśredniania (3.29) jest definiowana przez zaburzenie dyfuzyjne z niezerowym przesunięciem (3.31). Jeśli jednak $C_0(u,x) = C_0(x)$, tj. nie zależy od u, to oczywiście $b(u) = 0$.

Twierdzenie 3.3. *Niech funkcja Lapunowa $V(u), u \in R^d$, zapewnia wykładniczą stabilność układu uśrednionego* (3.29)

$C1$: $C(u)V'(u) \leqslant -c_0 V(u), c_0 > 0.$

Niech spełnione będą również dodatkowe warunki

$C2$: $| C_0(u,x)R_0[C_0(u,x)V'(u)]' | \leqslant c_1(1+V(u)),$

$C3$: $| C_i(u,x)R_0[\widetilde{C}(u,x)V'(u)]' | \leqslant c_2(1+V(u)),\ i = 0,1,$

$C4$: $| C_i(u,x)R_0[C_0(u,x)R_0[C_0(u,x)V'(u)]']' | \leqslant c_3(1+V(u)),\ i = 0,1,$

gdzie

$$\widetilde{C}(u,x) := C(u,x) - C(u), C_1(u,x) := C(u,x).$$

Dodatkowo niech funkcja rozkładu $G_x(t), x \in X$, spełnia warunek Cramera jednostajnie dla $x \in X$

$C5$: $\sup\limits_{x\in X} \int\limits_0^\infty e^{ht}\overline{G}_x(t)dt \leqslant H < \infty, h > 0.$

Wreszcie niech funkcja normalizująca $a(t)$ będzie monotonicznie malejąca, ograniczona i spełniająca warunki

$C6$: $\int\limits_0^\infty a(t)dt = \infty, \int\limits_0^\infty a^2(t)dt < \infty,$

$$a(t+\varepsilon^2 s)/a(t) \leqslant 1,$$

$$\mid a_t'(t+\varepsilon^2 s)/a^2(t) \mid \leqslant A < \infty,$$

$$\mid a_t''(t+\varepsilon^2 s)/a^3(t) \mid \leqslant A < \infty, \forall t, s \geqslant 0.$$

Następnie dla każdego dodatniego $\varepsilon \leqslant \varepsilon_0$, gdzie ε_0 jest wystarczająco mały, PAS (3.27) *dąży z prawdopodobieństwem 1 do dokładnie jednego punktu równowagi uśrednionej ewolucji* (3.29)

$$P\{\lim_{t\to\infty} u^\varepsilon(t) = 0\} = 1.$$

Rozważmy rozszerzony proces odzyskiwania Markowa (RPOM)

$$u_n^\varepsilon = u^\varepsilon(\tau_n^\varepsilon), x_n^\varepsilon = x(\tau_n^\varepsilon), \tau_n^\varepsilon = \varepsilon^2\tau_n, n \geqslant 0. \tag{3.33}$$

Tutaj $\tau_n := \sum\limits_{k=1}^{n} \theta_k$, $n \geqslant 0$, $\tau_0 = 0$ oznaczają momenty odnowy procesu półmarkowowskiego.

Rozważmy także klasę niejednorodnych półgrup $\boldsymbol{C}_{t+s}^{\varepsilon,t}(x)$, $t \geqslant 0, s \geqslant 0$, $x \in X$, która jest generowana przez towarzyszącą PAS

$$du_x(t)/dt = a(t)C^\varepsilon(u_x(t),x), x \in X, u_x(t_0) = u, \tag{3.34}$$

to znaczy na funkcjach testowych $\varphi(u) \in C^1(R^d)$ mamy

$$\boldsymbol{C}_{t+s}^{\varepsilon,t}(x)\varphi(u) = \varphi(u_x(t+s)), u_x(t) = u.$$

Tutaj używamy skróconego zapisu rozwiązania układu (3.34)

$$u_x(t+s) := u_x(t+s;u), u_x(t;u) = u.$$

Operator generujący $\boldsymbol{C}_t^{\varepsilon}(x)$ półgrup $\boldsymbol{C}_{t+s}^{\varepsilon,t}(x)$ jest wyznaczany przez formuły

$$\boldsymbol{C}_t^{\varepsilon}(x)\varphi(u) = a(t)\boldsymbol{C}^{\varepsilon}(x)\varphi(u),$$

gdzie

$$\boldsymbol{C}^{\varepsilon}(x)\varphi(u) := C^{\varepsilon}(u,x)\varphi'(u)$$

oraz

$$\varphi'(u) := \{\partial\varphi(u)/\partial u_k, k = \overline{1,d}\},$$

lub w formie operatora

$$\boldsymbol{C}^{\varepsilon}(x)\varphi(u) = \boldsymbol{C}(x)\varphi(u) + \varepsilon^{-1}\boldsymbol{C}_0(x)\varphi(u), \tag{3.35}$$

gdzie

$$\boldsymbol{C}(x)\varphi(u) := C(u,x)\varphi'(u), \boldsymbol{C}_0(x)\varphi(u) := C_0(u,x)\varphi'(u).$$

Wiadomo, że operator kompensujący RPOM (3.33) jest określony przez relację

$$\boldsymbol{L}_t^{\varepsilon}\varphi(u,x,t) = \varepsilon^{-2}q(x)[\int_0^{\infty} G_x(ds)\boldsymbol{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x)\int_X P(x,dy)\varphi(u,y,t+\varepsilon^2 s) -$$

$$-\varphi(u,x,t)]. \tag{3.36}$$

Lemat 3.5. Operator kompensujący (3.36) na funkcjach testowych $\varphi(u,\cdot) \in C^2(R^d)$ ma reprezentację

$$\boldsymbol{L}_t^{\varepsilon}\varphi(u,x) = \varepsilon^{-2}\mathbf{Q}\varphi(u,x) + a(t)\mathbf{C}^{\varepsilon}(x)\mathbf{G}_t^{\varepsilon,0}(x)\mathbf{Q}_0\varphi(u,x). \tag{3.37}$$

Operator $\mathbf{G}_t^{\varepsilon,0}(x)$, $x \in X$, przyjmuje postać

$$\mathbf{G}_t^{\varepsilon,0}(x) := \int_0^{\infty} \overline{G}_x(s)\mathbf{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x)a_1(t,\varepsilon^2 s)ds, x \in X, \tag{3.38}$$

$$a_1(t,\varepsilon^2 s) := a(t+\varepsilon^2 s)/a(t),$$

a operator $\mathbf{Q}_0$ jest zdefiniowany przez operator $\mathbf{P}$.

$$\mathbf{Q}_0 := q(x)\mathbf{P}.$$

Dowód. Najpierw wybieramy w (3.36) generator $\mathbf{Q}$, używając składnika $\pm\varphi(u,y)$ w całce wewnętrznej

$$\boldsymbol{L}_t^{\varepsilon}\varphi(u,x) = \varepsilon^{-2}q(x)\int\limits_X P(x,dy)[\varphi(u,y)-\varphi(u,x)]+$$

$$+\varepsilon^{-2}q(x)\int\limits_0^{\infty} G_x(ds)[\boldsymbol{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x)-I]\boldsymbol{P}\varphi(u,x). \tag{3.39}$$

Pierwszy człon w (3.39) daje generator $\mathbf{Q}$, a z drugiego członu przez całkowanie przez części, biorąc pod uwagę równanie półgrupy w postaci

$$d\boldsymbol{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x) = \varepsilon^2\boldsymbol{C}_{t+\varepsilon^2 s}^{\varepsilon}(x)\boldsymbol{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x)ds,$$

mamy operator $\boldsymbol{G}_t^{\varepsilon,0}(x)$ w postaci (3.38). □

Lemat 3.6. Operator kompensujący (3.36) na funkcjach testowych $\varphi(u,\cdot)\in C^4(R^d)$ ma asymptotyczną reprezentację

$$\mathbf{L}_t^{\varepsilon}\varphi(u,x) = \varepsilon^{-2}\mathbf{Q}\varphi(u,x)+\varepsilon^{-1}a(t)\mathbf{C}_0(x)\mathbf{P}\varphi(u,x)+$$

$$+a(t)[\mathbf{C}(x)+a(t)\mu_2(x)\mathbf{C}_0^2(x)]\mathbf{P}\varphi(u,x)+\varepsilon\theta_L^{\varepsilon}(x,t)\mathbf{P}\varphi(u,x), \tag{3.40}$$

gdzie

$$\mu_2(x) = g_2(x)/(2g(x)),$$

$$\theta_L^{\varepsilon}(x,t) = a^2(t)\mu_2(x)[2\mathbf{C}(x)\mathbf{C}_0(x)+\mathbf{C}_{\varepsilon}(x)\frac{a'(t)}{a^2(t)}+$$

$$+\varepsilon[\mathbf{C}^2(x)+a(t)\frac{2}{g_2(x)}\mathbf{C}_{\varepsilon}(x)[\mathbf{C}_{\varepsilon}(x)\mathbf{G}_t^{\varepsilon,21}(x)+\varepsilon\mathbf{G}_t^{\varepsilon,22}(x)]]]. \tag{3.41}$$

Ponadto

$$\mathbf{C}_{\varepsilon}(x) := \varepsilon\mathbf{C}^{\varepsilon}(x) = \varepsilon\mathbf{C}(x)+\mathbf{C}_0(x), \tag{3.42}$$

$$\mathbf{G}_t^{\varepsilon,21}(x) := \int\limits_0^{\infty}\overline{G}_x^{(3)}(s)\mathbf{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x)a_1(t,\varepsilon^2 s)[2a_2(t,\varepsilon^2 s)+$$

$$+a_1^2(t,\varepsilon^2 s)\mathbf{C}^{\varepsilon}(x)]ds, x\in X, \tag{3.43}$$

$$a_2(t,\varepsilon^2 s) := a_t'(t+\varepsilon^2 s)/a^2(t),$$

$$\mathbf{G}_t^{\varepsilon,22}(x) := \int\limits_0^{\infty}\overline{G}_x^{(3)}(s)\mathbf{C}_{t+\varepsilon^2 s}^{\varepsilon,t}(x)[a_3(t,\varepsilon^2 s)+a_1^2(t,\varepsilon^2 s)\mathbf{C}^{\varepsilon}(x)]ds, \tag{3.44}$$

$$a_3(t,\varepsilon^2 s) := \frac{a''(t+\varepsilon^2 s)}{a^3(t)},$$

$$\overline{G}_x^{(3)}(t) := \int_t^\infty \overline{G}_x^{(2)}(s)ds.$$

Zauważmy, że zgodnie z warunkiem $C6$ twierdzenia i reprezentacji (3.41)–(3.44) operator resztowy $\theta_L^\varepsilon(x,t)$ w (3.41) jest nieistotny w przypadku funkcji testowych $\varphi(u) \in C^4(R^d)$

$$\|\theta_L^\varepsilon(x,t)\varphi(u)\| \to 0, \varepsilon \to 0.$$

Dowód. Całkując przez części z (3.38), otrzymujemy

$$\boldsymbol{G}_t^{\varepsilon,0}(x) = g(x)I + \varepsilon^2 a(t)\boldsymbol{G}_t^{\varepsilon,1}(x), \tag{3.45}$$

gdzie

$$\boldsymbol{G}_t^{\varepsilon,1}(x) = \boldsymbol{G}_t^{\varepsilon,11}(x) + \boldsymbol{G}_t^{\varepsilon,12}(x) \tag{3.46}$$

oraz

$$\boldsymbol{G}_t^{\varepsilon,11}(x) := \boldsymbol{C}^\varepsilon(x) \int_0^\infty \overline{G}_x^{(2)}(s)\boldsymbol{C}_{t+\varepsilon^2 s}^\varepsilon(x)a_1^2(t,\varepsilon^2 s)ds, \tag{3.47}$$

$$\boldsymbol{G}_t^{\varepsilon,12}(x) := \int_0^\infty \overline{G}_x^{(2)}(s)\boldsymbol{C}_{t+\varepsilon^2 s}^\varepsilon(x)a_2(t,\varepsilon^2 s)ds. \tag{3.48}$$

Podobnie, całkując przez części z (3.47), otrzymujemy

$$\boldsymbol{G}_t^{\varepsilon,11}(x) = (g_2(x)/2)\boldsymbol{C}^\varepsilon(x) + \varepsilon^2 a(t)\boldsymbol{C}^\varepsilon(x)\boldsymbol{G}_t^{\varepsilon,21}(x), \tag{3.49}$$

gdzie $\boldsymbol{G}_t^{\varepsilon,21}(x)$ oblicza się ze wzoru (3.43).

Dla (3.48) całkowanie przez części daje

$$\boldsymbol{G}_t^{\varepsilon,12}(x) = (g_2(x)/2)\frac{a'(t)}{a^2(t)}I + \varepsilon^2 a(t)\boldsymbol{G}_t^{\varepsilon,22}(x), \tag{3.50}$$

gdzie $\boldsymbol{G}_t^{\varepsilon,22}(x)$ oblicza się ze wzoru (3.44).

Używając reprezentacji (3.49) i (3.50) w (3.46), a następnie w (3.45) i biorąc pod uwagę definicję operatora $\boldsymbol{C}^\varepsilon(x)$ (3.35), z (3.37) mamy (3.40).

□

Kluczowym punktem dowodu twierdzenia 3.3 jest rozwiązanie problemu zaburzenia osobliwego dla obciętego operatora kompensującego z (3.40), mianowicie

dla operatora

$$\boldsymbol{L}_{t0}^{\varepsilon} = \varepsilon^{-2}\mathbf{Q} + \varepsilon^{-1}a(t)\boldsymbol{C}_0(x)\,\boldsymbol{P} + a(t)[\boldsymbol{C}(x) + a(t)\mu_2(x)\boldsymbol{C}_0^2(x)]\boldsymbol{P}, \tag{3.51}$$

na zaburzonej funkcji Lapunowa

$$V^{\varepsilon}(u,x) = V(u) + \varepsilon a(t)V_1(u,x) + \varepsilon^2 a(t)V_2(u,x). \tag{3.52}$$

Lemat 3.7. Operator kompensujący $\mathbf{L}_{t0}^{\varepsilon}$ na zaburzonej funkcji Lapunowa $V^{\varepsilon}(u,x)$, przy $V(u) \in C^4(R^d)$, ma reprezentację

$$\mathbf{L}_{t0}^{\varepsilon}V^{\varepsilon}(u,x) = \mathbf{L}_t V(u) + \varepsilon\theta_0^{\varepsilon}(t)V(u), \tag{3.53}$$

gdzie

$$\mathbf{L}_t V(u) = a(t)\mathbf{C}V(u) + a^2(t)\mathbf{L}_0 V(u), \tag{3.54}$$

$$\mathbf{C}V(u) = C(u)V'(u),$$

$$\mathbf{L}_0 V(u) = b(u)V'(u) + \frac{1}{2}Sp[B(u)V''(u)], \tag{3.55}$$

gdzie $b(u)$ oraz $B(u)$ obliczamy przez (3.31) i (3.32).

Operator resztowy $\theta_0^{\varepsilon}(t)V(u)$ spełnia warunek $|\theta_0^{\varepsilon}(t)V(u)| \leqslant C_{\theta}$.

Dowód. Zgodnie z lematem 2.3 rozwiązanie problemu zaburzenia osobliwego dla operatora (3.51) na funkcjach (3.52) prowadzi do rozwinięcia

$$\begin{aligned}
\boldsymbol{L}_{t0}^{\varepsilon}V^{\varepsilon}(u,x) = [\varepsilon^{-2}Q+ \\
+\varepsilon^{-1}a(t)\boldsymbol{C}_0(x)\,\boldsymbol{P} + a(t)[\boldsymbol{C}(x) + a(t)\mu_2(x)\boldsymbol{C}_0^2(x)]\boldsymbol{P}QV(u)] \\
[V(u) + \varepsilon a(t)V_1(u,x) + \varepsilon^2 a(t)V_2(u,x)] = \\
= \varepsilon^{-2}QV(u) + \varepsilon^{-1}a(t)[QV_1(u,x) + \boldsymbol{C}_0(x)V(u)] + \\
+a(t)QV_2(u,x) + a^2(t)\boldsymbol{C}_0(x)\boldsymbol{P}V_1(u,x) + a(t)\boldsymbol{C}(x)\boldsymbol{P}V(u) + \\
+a^2(t)\mu_2(x)\boldsymbol{C}_0^2(x)\boldsymbol{P}V(u) + \theta_0^{\varepsilon}(t)V(u),
\end{aligned} \tag{3.56}$$

ze składnikiem resztowym mającym postać

$$\begin{aligned}
\theta_0^{\varepsilon}(t)V(u) = a^2(t)[[\varepsilon\boldsymbol{C}(x) + \boldsymbol{C}_0(x)]\boldsymbol{P}V_2(u,x) + \boldsymbol{C}(x)\boldsymbol{P}V_1(u,x)] + \\
+a^2(t)\mu_2(x)\boldsymbol{C}_0^2(x)\boldsymbol{P}[V_1(u,x) + \varepsilon V_2(u,x)].
\end{aligned} \tag{3.57}$$

Zatem z (3.56) mamy układ równań

$$\mathbf{Q}V(u) = 0. \tag{3.58}$$

$$\mathbf{Q}V_1(u,x) + \boldsymbol{C}_0(x)V(u) = 0, \tag{3.59}$$

$$a(t)\mathbf{Q}V_2(u,x) + a^2(t)\boldsymbol{C}_0(x)\boldsymbol{P}V_1(u,x) + a(t)\boldsymbol{C}(x)\boldsymbol{P}V(u) + \\ + a^2(t)\mu_2(x)\boldsymbol{C}_0^2(x)\boldsymbol{P}V(u) = \boldsymbol{L}_tV(u). \tag{3.60}$$

Równanie (3.58) jest prawdziwe, ponieważ $V(u)$ nie zależy od $x \in X$. Warunek równowagi (3.28) zapewnia rozwiązywalność równania (3.59), którego rozwiązanie można podać w formie

$$V_1(u,x) = \mathbf{R}_0\boldsymbol{C}_0(x)V(u). \tag{3.61}$$

Podstawianie (3.61) do (3.60) daje

$$\boldsymbol{Q}V_2(u,x) + \boldsymbol{L}_t(x)V(u) = \boldsymbol{L}_tV(u),$$

gdzie

$$\boldsymbol{L}_t(x) = a(t)\boldsymbol{C}(x) + a^2(t)\boldsymbol{L}_0(x) \tag{3.62}$$

i

$$\boldsymbol{L}_0(x) := \boldsymbol{C}_0(x)\boldsymbol{P}\mathbf{R}_0\boldsymbol{C}_0(x) + \mu_2(x)\boldsymbol{C}_0^2(x). \tag{3.63}$$

Przy obliczaniu pierwszego składnika w (3.63) skorzystamy z rozkładu

$$\boldsymbol{P}\mathbf{R}_0 = \mathbf{R}_0 + g(x)[\Pi - I].$$

Otrzymujemy

$$\boldsymbol{L}_0(x) = \boldsymbol{C}_0(x)R_0\boldsymbol{C}_0(x) - g(x)\boldsymbol{C}_0^2(x) + \mu_2(x)\boldsymbol{C}_0^2(x) = \\ = \boldsymbol{C}_0(x)R_0\boldsymbol{C}_0(x) + \mu(x)\boldsymbol{C}_0^2(x). \tag{3.64}$$

W ten sposób otrzymujemy $\boldsymbol{L}_t$ w (3.53), korzystając z postaci $\boldsymbol{L}_t = \Pi\boldsymbol{L}_t(x)\Pi$. Zatem używamy (3.64) w (3.62) i z tego ostatniego mamy

$$\boldsymbol{L}_t = a(t)\Pi\boldsymbol{C}(x)\Pi + a^2(t)\Pi\boldsymbol{L}_0(x)\Pi. \tag{3.65}$$

Pierwszy wyraz w (3.65) daje pierwszy wyraz w (3.54). Obliczenie drugiego wyrazu w (3.65) prowadzi do drugiego wyrazu w (3.54).

Wreszcie z (3.57), (3.61) i z faktu, że

$$V_2(u,x) = R_0[\boldsymbol{L}_t(x) - \boldsymbol{L}_t]V(u), \tag{3.66}$$

otrzymujemy ograniczenie składnika resztowego $\theta_0^\varepsilon(t)V(u)$ dla $V(u) \in C^4(R^d)$.
□

Lemat 3.8. Rozwiązanie problemu zaburzenia osobliwego dla operatora (3.40) na funkcjach testowych (3.52) z $V(u) \in C^4(R^d)$ ma reprezentację

$$\mathbf{L}_t^\varepsilon V^\varepsilon(u,x) = \mathbf{L}_t V(u) + \varepsilon \mathbf{H}_L^\varepsilon(x,t)V(u), \tag{3.67}$$

gdzie

$$\mathbf{H}_L^\varepsilon(x,t)V(u) = [\theta_L^\varepsilon(x,t) + \theta_0^\varepsilon(t)]V(u) + \\ + \varepsilon\theta_L^\varepsilon(x,t)[V_1(u,x) + \varepsilon V_2(u,x)]. \tag{3.68}$$

Tutaj $\theta_L^\varepsilon(x,t)$ i $\theta_0^\varepsilon(t)$ mają odpowiednio reprezentacje (3.41) i (3.57), a $V_1(u,x)$ i $V_2(u,x)$ reprezentacje (3.61) i (3.66).

Dowód. Na podstawie (3.40) i (3.51), dla operatora $\boldsymbol{L}_t^\varepsilon$ mamy

$$\boldsymbol{L}_t^\varepsilon = \boldsymbol{L}_{t0}^\varepsilon + \varepsilon\theta_L^\varepsilon(x,t).$$

Zatem, biorąc pod uwagę (3.53) dla operatora kompensującego $\boldsymbol{L}_t^\varepsilon$, mamy reprezentację (3.67). □

Dla zakończenia dowodu twierdzenia 3.3 zauważmy, że z warunków $C2$–$C6$ twierdzenia i reprezentacji (3.55) otrzymujemy oszacowanie

$$|\boldsymbol{L}_0 V(u)| \leqslant c(1+V(u)) \tag{3.69}$$

i z reprezentacji (3.68) oszacowanie

$$|\boldsymbol{H}_L^\varepsilon(x,t)V(u)| \leqslant c(1+V(u)). \tag{3.70}$$

Korzystając z warunku $C1$ twierdzenia, oszacowania (3.69), (3.70) i z reprezentacji (3.67), otrzymujemy oszacowanie dla operatora kompensującego

$$\boldsymbol{L}_t^\varepsilon V^\varepsilon(u,x) \leqslant -\delta a(t)V(u). \tag{3.71}$$

Kluczowa nierówność (3.71) i szacunki (3.69)–(3.70) pozwalają uzupełnić dowód twierdzenia 3.3, korzystając z wyników twierdzenia Nevelsona-Hasminskiego [33].

Literatura

[1] Anisimov V.V., *Limit Theorems for Switching Processes and Their Applications*, „Cybernetics", 1978, 14(6), p. 917–929.

[2] Anisimov V.V., *Limit Theorems for Switching Processes*, „Theory of Probability and Mathematical Statistics", 1988, vol. 37, p. 1–5.

[3] Anisimov V.V., *Switching Processes: Averaging Principle, Diffusion Approximation and Applications*, „Acta Applicandae Mathematicae", 1995, vol. 40, p. 95–141.

[4] Anisimov, V.V,. *Averaging Methods for Transient Regimes in Overloading Retrial Queuing Systems*, „Mathematical and Computing Modelling", 1999, vol. 30(3), p. 65–78.

[5] Anisimov V.V., *Switching Processes in Queueing Models*, Wiley-ISTE, London, 2008.

[6] Blankenship G.L., Papanicolaou G.C., *Stability and Control of Stochastic Systems with Wide Band Noise Disturbances*, „SIAM Journal on Applied Mathematics", 1978, vol. 34, p. 437–476.

[7] Боголюбов М.М., *О некоторих статистических методах в математической физике*, Изд. АН УССР, Львов, 1945.

[8] Chabanyuk Ya. M., *Continuous Stochastic Approximation with Semi-Markov Switchings in the Diffusion Approximation Scheme*, „Cybernetics and Systems Analysis", 2007, vol. 43, p. 605–612.

[9] Chabanyuk Ya.M., *Convergence of a Jump Procedure in a Semi-Markov Environment in Diffusion-Approximation Scheme*, „Cybernetics and Systems Analysis", 2007, vol. 43, p. 866–875.

[10] Chabanyuk Ya.M., *Stability of a Dynamical System with Semi-Markov Switchings under Conditions of Diffusion Approximation*, „Ukrainian Mathematical Journal", 2007, vol. 59, p. 1441–1452.

[11] Chabanyuk Y., Nikitin A., Khimka U., *Asymptotic Analyses for Complex Evolutionary Systems with Markov and Semi-Markov Switching Using Approximation Schemes*, Wiley-ISTE, 2020.

[12] Driml M., Nedoma J., *Stochastic approximations for continuous random processes*, [W:] Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 1960, p. 145–158.

[13] Gorun P. P., Chabanyuk Y. M., *Asymptotic Behavior of a Modified Stochastic Optimization Procedure in an Averaging Scheme*, „Cybernetics and Systems Analysis", 2015, vol. 51(6), p. 956–964.

[14] Griego R., Hersh R., *Random Evolutions, Markov Chains, and Systems of Partial Differential Equations*, „Proceedings of the National Academy of Sciences of the United States of America", 1969, vol. 62(2), p. 305–308.

[15] Khimka U.T., Chabanyuk Ya.M., *A Difference Stochastic Optimization Procedure with Impulse Perturbation*, „Cybernetics and Systems Analysis", 2013, vol. 49(5), p. 768–773.

[16] Kiefer E., Wolfowitz J., *Stochastic Estimation of the Maximum of a Regression Function*, „The Annals of Mathematical Statistics", 1952, vol. 23(3) , p. 462–466.

[17] Kinash A., Chabanyuk Ya., Khimka U., *Asymptotic Dissipativity of the Diffusion Process in the Asymptotic Small Diffusion Scheme*, „Journal of Applied Mathematics and Computational Mechanics"', 2015, 14(4), p. 93–103.

[18] Korolyuk V.S., *Stability of Stochastic Systems in the Diffusion-Approximation scheme*, „Ukrainian Mathematical Journal", 1998, vol. 50, p. 36–47.
[19] Korolyuk, V. S., *Problem of Large Deviations for Markov Random Evolutions with Independent Increments in the Scheme of Asymptotically Small*, „Ukrainian Mathematical Journal", 2010, vol. 62, p. 739–747.
[20] Korolyuk, V.S., Chabanyuk, Ya.M., *Stability of a Dynamical System with Semi-Markov Switchings under Conditions of Stability of the Averaged System*, „Ukrainian Mathematical Journal", 2002, vol. 54, p. 239–252.
[21] Korolyuk V.S., Korolyuk V.V., *Stochastic Models of Systems*, Kluwer, Dordrecht. 1999.
[22] Korolyuk V.S., Korolyuk V.V., Limnios N., *Queueing Systems with Semi-Markov Flow in Average and Diffusion Approximation Schemes*, „Methodology and Computing in Applied Probability", 2009, vol. 11, p. 201–209.
[23] Korolyuk V.S., Limnios N., *Evolutionary systems in an asymptotic split state space*, [W:] Recent Advances in Reliability Theory: Methodology, Practice and Inference, N. Limnios M. Nikulin (red.), Birkhauser, Boston, 2000, p. 145–161.
[24] Korolyuk V.S., Limnios N., *Average and Diffusion Approximation for Evolutionary Systems in an Asymptotic Split Phase Space*, „Annals of Applied Probability", 2004, vol. 14(1), p. 489–516.
[25] Korolyuk V.S, Limnios N., *Stochastic Systems in Merging Phase Space*, World Scientific, Singapore, 2005.
[26] Korolyuk V.S., Portenko N., Syta H. (red.), *Skorokhod's Ideas in Probability Theory*, Institute of Mathematics, Kiev, 2000.
[27] Korolyuk V. S., Limnios N., Samoilenko I.V., *Poisson approximation of recurent process with locally independent increments and semi-Markov switching – toward application in reliability*, [W:] Advances in Degradation Modeling, Nikulin M.S., Limnios N., Balakrishnan N., Kahle W., Huber-Carol C. (red.), 2010, p. 105–116.
[28] Korolyuk, V.S., Swishchuk, A.V., *Random Evolutions*, Kluwer Academic Publishers, Dordrecht, 1994.
[29] Korolyuk V. S., Swishchuk A., *Evolution of System in Random Media*, CRC Press, 1995.
[30] Kushner H.J., *Optimality Conditions for the Average Cost per Unit Time Problem with a Diffusion Model*, „SIAM Journal of Control and Optimization", 1978, vol. 16(2), p. 330–346.
[31] Ljung L., Pflug G., Walk H., *Stochastic Approximation and Optimization of Random Systems*, Birkhauser Verlag, Basel, 1992.
[32] Митропольский Ю.О., Самойленко А.М., *Математические проблеми нелинейной механики*, Киев, Наукова думка, 1987.
[33] Невельсон М.Б., Хасьминский Р.З., *Стохастическая аппроксимация и рекурентное оценивание*, Москва, Наука, 1972.
[34] Robbins H., Monro S., *A Stochastic Approximation Method*, „The Annals of Mathematical Statistics", 1951, vol. 22, p. 400–407.
[35] Samoilenko I.V., Chabanyuk Y.M., Nikitin V., Khimka U.T., *Differential Equations with Small Stochastic Additions Under Poisson Approximation Conditions*, „Cybernetics and Systems Analysis", 2017, vol. 53(3), p. 410–416.

[36] Skorokhod A.V., *Asymptotic Methods in the Theory of Stochastic Differential Equations*, Translations of Mathematical Monographs, vol. 78, American Mathematical Society, Providence, 1989.
[37] Stroock D.W., Varadhan S.R.S., *Multidimensional Diffusion Processes*, Springer-Verlag, Berlin, 1979.
[38] Sviridenko M.N., *Martingale Approach to Limit Theorems for Semi–Markov Processes*, „Theory of Probability & Its Applications", 1990, vol. 34(3), p. 540–545.

Izolda Gorgol [1], **Faustyna Smarzewska** [2]

Model Bianconi–Barabásiego w ujęciu teoretycznym

Streszczenie

Rozdział dotyczy problematyki analizy matematycznej sieci złożonych utworzonych na podstawie modelu Bianconi–Barabásiego. Zawarte są w nim informacje dotyczące dynamiki stopni węzłów w sieciach opartych na modelach Barabásiego–Albert i Bianconi–Barabásiego oraz wpływu rozkładu stopni wierzchołków oraz parametrów różnorodności adaptacyjnej na cechy charakteryzujące sieć.

Słowa kluczowe: *modele sieci złożonych, dynamika stopni węzłów, różnorodność adaptacyjna*

Abstract

The chapter deals with the problem of mathematical analysis of complex networks formed on the basis of the Bianconi–Barabási model. It includes information on the dynamics of node degrees in networks based on the Barabási-Albert and Bianconi–Barabási models, as well as the influence of the distribution of vertex degrees and adaptive diversity parameters on the characteristics of the network.

Keywords: *complex network model, degree dynamics, fitness model*

Wstęp

Nauka o sieciach złożonych powstała, aby znaleźć odpowiedź na pytanie, w jaki sposób działają sieci rzeczywiste. Jej początki sięgają 1959 r., kiedy to dwóch węgierskich matematyków – P. Erdős oraz A. Rényi – w artykule [8] opublikowało model, który miał odwzorowywać mechanizm tworzenia się sieci. Zakładali oni, że pomiędzy ustaloną z góry liczbą elementów należy utworzyć połączenia, a najlepszym sposobem na osiągnięcie celu jest zrobienie tego w sposób losowy. Zamiast rozpatrywania jednego układu skupili się oni na zbiorze sieci o zadanej liczbie wierzchołków, które różniły się występowaniem krawędzi. W ten sposób powstała koncepcja grafów losowych, a do powstającej analizy sieci po raz pierwszy wykorzystany został rachunek prawdopodobieństwa, co pozwoliło na uwzględnienie w badaniach przypadkowości [9].

Kiedy w drugiej połowie XX wieku pojawił się internet, przed badaczami pojawiła się możliwość nie tylko zbierania i analizowania, ale również wymieniania się dużą ilością danych dotyczących sieci występujących w rzeczywistości. Wówczas węgierscy fizycy A.L. Barabási i R. Albert dokonali ciekawego odkrycia. Przyjrzeli

[1] Katedra Matematyki Stosowanej, Politechnika Lubelska
[2] Koło Naukowe „Kwaternion", Politechnika Lubelska

się oni kilku sieciom rzeczywistym i zauważyli, że w sieciach tych występują „huby", tj. węzły o znacznie większej liczbie połączeń niż pozostałe, co nie pokrywało się ze znanym dotąd modelem grafu losowego. W wyniku tych obserwacji w 1999 r. w artykule [4] opublikowano koncepcję sieci bezskalowych. Pojawił się tam również model Barabásiego–Albert, w którym uwzględnione zostały mechanizmy ewolucji sieci niebrane wcześniej pod uwagę w koncepcji Erdősa i Rényiego.

Jednakże w miarę upływu czasu zaczęto zauważać niedoskonałości zaproponowanego rozwiązania. Model Barabásiego–Albert zakłada, że bardziej prawdopodobne jest przyłączanie się nowych węzłów do tych już posiadających wiele połączeń – innymi słowy, nowe elementy dołączające do struktury nie mają szans w porównaniu ze swoimi starszymi odpowiednikami. Tłumaczyło to, w jaki sposób popularność zyskały wyszukiwarki internetowe Alta Vista oraz Inktomi, które w sieci www pojawiły się odpowiednio w 1995 r. oraz 1996 r., oraz dlaczego udało im się zdominować konkurencję. Jednak kiedy w 1998 r. do sieci dołączona została kolejna wyszukiwarka, nie tylko w stosunkowo krótkim czasie dorównała popularnością swoim odpowiednikom, ale dzięki niezwykle szybkiemu zdobywaniu nowych połączeń już w 2000 r. stała się największym „hubem" sieci www [2]. Stronę tę znamy do dziś pod nazwą Google. Wobec takich obserwacji kontynuowano badania nad własnościami sieci rzeczywistych, aby móc ulepszyć model Barabásiego–Albert, aż w 2001 r. włoska matematyczka G. Bianconi i A.L. Barabási opublikowali artykuł [5] wprowadzający do nauki o sieciach złożonych model Bianconi–Barabásiego.

Dzięki zastosowanej przez nich modyfikacji, w nowym modelu prawdopodobieństwo przyłączenia do węzła z sieci kolejnego wierzchołka nie zależy już tylko od tego, jak wiele połączeń posiada dany węzeł, ale też od tego, jaka jest jego jakość z punktu widzenia całej sieci. Innymi słowy w modelu uwzględniony został dodatkowo parametr reprezentujący osobliwą własność każdego z wierzchołków, która sprawia, że jest on w stanie tworzyć nowe połączenia. Doskonałym przykładem jest tu kontynuacja historii związanej z największym „hubem" sieci www. Otóż 6 lat po wprowadzeniu Google i 4 lata po uzyskaniu przez tę wyszukiwarkę tytułu największego „hubu", tj. w 2004 r., do sieci została dołączona strona internetowa firmy Facebook. W tym czasie istniało już ponad 50 mln stron internetowych. Mimo to w 2011 r. właśnie ten portal społecznościowy został uznany za węzeł o największej liczbie połączeń [2]. Warto zauważyć, że w sieci w tym czasie znajdowało się już ok. 346 mln stron [15]. Okazuje się zatem, że nie czas obecności węzłów w sieci decydował o ostatecznie odniesionym sukcesie. Kluczem stał się satysfakcjonujący i uzależniający charakter usług, jakie oferował Facebook. Pomimo pojawienia się w sieci później od pozostałych, był on „lepiej dopasowany" do potrzeb użytkownika, dzięki czemu zyskał światową popularność i stał się sztandarowym przykładem zjawiska *fit-get-richer*.

1. Bezskalowość

Jednym z ważniejszych pojęć wykorzystywanym do określania charakterystycznych cech rozważanej sieci jest rozkład stopni wierzchołków. Informuje on o prawdopodobieństwie, z jakim losowo wybrany z sieci wierzchołek będzie miał stopień k.

Definicja 1.1. [13] Rozkładem stopni wierzchołków nazywamy rozkład prawdopodobieństwa dany wzorem

$$P(k) = \frac{|\{i : i \in V \wedge k_i = k\}|}{N},$$

gdzie $|\{i : i \in V \wedge k_i = k\}|$ oznacza liczbę wierzchołków grafu, których stopień jest równy k.

Ponieważ $P(k)$ jest prawdopodobieństwem, oznacza to, że

$$\sum_{k=0}^{k_{max}} P(k) = 1, \tag{1.1}$$

gdzie k_{max} oznacza największy stopień wierzchołka uzyskiwany w sieci.

Średni stopień wierzchołka możemy wówczas określić jako wartość oczekiwaną daną wzorem

$$\langle k \rangle = \sum_{k=0}^{k_{max}} k \cdot P(k). \tag{1.2}$$

Pod koniec XX wieku Barabási i Albert odkryli, że w sieciach rzeczywistych występuje potęgowy rozkład stopni wierzchołków, który można zapisać jako

$$P(k) = \frac{C}{k^{\alpha}}, \tag{1.3}$$

gdzie C oznacza pewną stałą, a $\alpha > 0$ to wykładnik charakterystyczny rozkładu. Sieć utworzoną w taki sposób cechuje występowanie zjawiska bezskalowości, w którym zauważalny jest brak powtarzalności, a także duża zmienność otrzymywanych wyników [9].

Przyjmijmy, że rozpatrujemy zmienną losową X. Jeżeli przyjmowane przez nią wartości należą do zbioru przeliczalnego, nazywamy ją zmienną dyskretną. Jeśli natomiast wartości te będą należały do zbioru nieprzeliczalnego – X będzie zmienną ciągłą. Załóżmy, że prawdopodobieństwo wystąpienia danej wartości zmiennej losowej w rozkładzie potęgowym jest dane przez

$$P(x) = P[X = x] = Cx^{-\alpha}. \tag{1.4}$$

Warunek unormowania mówi o tym, że suma prawdopodobieństw realizacji zmiennej losowej jest równa 1.

W przypadku dyskretnej zmiennej losowej potęgowy rozkład prawdopodobieństwa można zapisać jako

$$P(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})}, \tag{1.5}$$

w którym $\frac{1}{\zeta(\alpha,x_{min})}$ jest stałą C wyznaczoną z dyskretnego warunku unormowania, określonego dla wykładnika $\alpha > 1$ następująco

$$\sum_{x=x_{min}}^{\infty} P(x) = C \cdot \sum_{x=x_{min}}^{\infty} x^{-\alpha} = C \cdot \sum_{n=0}^{\infty} (n + x_{min})^{-\alpha} = C \cdot \zeta(\alpha, x_{min}) = 1, \tag{1.6}$$

gdzie x_{min} to najmniejsza wartość przyjmowana przez zmienną X, natomiast $\zeta(\alpha, x_{min})$ jest funkcją zeta Hurwitza.

W przypadku ciągłej zmiennej losowej prawdopodobieństwo $P(x)$ jest wyrażone wzorem

$$P(x) = \frac{\alpha - 1}{x_{min}} \cdot \left(\frac{x}{x_{min}} \right)^{-\alpha}, \tag{1.7}$$

gdzie postać C dla wykładnika $\alpha > 1$ wynika z warunku unormowania o następującej postaci

$$\int_{x_{min}}^{\infty} P(x)dx = C \cdot \int_{x_{min}}^{\infty} x^{-\alpha} dx = C \cdot \left[\frac{x^{-\alpha+1}}{-\alpha+1} \right]_{x_{min}}^{\infty} = C \cdot \frac{x_{min} \cdot (x_{min})^{-\alpha}}{\alpha - 1} = 1. \tag{1.8}$$

Warunek unormowania nie będzie spełniony, gdy wykładnik $\alpha \leqslant 1$. Ze względu na trudności obliczeniowe występujące w przypadku sumowania funkcji potęgowych, często stosuje się przybliżenie oparte na zastąpieniu sumy całkowaniem. Na podstawie reguły sumacyjnej Poissona postaci

$$\sum_{x=0}^{\infty} F(x) - \int_{0}^{\infty} F(x)dx = \frac{1}{2} F(0) + 2 \sum_{n=1}^{\infty} \int_{0}^{\infty} F(x) \cos(2n\pi x) dx \tag{1.9}$$

uzyskujemy oszacowanie błędu wykonanego przybliżenia.

Wartość oczekiwana zmiennej losowej X o rozkładzie potęgowym danym wzorem (1.7) wynosi

$$\langle x \rangle = \int_{x_{min}}^{\infty} xP(x)dx = C \cdot \int_{x_{min}}^{\infty} x^{-\alpha+1} dx. \tag{1.10}$$

Gdy $\alpha > 2$, wartość średnia (1.10) rozkładów prawdopodobieństwa o charakterze

potęgowym jest skończona. Korzystając z C otrzymanego w (1.8), można ją zapisać jako

$$\langle x\rangle = C\cdot\left[\frac{x^{-\alpha+2}}{-\alpha+2}\right]_{x_{min}}^{\infty} = C\cdot\frac{-x_{min}^{-\alpha+2}}{-\alpha+2} = \frac{\alpha-1}{x_{min}^{-\alpha+1}}\cdot\frac{x_{min}^{-\alpha+2}}{\alpha-2} = \frac{\alpha-1}{\alpha-2}\cdot x_{min}. \quad (1.11)$$

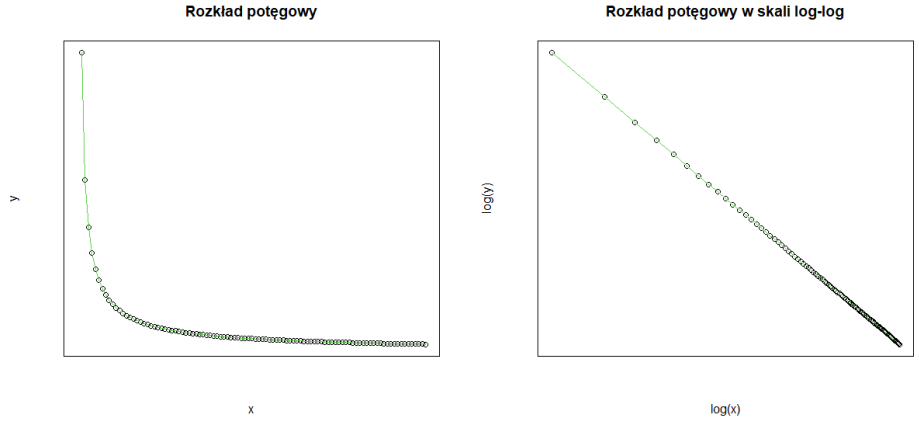
Gdy wykładnik charakterystyczny $\alpha \leqslant 2$, wartość oczekiwana jest nieskończona. Z kolei kiedy $\alpha > 3$, całka określona dla drugiego momentu jest zbieżna i wyrażona wzorem

$$\langle x^2\rangle = \int_{x_{min}}^{\infty} x^2 P(x)dx = C\cdot\int_{x_{min}}^{\infty} x^{-\alpha+2}dx = \frac{\alpha-1}{\alpha-3}\cdot x_{min}^2. \quad (1.12)$$

Natomiast, gdy parametr $\alpha \leqslant 3$, całka ta jest rozbieżna, przez co zarówno wariancja σ^2 dana wzorem $\langle x^2\rangle - \langle x\rangle^2$, jak też odchylenie standardowe $\sigma = \sqrt{\langle x^2\rangle - \langle x\rangle^2}$ są nieskończone. W takiej sytuacji, nawet jeżeli wartość średnia jest dobrze określona, wnioskowanie na jej podstawie zostaje obarczone ogromnym błędem, ponieważ posiada nieskończoną niepewność oszacowania. Po uogólnieniu okazuje się, że momenty $\langle x^m\rangle$ rozkładów potęgowych są rozbieżne, gdy spełniają zależność $\alpha \leqslant m+1$.

Podsumowując, gdy w rozkładach potęgowych wykładnik charakterystyczny wynosi $\alpha \leqslant 2$, bezskalowość objawia się w niemożliwej do policzenia wartości średniej $\langle x\rangle$. Gdy $\alpha \leqslant 3$, zjawisko to jest zauważalne w nieskończonej niepewności pomiarowej σ. Natomiast gdy α przyjmuje większe wartości, niemierzalność może dotyczyć niemożliwości określenia błędu niepewności. Z tego powodu układy takie nazywamy niemierzalnymi lub też bezskalowymi (ang. *scale-free*). Bezskalowość występuje tylko wtedy, gdy układ podlega prawom potęgowym, dlatego też pojęcie rozkładów bezskalowych wykorzystywane jest do określania rozkładów prawdopodobieństwa o charakterze potęgowym.

W układach rzeczywistych bezskalowość jest zauważalna poprzez występowanie zdarzeń ekstremalnych wśród tych mających małe oraz średnie rozmiary. Przykładowo, obok ludzi o przeciętnych dochodach, istnieją osoby o zarobkach porównywalnych z dochodami państw. Wykres rozkładu potęgowego ilustruje wspomnianą zależność. Wykresy te po przedstawieniu w skali podwójnie logarytmicznej przyjmują formę linii prostej. Możemy powiedzieć, że w sieciach charakteryzowanych przez potęgowy rozkład stopni wierzchołków zauważalny jest tzw. efekt sieci bezskalowej. Oznacza to, że występuje w nich wiele wierzchołków o małej liczbie połączeń, a zarazem niewiele „hubów", tj. węzłów o bardzo dużym stopniu. Podczas analizy sieci rzeczywistych korzysta się z przedstawiania rozkładu stopni węzłów w postaci wykresu w skali podwójnie logarytmicznej, by lepiej uwidocznić ich potęgowy charakter.



Rysunek 1. Wykres przedstawiający kształt rozkładu potęgowego w skali tradycyjnej oraz podwójnie logarytmicznej
Źródło: Opracowanie własne

2. Model Barabásiego–Albert

Model Barabásiego–Albert (BA) stanowi podstawę modelu Bianconi–Barabásiego (BB), będącego tematem pracy. Procedury konstrukcyjne obu modeli są do siebie bardzo podobne, a rozumowanie przeprowadzane w metodzie czasu ciągłego modelu BA, dzięki występującej analogii, wykorzystywane jest podczas określenia dynamiki stopni w modelu BB. Z tego względu model Barabásiego–Albert został szczegółowo opisany w pracy. Aby pełniej zrozumieć mechanizmy występujące w tym podstawowym modelu związanym z sieciami ewoluującymi, warto najpierw zapoznać się z głównymi cechami tej klasy modeli.

Złożone systemy rzeczywiste charakteryzują się tym, że ewoluują w miarę upływu czasu. Oznacza to, że stale zmieniają swoją strukturę poprzez dodawanie nowych węzłów i połączeń lub usuwanie ich. Zmiany zachodzące na poziomie pojedynczych węzłów mogą wydawać się mało znaczące, jednak wpływają na topologię całego układu. Dlatego też poznanie dynamiki przemian występujących w sieci jest tak ważne [5].

W dziedzinie nauki o sieciach złożonych występują trzy klasy modeli: modele statyczne (ang. *static models*), generatywne (ang. *generative models*) i modele sieci ewoluujących (ang. *evolving network models*) [2]. Przykładem modelu statycznego jest model Erdősa-Rényiego, w którym pomiędzy stałą liczbą wierzchołków w sposób losowy tworzone są połączenia, a otrzymanie każdego z możliwych grafów jest równie prawdopodobne [8]. W modelu tym zakłada się, że każdy wierzchołek

łączony jest z podobną liczbą innych wierzchołków oraz nie występują w nim węzły o stopniu znacząco odbiegającym od średniego stopnia wierzchołków [9]. Korzystając z niego, można sprawdzić, czy daną cechę sieci złożonej jesteśmy w stanie wyjaśnić za pomocą czysto losowego przyłączania się węzłów. Z kolei modele Barabásiego–Albert i Bianconi–Barabásiego należą do modeli sieci ewoluujących. Klasę tę charakteryzuje zmieniająca się w czasie topologia oraz prawdopodobieństwo $P(k)$ zależne od procesów przyczyniających się do wzrostu sieci. Dlatego też, jeśli chcemy poznać nie tylko rolę sieci w kształtowaniu się danych zjawisk, ale też samo źródło powstania danej cechy występującej w sieci, należy przyjrzeć się modelom uwzględniającym te procesy [2].

Albert-László Barabási oraz Réka Albert w swoim artykule [4] z 1999 r. stwierdzili, iż do otrzymania potęgowego rozkładu stopni wierzchołków potrzeba działania dwóch mechanizmów, jakimi są:

— wzrost sieci (ang. *growth*) – tzn. nieustanne poszerzanie się sieci poprzez dodawanie nowych węzłów,
— liniowa reguła preferencyjnego dołączania wierzchołków (ang. *linear preferential attachment*) – tj. zasada, w myśl której węzły obecne w sieci przyłączają nowe wierzchołki z prawdopodobieństwem liniowo proporcjonalnym do swojego stopnia, a liczba powstających wówczas krawędzi rośnie proporcjonalnie do liczby węzłów.

Takie podejście zakłada, że nowe wierzchołki z większym prawdopodobieństwem będą przyłączane do węzłów ze stosunkowo dużym stopniem, a więc do „hubów". Obrazuje to następująca sytuacja. Gdy nowa osoba dołącza do danej grupy, z większym prawdopodobieństwem nawiąże znajomość z kimś rozpoznawalnym, tzn. osobą posiadającą wielu znajomych, niż z osobą trzymającą się na uboczu. Model ten tłumaczy też zjawisko „przewagi pierwszego gracza" (ang. *first-mover advantage*), a więc sytuację, gdy o sukcesie węzła w sieci decyduje to, jak wcześnie się w niej pojawił [3]. Reguła preferencyjnego dołączania wierzchołków pozwala również na zobrazowanie zjawiska „bogacenia się najbogatszych" (ang. *rich-get-richer*), gdzie wierzchołki o stopniach większych niż stopnie pozostałych węzłów mają większe szanse na nawiązanie nowego połączenia. Wspomnianą regułę możemy zaobserwować w sieciach rzeczywistych, takich jak np. internet oraz sieć cytowań [9], gdzie do routerów z przyłączonymi wieloma urządzeniami częściej dołączane są kolejne urządzenia, a popularne prace zyskują jeszcze większą popularność.

2.1. Procedura konstrukcyjna

Proces konstrukcji sieci Barabásiego–Albert składa się z następujących po sobie etapów:

— W chwili $t_0 = 0$ sieć jest grafem pełnym o $n_0 \geqslant 1$ wierzchołkach.

— W kroku czasowym $t = 1$ do sieci włączony zostaje jeden wierzchołek o indeksie $n = n_0 + t$.
— Dodany węzeł łączy się z $m \leqslant n_0$ już istniejącymi w sieci węzłami zgodnie z liniową regułą preferencyjnego dołączania wierzchołków. Oznacza to, że w chwili t nowy wierzchołek zostanie przyłączony do i-tego wierzchołka z sieci z prawdopodobieństwem

$$\Pi(k_{i,t-1}) = \frac{k_{i,t-1}}{\sum_{j=1}^{n-1} k_{j,t-1}} = \frac{k_{i,t-1}}{2E_{t-1}}, \tag{2.1}$$

gdzie
— $k_{i,t-1}$ – stopień i-tego wierzchołka należącego do sieci w chwili $t-1$ poprzedzającej dodanie nowego węzła,
— E_{t-1} – liczba wszystkich krawędzi w sieci w chwili $t-1$.
— Proces jest powtarzany dla kolejnych kroków czasowych $t = 2, 3, ...$, a zakończony może zostać w dowolnej chwili t.

Nowy wierzchołek musi dołączyć się do już istniejących (nie może zostać wierzchołkiem izolowanym lub przyłączyć się do węzłów spoza sieci), stąd suma prawdopodobieństw, że w chwili t zostanie on przyłączony do węzłów z sieci, wynosi 1

$$\sum_{i=1}^{n-1} \Pi(k_{i,t-1}) = 1. \tag{2.2}$$

W przypadku bardzo dużych sieci, gdzie $t \gg 1$, stosowane są przybliżenia liczby krawędzi $|E_t|$ i liczby wierzchołków N_t w sieci w chwili t, pomijające występujące początkowo w grafie połączenia

$$|E_t| = mt + \binom{n_0}{2} \approx mt, \tag{2.3}$$

$$N_t = t + n_0 \approx t. \tag{2.4}$$

Za każdym razem, gdy dodawany jest do sieci nowy wierzchołek, tworzonych jest m nowych połączeń, co w grafie nieskierowanym powoduje zwiększenie sumy stopni wierzchołków o $2m$. Przemnożenie tego efektu z przybliżonym rozmiarem sieci 2.4 dla $t \gg 1$ sprawia, że sumę stopni wierzchołków można wówczas zapisać jako

$$\sum_{i=1}^{N_t} k_{i,t} = 2|E_t| \approx 2mt. \tag{2.5}$$

Pozwala to na zapisanie reguły preferencyjnego przyłączania wierzchołków przy dodawaniu nowego węzła w chwili $t+1$ w następujący sposób

$$\Pi(k_{i,t}) = \frac{k_{i,t}}{\sum\limits_{j=1}^{N_t} k_{j,t}} \approx \frac{k_{i,t}}{2mt}. \tag{2.6}$$

Symulacje przeprowadzone przez Barabásiego i Albert w 1999 r. wskazują na to, że zjawisko bezskalowości, omówione w poprzednim podrozdziale, może pojawić się w układzie tylko wtedy, gdy $\Pi(k)$ jest wprost proporcjonalne do k^{α}, gdzie $\alpha = 1$ – innymi słowy tylko wtedy, gdy reguła preferencyjnego przyłączania wierzchołków jest liniowa [4].

2.2. Metoda czasu ciągłego

Podrozdział został opracowany na podstawie pozycji [3, 9] literatury. Metoda czasu ciągłego (ang. *continuum theory*) wykorzystywana jest w modelach Barabásiego–Albert oraz Bianconi–Barabásiego do uzyskania uśrednionego rozkładu stopni wierzchołków. Ze względu na analogie występujące pomiędzy oboma modelami, zostało tu opisane rozumowanie przeprowadzane podczas rozpatrywania modelu BA. W rozważanej metodzie zakłada się, że czas t oraz stopnie węzłów $k_i(t)$ zmieniają się ciągle, tzn. nie ulegają zmianom skokowym. Prawdopodobieństwo $\Pi(k_i)$ dołączenia nowego wierzchołka do i-tego węzła sieci może być interpretowane jako tempo zmiany stopnia $k_i(t)$ w czasie [1]. Przez to, że wartości stopni węzłów ulegają zmianom o drobne ułamki, w metodzie tej nie mówimy o prawdziwych wartościach stopni, a zamiast tego posługujemy się średnimi stopniami wierzchołków, zmieniającymi się w czasie.

Przez krok czasowy będziemy rozumieli Δt. Szybkość, z jaką zmianie będzie ulegał średni stopień i-tego węzła, określamy jako $\frac{\Delta k_i}{\Delta t} \approx \frac{\partial k_i}{\partial t}$. Zmiana ta jest uzależniona od tego, jak wiele połączeń będzie tworzonych z węzłem i-tym. Warto zaznaczyć, że w tej metodzie uwzględniono, iż możliwe jest powstanie więcej niż jednego połączenia do danego węzła, gdy liczba powstających w każdym kroku czasowym krawędzi m jest większa od 1, czyli są dopuszczalne krawędzie wielokrotne. W dalszym rozumowaniu wykorzystywany jest rozkład dwumianowy

$$p(k,n) = \binom{n}{k} p^k (1-p)^k \tag{2.7}$$

o prawdopodobieństwach: sukcesu p, porażki $(1-p)$ oraz o k sukcesach i $n-k$ porażkach wśród n prób. Jego wartość oczekiwana wyrażana jest wzorem

$$\langle k \rangle = \sum_{k=0}^{n} k \cdot p(k,n) = np. \tag{2.8}$$

Korzystając z rozkładu dwumianowego, określamy, że prawdopodobieństwo uzyskania przez i-ty wierzchołek liczby krawędzi równej l spośród m tworzonych połączeń wynosi

$$p_i(l,m) = \binom{m}{l} \Pi(k_i)^l (1-\Pi(k_i))^{m-l}, \tag{2.9}$$

gdzie $\Pi(k_i)$ oznacza prawdopodobieństwo sukcesu, tj. połączenia krawędzi ze wspomnianym węzłem, a jednocześnie jest opisaną w (2.1) regułą preferencyjnego dołączania wierzchołków. Zmianę średniego stopnia i-tego wierzchołka w czasie możemy zapisać jako średnią wartość liczby przyłączonych krawędzi l spośród wszystkich tworzonych połączeń m. Korzystając z przybliżenia (2.6) dla modelu BA, otrzymujemy

$$\frac{\partial k_i}{\partial t} = \langle l \rangle = \sum_{l=0}^{m} l \cdot p(l,m) = m \cdot \Pi(k_i) = m \cdot \frac{k_i}{\sum_{j=1}^{N} k_j} \approx m \cdot \frac{k_i}{2mt} = \frac{k_i}{2t} = \frac{1}{2} \cdot \frac{k_i}{t}. \tag{2.10}$$

Rozwiązujemy zatem równanie różniczkowe I rzędu o zmiennych rozdzielonych z (2.10), pamiętając, że stopień wierzchołka jest funkcją czasu $k_i = k_i(t)$. Dla uogólnienia zapisu obecny w (2.10) ułamek $\frac{1}{2}$ oznaczymy za pomocą symbolu β.

$$\begin{aligned}
\frac{\partial k_i}{\partial t} &= \beta \cdot \frac{k_i}{t}, \\
\frac{\partial k_i}{k_i} &= \beta \cdot \frac{\partial t}{t}, \\
\int \frac{1}{k_i} dk_i &= \beta \int \frac{1}{t} dt, \\
\ln|k_i| &= \beta \ln|t| + C,\ k_i > 0,\ t > 0, \\
k_i &= \exp(\ln t^{\beta} + C), \\
k_i &= t^{\beta} \cdot D,\ D = e^{C}.
\end{aligned} \tag{2.11}$$

Średni stopień węzła i w chwili jego dołączenia do sieci t_i jest równy liczbie połączeń, które wtedy utworzył, stąd określamy warunek początkowy $k_i(t_i) = m$. Po jego

uwzględnieniu otrzymujemy stałą D.

$$\begin{aligned} &k_i = t^{\beta} \cdot D,\ k_i(t_i) = m, \\ &m = t_i^{\beta} \cdot D, \\ &D = \frac{m}{t_i^{\beta}}. \end{aligned} \tag{2.12}$$

Podstawiając obliczoną stałą D do otrzymanego w (2.11) rozwiązania ogólnego, otrzymujemy rozwiązanie szczególne – informację o tym, jak można określić średni stopień i-tego węzła w zależności od innych zmiennych. Jest to

$$k_i(t) = m \cdot \left(\frac{t}{t_i}\right)^{\beta}, \tag{2.13}$$

gdzie β jest dynamicznym wykładnikiem (ang. *dynamic exponent*) czasowej ewolucji węzła. Na początku przyjęliśmy, że $\beta = \frac{1}{2}$, zatem dla modelu BA możemy użyć zapisu

$$k_i(t) = m \cdot \sqrt{\frac{t}{t_i}}. \tag{2.14}$$

W metodzie czasu ciągłego wykorzystywana jest aproksymacja ciągła (ang. *continuum approximation*). Aby na jej podstawie wyznaczyć przybliżony wzór zależności występującej w rozkładzie stopni wierzchołków, obliczany jest skumulowany rozkład (ang. *cumulative distribution function*) $P^c(k)$

$$P^c(k) = P(k_i(t) \leqslant k) = 1 - P(k_i(t) > k), \tag{2.15}$$

który zawiera informację o tym, z jakim prawdopodobieństwem średni stopień losowo wybranego z sieci węzła będzie mniejszy bądź równy k. Formalnie używane jest pojęcie dystrybuanty, natomiast w naukach empirycznych dla podkreślenia, że zmienna losowa jest dyskretna, stosowane jest pojęcie rozkładu skumulowanego. Najpierw wyznaczamy liczbę węzłów, których średni stopień $k_i(t) > k$. Korzystając z (2.13), możemy zapisać

$$\begin{aligned} &m \cdot \left(\frac{t}{t_i}\right)^{\beta} > k, \\ &\frac{m}{k} \cdot t^{\beta} > t_i^{\beta}, \end{aligned} \tag{2.16}$$

wobec czego otrzymujemy

$$\left(\frac{m}{k}\right)^{\frac{1}{\beta}} \cdot t > t_i. \tag{2.17}$$

Zatem, aby średni stopień i-tego wierzchołka $k_i(t)$ był większy od k, czas jego dołączenia do sieci t_i musi być mniejszy niż $t \cdot \left(\frac{m}{k}\right)^{\frac{1}{\beta}}$. W procedurze konstrukcyjnej modelu BA zakłada się, że do sieci dołączane jest po jednym węźle w równych odstępach czasu, dlatego też węzłów spełniających podaną nierówność będzie tyle, ile ich możliwych czasów t_i dołączenia do układu. Z tego powodu Barabási w swojej książce [3] wskazuje, że można przyjąć, iż $t\left(\frac{m}{k}\right)^{\frac{1}{\beta}}$ oznacza liczbę węzłów o średnim stopniu większym niż k. Wówczas, biorąc pod uwagę przybliżenie (2.4) liczby węzłów sieci przez t w granicy dużych N, skumulowany rozkład można zapisać jako

$$\begin{aligned} P^c(k) &= P(k_i(t) \leqslant k) = 1 - P(k_i(t) > k) \approx \\ &\approx 1 - \frac{t \cdot \left(\frac{m}{k}\right)^{\frac{1}{\beta}}}{N} = 1 - \frac{t \cdot \left(\frac{m}{k}\right)^{\frac{1}{\beta}}}{t+n_0} \approx 1 - \frac{t \cdot \left(\frac{m}{k}\right)^{\frac{1}{\beta}}}{t} = 1 - \left(\frac{m}{k}\right)^{\frac{1}{\beta}}. \end{aligned} \tag{2.18}$$

Pochodna z uzyskanego wyrażenia (2.18) pozwala na uzyskanie uśrednionego rozkładu stopni wierzchołków $P(k)$

$$P(k) = \frac{\partial P^c(k)}{\partial k} = -m^{\frac{1}{\beta}} \cdot \left(-\frac{1}{\beta}\right) \cdot k^{-\frac{1}{\beta}-1} = \frac{1}{\beta} \cdot \frac{m^{\frac{1}{\beta}}}{k^{\frac{1}{\beta}+1}} = \frac{1}{\beta} \cdot \frac{m^{\frac{1}{\beta}}}{k^{\alpha}}, \tag{2.19}$$

gdzie β jest wykładnikiem dynamicznym ewolucji węzła w czasie, natomiast

$$\alpha = \frac{1}{\beta} + 1 \tag{2.20}$$

jest wykładnikiem charakterystycznym potęgowego rozkładu stopni wierzchołków (ang. *connectivity exponent*). Dla modelu BA wykładnik $\beta = \frac{1}{2}$, stąd otrzymujemy

$$P(k) = \frac{\partial P^c(k)}{\partial k} = \frac{2m^2}{k^3}. \tag{2.21}$$

Do uzyskania dokładnego rozkładu $P(k)$ wykorzystywana jest ścisła metoda równania master. W wyniku jej zastosowania uzyskuje się precyzyjną informację o rozkładzie stopni węzłów, którą sformułowano następująco: w sieci BA, gdzie m oznacza liczbę połączeń tworzonych w każdym kroku czasowym, prawdopodobieństwo tego, że dany węzeł będzie posiadał stopień $k \geqslant m$ wynosi

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \propto k^{-3}. \tag{2.22}$$

2.3. Wybrane własności

W sieci utworzonej na podstawie modelu Barabásiego–Albert:
— średni stopień wierzchołka (wartość oczekiwana krotności węzła) wynosi

$$\langle k \rangle = \sum_{k=m}^{\infty} k \cdot P(k) = 2m(m+1) \sum_{k=m}^{\infty} \frac{k}{k(k+1)(k+2)} = 2m, \tag{2.23}$$

— średnica sieci przy wystarczająco dużej liczbie węzłów N oraz $m > 1$ dołączanych w każdym kroku krawędziach jest równa

$$d(N) = \frac{\ln(N)}{\ln(\ln(N))}, \tag{2.24}$$

— średnia długość ścieżki jest wyrażona wzorem

$$\langle p(N,m,\gamma) \rangle = \frac{\ln(N) - \ln(\frac{m}{2} - 1 - \gamma)}{\ln(\ln(N)) + \ln(\frac{m}{2} + 1,5)}, \tag{2.25}$$

gdzie $\gamma \approx 0,5772$ oznacza stałą Eulera.

Model BA wykorzystywany jest do opisu potęgowego rozkładu stopni wierzchołków w sieciach ewoluujących. Stanowi on model minimalny, dlatego tylko część własności występujących w rzeczywistości zostaje przez niego przedstawiona. Przykładowo wykładnik charakterystyczny α opisujący $P(k) \propto k^{-\alpha}$ w sieciach rzeczywistych najczęściej przyjmuje wartości z przedziału $(2,3)$, podczas gdy w modelu BA $\alpha = 3$ [9]. Kolejną z występujących niedoskonałości jest przyjęcie założenia, że wszystkie węzły włączane do sieci posiadają takie same właściwości, co powoduje, że prawdopodobieństwo przyłączenia nowego węzła do wierzchołków o tym samym stopniu nie jest różnicowane pod kątem ich odmienności. Procedura konstrukcyjna modelu BA uważana jest za stosunkowo prostą, co pozwala na rozszerzanie jej o kolejne reguły i wprowadzanie modyfikacji. Jednym z przykładów jej ulepszenia jest model Bianconi–Barabásiego, omówiony szerzej w dalszej części pracy.

3. Model Bianconi–Barabásiego

W modelu BA średni stopień wierzchołka sieci $k_i(t)$ zależy wprost proporcjonalnie od pierwiastka czasu t jego obecności w układzie zgodnie z zależnością (2.13). Oznacza to, że węzłom, które dołączono do sieci najwcześniej, zostanie przyporządkowanych najwięcej połączeń. Wynika z tego, że stopień wierzchołków przyłączonych do sieci później nigdy nie przewyższy stopnia węzłów najstarszych. W ten sposób o sukcesie danego elementu decyduje jedynie jego czas pojawienia się

w układzie. Kłóci się to z dokonanymi przez Bianconi i Barabásiego obserwacjami sieci występującymi w rzeczywistości. Istnieją bowiem nastolatkowie posiadający o wiele więcej znajomych niż osoby starsze. Zdarzają się też sytuacje, gdy nowe firmy i strony internetowe szybko zyskują popularność, podczas gdy inne, występujące długo w danym środowisku ostatecznie nie wybijają się ponad konkurencję. Przykładem jest wspomniany już wcześniej sukces platformy Facebook, która została dodana do sieci stosunkowo niedawno, co nie przeszkodziło jednak w zrzeszeniu przez nią znacznie większej liczby użytkowników niż powstałe przed nią serwisy społecznościowe. W ten sposób zauważono, iż cechy danego elementu układu są również ważne i mogą przyczynić się do jego rozwoju. Rozważaną koncepcję zawiera model Bianconi–Barabásiego. Stanowi on udoskonalenie modelu Barabásiego–Albert. Uwzględniona w nim została tzw. różnorodność adaptacyjna (ang. *fitness*) wierzchołków, będąca jakością każdego węzła z punktu widzenia sieci. Pozwala ona na skalowanie stopnia węzła, do którego jest przyporządkowana, dzięki czemu umożliwia dokładniejsze opisanie ewolucji układów rzeczywistych [6].

Ze względu na to, że model BB stanowi modyfikację modelu BA, on również jest oparty na dwóch mechanizmach, jakimi są wzrost sieci oraz reguła preferencyjnego przyłączania wierzchołków.

Wzrost sieci wiąże się z dodawaniem do grafu w każdym kroku czasowym nowego wierzchołka, który tworzy m połączeń z węzłami obecnymi w sieci zgodnie z procedurą konstrukcyjną modelu BA, opisaną w podrozdziale 2.1. Modyfikację stanowi fakt, że każdemu wierzchołkowi w momencie dołączania do sieci przypisywany jest na stałe parametr różnorodności adaptacyjnej η (w skrócie nazywany parametrem adaptacyjnym). Jest on utożsamiany ze specjalnymi właściwościami węzła, które wpływają na tworzenie pomiędzy nim a innymi węzłami nowych połączeń. Parametr ten jest zmienną losową otrzymaną z wybranego rozkładu gęstości prawdopodobieństwa $\rho(\eta)$.

W modelu BB reguła preferencyjnego przyłączania wierzchołków, znana z modelu BA, również wzbogacona jest o parametr różnorodności adaptacyjnej η. Wówczas prawdopodobieństwo dołączenia nowego wierzchołka do i-tego węzła sieci w chwili $t+1$ wyrażone jest wzorem

$$\Pi(k_{i,t}, \eta_i) = \frac{k_{i,t} \cdot \eta_i}{\sum\limits_{j=1}^{N_t} k_{j,t} \cdot \eta_j}, \tag{3.1}$$

gdzie

— $k_{i,t}$ to stopień i-tego wierzchołka w chwili t przed dodaniem nowego węzła do sieci,

— η_i oznacza niezmienny w czasie parametr adaptacyjny i-tego węzła,

— N_t jest liczbą wszystkich wierzchołków w sieci przed dodaniem nowego węzła. Zmienną $k_{i,t}$ mówiącą o stopniu danego wierzchołka możemy interpretować jako jego widoczność w sieci w chwili t. Im widoczniejszy węzeł, tym większe prawdopodobieństwo zdobycia przez niego nowych krawędzi. Wpływ parametru adaptacyjnego η_i najlepiej zauważalny jest w sytuacji, gdy w sieci posiadamy dwa wierzchołki o tym samym stopniu. Wówczas bardziej prawdopodobne jest, że nowy węzeł przyłączy się do tego z nich, który posiada większą wartość parametru różnorodności adaptacyjnej η.

3.1. Dynamika stopni

Podrozdział został opracowany na podstawie pozycji [2, 5] literatury. Aby dowiedzieć się, w jaki sposób zmianom podlegają stopnie wierzchołków sieci, musimy poznać dynamikę ich ewolucji. W tym celu wykorzystujemy opisaną poprzednio metodę czasu ciągłego (ang. *continuum theory*). Najpierw zapisujemy, w jaki sposób średni stopień $k_i(t)$ i-tego wierzchołka zmienia się w czasie na podobieństwo wzoru występującego w (2.10), a następnie, korzystając ze wzoru reguły preferencyjnego dołączania wierzchołków w modelu BB (3.1), otrzymujemy

$$\frac{\partial k_i}{\partial t} = m \cdot \Pi(k_i, \eta_i) = m \frac{k_i \eta_i}{\sum_{j=1}^{N} k_j \eta_j}. \tag{3.2}$$

Warto zauważyć, że jeżeli wszystkie parametry różnorodności adaptacyjnej η są równe 1, wówczas reguła (3.2) upraszcza się do formy (2.10) znanej z dynamiki stopni modelu BA. Taki rozkład możemy zapisać jako

$$\rho(\eta) = \delta(\eta - 1), \tag{3.3}$$

gdzie funkcja δ oznacza deltę Diraca, a więc funkcję, o której nieformalnie możemy powiedzieć, że przyjmuje wartość jedynie w zerze – stąd η może być równa tylko 1.

Przez analogię do (2.13) zakładamy, że zmiana średniego stopnia i-tego wierzchołka w czasie w modelu BB również podlega prawu potęgowemu, jednak w tym przypadku zależy też od parametru jego różnorodności adaptacyjnej η_i. Wobec tego przyjmujemy, że w modelu będzie zachodzić wielokrotne skalowanie (ang. *multiscaling*). Innymi słowy, uznajemy, że dynamiczny wykładnik β związany z ewolucją i-tego węzła w czasie będzie zależał od konkretnej wartości parametru adaptacyjnego η_i nadanej temu węzłowi, co zapisujemy jako $\beta(\eta_i)$. Wówczas wzór na średni stopień i-tego węzła w zależności od czasu t i czasu t_i dołączenia tego węzła do

sieci będzie wyrażał się wzorem

$$k_i(t,t_i,\eta_i) = m \cdot \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}. \tag{3.4}$$

Wykorzystując (3.4) w równaniu różniczkowym (3.2), uzyskujemy postać dynamicznego wykładnika, którą można zapisać w postaci następującego twierdzenia.

Twierdzenie 3.1. *W modelu BB dynamiczny wykładnik $\beta(\eta_j)$ ewolucji j-tego węzła w czasie zależy wprost proporcjonalnie od parametru różnorodności adaptacyjnej węzła η_j i jest określony wzorem*

$$\beta(\eta_j) = \frac{\eta_j}{C}, \tag{3.5}$$

gdzie stała C została określona poprzez

$$C = \int_{-\infty}^{\infty} \eta_j \rho(\eta_j) \frac{1}{1-\beta(\eta_j)} d\eta_j. \tag{3.6}$$

Dowód. [5] Chcąc dowiedzieć się, jak wygląda postać $\beta(\eta_j)$, będziemy musieli przeprowadzić następujące rozważania.

W rozpatrywanym modelu przyjmujemy założenie, że stopień wierzchołka rośnie w czasie, stąd $\beta(\eta_j)$ nie może być równy 0, bo wówczas $k_j(t,t_j,\eta_j)$, zgodnie z (3.4), byłoby stałe ($k_j = m$). Parametr $\beta(\eta_j)$ nie może też być mniejszy od 0, ponieważ wówczas stopień j-tego wierzchołka malałby wraz z upływem czasu t, co kłóci się z założeniami modelu BB o wzroście sieci. Stopień wierzchołka k_j nie może rosnąć szybciej niż czas t, tzn. $\beta(\eta_j)$ nie może być większa od 1, ponieważ gdyby tak było, liczba krawędzi j-tego wierzchołka sieci mogłaby być większa niż wszystkie krawędzie obecne w sieci – co jest niemożliwe. Przyjęcie parametru $\beta(\eta_j) = 1$ oznaczałoby, że pierwszy węzeł sieci przejmowałby w każdym kroku wszystkie m nowo tworzone połączenia, podczas gdy pozostałe węzły sieci również zyskiwałyby krawędzie – co znów uznawane jest za sytuację niemożliwą. Stąd uznajemy, że dynamiczny wykładnik jest ograniczony za pomocą nierówności

$$0 < \beta(\eta_j) < 1. \tag{3.7}$$

Równanie różniczkowe (3.2) zawiera odwrotność sumy stopni wierzchołków ważoną wartościami parametrów adaptacyjnych. Dowiedzenie się, jaką inną postać może przyjąć to wyrażenie, pozwoli na uproszczenie obliczeń. Aby tego dokonać, Bianconi i Barabási proponują wykorzystanie średniej sumy iloczynów stopnia

wierzchołka oraz jego różnorodności adaptacyjnej $\langle \sum_{j=1}^{N} k_j \eta_j \rangle$ [5]. Średnią tę chcemy określić dla wszystkich możliwych realizacji parametru różnorodności adaptacyjnej η_j, przyjmującego wartości ciągłe.

Wartość oczekiwaną ciągłej zmiennej losowej możemy zapisać jako

$$\langle x \rangle = \int_{-\infty}^{\infty} x \cdot f(x) dx. \tag{3.8}$$

Parametr η_j jest zmienną losową, natomiast $\rho(\eta_j)$ jest jego gęstością rozkładu, dlatego

$$\langle \eta_j \rangle = \int_{-\infty}^{\infty} \eta_j \cdot \rho(\eta_j) d\eta. \tag{3.9}$$

Wartość oczekiwana sumy zmiennych losowych to suma wartości oczekiwanych tych zmiennych. Stopień wierzchołka nie jest zmienną losową, dlatego można zapisać go poza wartością oczekiwaną. W związku z tym, gdyby tylko k_j nie zależało od η_j, otrzymalibyśmy, że

$$\langle \sum_j k_j \eta_j \rangle = \sum_j k_j \langle \eta_j \rangle = \sum_j k_j \int_{-\infty}^{\infty} \eta_j \cdot \rho(\eta_j) d\eta_j. \tag{3.10}$$

Ze względu jednak na to, że stopień $k_j = k_j(t, t_j, \eta_j)$ zależy od parametru adaptacyjnego, każdy z wierzchołków dołącza w innym czasie $1 \leqslant t_j \leqslant t$ i korzystamy z założenia o ciągłości czasu z metody czasu ciągłego – sumę po indeksach wierzchołków j możemy zapisać w postaci całki po czasie dołączenia wierzchołka t_j

$$\langle \sum_j k_j \eta_j \rangle = \int_{-\infty}^{\infty} \eta_j \cdot \rho(\eta_j) \int_1^t k_j(t, t_j, \eta_j) dt_j d\eta_j. \tag{3.11}$$

Korzystając z (3.4), całkę zależną od t_j możemy zapisać jako

$$\begin{aligned} \int_1^t k_j(t, t_j, \eta_j) dt_j &= \int_1^t m \cdot \left(\frac{t}{t_j} \right)^{\beta(\eta_j)} dt_j = m \cdot t^{\beta(\eta_j)} \int_1^t t_j^{-\beta(\eta_j)} dt_j \\ &= m \cdot t^{\beta(\eta_j)} \left[\frac{t_j^{-\beta(\eta_j)+1}}{1-\beta(\eta_j)} \right]_1^t = m \cdot \frac{t - t^{\beta(\eta_j)}}{1-\beta(\eta_j)} = mt \cdot \frac{1 - t^{-(1-\beta(\eta_j))}}{1-\beta(\eta_j)}. \end{aligned} \tag{3.12}$$

Ponieważ $\beta(\eta_j)$ jest ułamkiem należącym do przedziału $(0,1)$ i zakładamy, że proces dołączania nowych węzłów do sieci jest nieskończony (tj. $t \to \infty$), przyjmowane jest, że $t^{\beta(\eta_j)}$ może być wartością zaniedbywalną w porównaniu do t w granicy dążącej do nieskończoności [5]. Aby móc to zapisać, przyjmujemy, że

$$\varepsilon = 1 - \max_{\eta_j} \beta(\eta_j) \tag{3.13}$$

gdzie $\max_{\eta_j} \beta(\eta_j)$ oznacza największy wykładnik dynamiczny $\in (0,1)$, dlatego również $\varepsilon \in (0,1)$. Wówczas uzyskujemy najmniejsze możliwe ε będące ułamkiem. Z tego powodu $t^{-\varepsilon} = \left(\frac{1}{t}\right)^{\varepsilon}$ będzie największą z możliwych do otrzymania wartości $t^{-(1-\beta(\eta_j))}$. Dlatego też zapisujemy, że górna granica asymptotyczna wyrażenia $t^{-(1-\beta(\eta_j))}$ będzie wynosić $O(t^{-\varepsilon})$, a więc wartości tego wyrażenia nie będą rosły szybciej niż $M \cdot t^{-\varepsilon}$, gdzie M jest pewną stałą. Wobec tego, uwzględniając (3.11) i (3.12), możemy zapisać, że gdy $t \to \infty$, wartość oczekiwana sumy z mianownika (3.2) będzie określona jako

$$\begin{aligned}\langle \sum_j k_j \eta_j \rangle &= \int_{-\infty}^{\infty} \eta_j \cdot \rho(\eta_j) d\eta_j \cdot mt \cdot \frac{1 - O(t^{-\varepsilon})}{1 - \beta(\eta_j)} = \\ &= \int_{-\infty}^{\infty} \eta_j \cdot \rho(\eta_j) \cdot \frac{1}{1-\beta(\eta_j)} d\eta_j \cdot mt \cdot (1 - O(t^{-\varepsilon})) = Cmt(1 - O(t^{-\varepsilon})) \to Cmt,\end{aligned} \tag{3.14}$$

gdzie

$$C = \int_{-\infty}^{\infty} \eta_j \cdot \rho(\eta_j) \cdot \frac{1}{1-\beta(\eta_j)} d\eta_j. \tag{3.15}$$

Wówczas, korzystając z przybliżenia (3.14), równanie różniczkowe zmiany średniego stopnia j-tego wierzchołka w czasie możemy zapisać w postaci znanej z (3.2) w następujący sposób

$$\frac{\partial k_j(t,t_j,\eta_j)}{\partial t} \approx m \cdot \frac{k_j(t,t_j,\eta_j) \cdot \eta_j}{Cmt} = \frac{\eta_j}{C} \cdot \frac{k_j(t,t_j)}{t}. \tag{3.16}$$

Rozwiązanie szczególne tego typu równania różniczkowego, które zostało uzyskane dzięki (2.11) i (2.12), będzie postaci (2.13). Do otrzymania takiej postaci dążymy, jeżeli tylko przyjmiemy, że $\frac{\eta_j}{C}$ jest wykładnikiem β. Zapisujemy zatem ostatecznie, że

$$\beta(\eta_j) = \frac{\eta_j}{C}. \tag{3.17}$$

□

Podstawiając wzór (3.5) do (3.6), otrzymujemy

$$C = \int_{-\infty}^{\infty} \eta_j \rho(\eta_j) \frac{1}{1 - \frac{\eta_j}{C}} d\eta_j. \tag{3.18}$$

Następnie, dzieląc stronami przez C, uzyskujemy

$$1 = \int_{-\infty}^{\infty} \eta_j \rho(\eta_j) \frac{1}{C \cdot (1 - \frac{\eta_j}{C})} d\eta_j. \tag{3.19}$$

Przekształcając prawą stronę powyższej równości i uwzględniając, że parametr różnorodności adaptacyjnej η_j może osiągać wartości od 0 do η_{max} (oznaczającego maksymalną wartość tego parametru w sieci) w następujący sposób

$$\int_0^{\eta_{max}} \eta_j \rho(\eta_j) \frac{1}{C \cdot (1 - \frac{\eta_j}{C})} d\eta_j = \int_0^{\eta_{max}} \rho(\eta_j) \frac{1}{\frac{1}{\eta_j} \cdot (C - \eta_j)} d\eta_j = \int_0^{\eta_{max}} \rho(\eta_j) \frac{1}{\frac{C}{\eta_j} - 1} d\eta_j, \tag{3.20}$$

otrzymujemy ostatecznie

$$\int_0^{\eta_{max}} \rho(\eta_j) \frac{1}{\frac{C}{\eta_j} - 1} d\eta_j = 1. \tag{3.21}$$

Jeżeli mianownik wyrażenia znajdującego się w całce (3.21) dla pewnej wartości η_j wynosiłby 0, wówczas wartość funkcji podcałkowej w tym punkcie byłaby nieskończona, co oznacza, że całkę tę moglibyśmy nazwać osobliwą. Jednak ze względu na to, że nierówność $\beta(\eta_j) < 1$ jest spełniona dla każdego η_j, zgodnie z (3.7), dla każdego j zachodzi nierówność

$$\frac{\eta_j}{C} < 1, \tag{3.22}$$

wobec czego C będzie większe również od największej wartości parametru η w układzie

$$\eta_{max} < C. \tag{3.23}$$

Z tego względu $\frac{C}{\eta_j}$ nigdy nie osiągnie wartości 1, a więc całka z równania (3.21) w swoich granicach całkowania nie będzie osobliwa.

Z drugiej strony sumę iloczynów stopni wierzchołków i parametrów ich różnorodności adaptacyjnej w sieci możemy oszacować z góry, korzystając z oszacowania

sumy stopni wierzchołków z (2.5)

$$\sum_{j=1}^{N} k_j \eta_j \leqslant \eta_{max} \sum_{j=1}^{N} k_j \approx \eta_{max} \cdot 2mt. \tag{3.24}$$

Wstawiając w miejsce sumy jej wartość oczekiwaną oszacowaną w (3.14), otrzymujemy, że

$$Cmt(1 - O(t^{-\varepsilon})) \approx Cmt \leqslant \eta_{max} \cdot 2mt, \tag{3.25}$$

z czego uzyskujemy

$$C \leqslant 2 \cdot \eta_{max}. \tag{3.26}$$

Łącząc nierówności (3.23) i (3.26), otrzymujemy przybliżone granice, w których powinna znajdować się stała C

$$\eta_{max} < C \leqslant 2 \cdot \eta_{max}. \tag{3.27}$$

Zależność pomiędzy parametrem różnorodności adaptacyjnej węzła a tempem wzrostu stopnia wierzchołka dana jest wzorem (3.4). Po obustronnym zlogarytmowaniu równania otrzymujemy, że

$$\begin{aligned} \ln(k(t,t_i,\eta_i)) &= \ln\left(m \cdot \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}\right), \\ \ln(k(t,t_i,\eta_i)) &= \ln(t^{\beta(\eta_i)}) + \ln\left(\frac{m}{t_i^{\beta(\eta_i)}}\right), \\ \ln(k(t,t_i,\eta_i)) &= \beta(\eta_i) \cdot \ln(t) + B_i, \end{aligned} \tag{3.28}$$

gdzie $B_i = \ln\left(\frac{m}{t_i^{\beta(\eta_i)}}\right)$ jest parametrem niezależnym od czasu t.

Innym sposobem zapisania powyższej równości jest

$$\begin{aligned} \ln(k(t,t_i,\eta_i)) &= \ln\left(m \cdot \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}\right), \\ \ln(k(t,t_i,\eta_i)) &= \ln(m) + \beta(\eta_i) \cdot \ln\left(\frac{t}{t_i}\right). \end{aligned} \tag{3.29}$$

Wówczas jeżeli znamy stopień $k(t,t_i,\eta_i)$ i-tego wierzchołka w chwili t, chwilę jego dołączenia do sieci t_i oraz rozpatrywany czas t, jesteśmy w stanie oszacować parametry m i $\beta(\eta_i)$. Dokonujemy tego na podstawie dopasowania funkcji liniowej do wykresu zależności logarytmu stopnia wierzchołka $\ln(k(t,t_i,\eta_i))$ od logarytmu

względnego czasu $\ln\left(\frac{t}{t_i}\right)$. W tak uzyskanym wzorze funkcji o formie $y = ax + b$ otrzymana stała b odpowiada $\ln(m)$, natomiast a jest nachyleniem wykresu $\beta(\eta_i)$.

Oznacza to, że jeżeli jesteśmy w stanie prześledzić ewolucję stopni w czasie dla wystarczająco dużej liczby węzłów, będziemy też w stanie określić rozkład dynamicznego wykładnika $\beta(\eta)$. Wykładnik ten jest z kolei wprost proporcjonalny do parametru η zgodnie z zależnością (3.5), dlatego też rozkład $\beta(\eta)$ będzie odpowiadał rozkładowi $\rho(\eta)$. W ten sposób, badając wzrost sieci, można uzyskać informacje na temat parametrów adaptacyjnych [2].

Aby poznać zmianę i-tego stopnia w czasie, przyjmujemy oznaczenia t_{pocz} – czas rozpoczęcia obserwacji, t_{konc} – czas zakończenia obserwacji oraz $t_{pocz} < t_{konc}$ i wyznaczamy

$$\begin{aligned} k_{i,t_{konc}} - k_{i,t_{pocz}} &= \ln(k(t_{konc}, t_i, \eta_i)) - \ln(k(t_{pocz}, t_i, \eta_i)) = \\ &= \beta(\eta_i) \cdot \ln(t_{konc}) + B_i - \beta(\eta_i) \cdot \ln(t_{pocz}) - B_i = \beta(\eta_i) \cdot \ln\left(\frac{t_{konc}}{t_{pocz}}\right). \end{aligned} \tag{3.30}$$

Wówczas otrzymujemy, że zmiana stopnia wierzchołka w czasie

$$\frac{\Delta k}{\Delta t} = \frac{k_{i,t_{konc}} - k_{i,t_{pocz}}}{t_{konc} - t_{pocz}} = \beta(\eta_i) \cdot \frac{\ln(t_{konc} - t_{pocz})}{t_{konc} - t_{pocz}} \tag{3.31}$$

zależy liniowo od dynamicznego wykładnika $\beta(\eta_i)$. Oznacza to, że parametr $\beta(\eta_i)$ wpływa na nachylenie wykresu funkcji zmiany stopnia węzła w czasie, co można zaobserwować na rysunku 2.

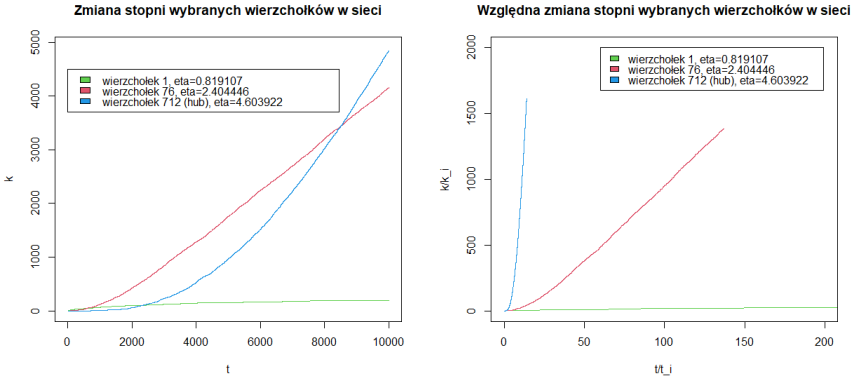
Względna różnica stopni dwóch wierzchołków i, j może być zapisana jako

$$\begin{aligned} \frac{k(t, t_j, \eta_j) - k(t, t_i, \eta_i)}{k(t, t_i, \eta_i)} &= \frac{m \cdot \left(\frac{t}{t_j}\right)^{\beta(\eta_j)} - m \cdot \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}}{m \cdot \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}} = \frac{t^{\beta(\eta_j)} \cdot t_i^{\beta(\eta_i)}}{t_j^{\beta(\eta_j)} \cdot t^{\beta(\eta_i)}} - 1 = \\ &= t^{\beta(\eta_j) - \beta(\eta_i)} \cdot \frac{t_i^{\beta(\eta_i)}}{t_j^{\beta(\eta_j)}} - 1 = t^{\frac{\eta_j - \eta_i}{C}} \cdot \left(\frac{t_i^{\eta_i}}{t_j^{\eta_j}}\right)^{\frac{1}{C}} - 1. \end{aligned} \tag{3.32}$$

Gdy dwa wierzchołki – i oraz j – zostają dołączone do sieci w tym samym kroku czasowym $t_i = t_j$, ale posiadają inne parametry adaptacyjne $\eta_i \neq \eta_j$, względna różnica ich stopni w granicach dużych czasów $t \to \infty$ jest uznawana za proporcjonalną do czasu z wykładnikiem potęgowym uwzględniającym różnicę w ich parametrach

adaptacyjnych [2]

$$\frac{k(t,t_j,\eta_j)-k(t,t_i,\eta_i)}{k(t,t_i,\eta_i)} = t^{\frac{\eta_j-\eta_i}{C}} \cdot \left(\frac{t_i^{\eta_i}}{t_i^{\eta_j}}\right)^{\frac{1}{C}} - 1 = t^{\frac{\eta_j-\eta_i}{C}} \cdot t_i^{\frac{\eta_i-\eta_j}{C}} - 1 \propto t^{\frac{\eta_j-\eta_i}{C}}. \quad (3.33)$$



Rysunek 2. Wykres zmiany oraz względnej zmiany stopni wierzchołków o parametrach adaptacyjnych $\eta_1 = 0.819107$, $\eta_{76} = 2.404446$, $\eta_{712} = 4.603922$ występujących w sieci BB o parametrach `N` $= 10000$, `num_seed`$= 3$, `multiple_node`$= 1$, `m`$= 3$, `mode_f = "power_law"`, `s`$= 10$

Źródło: Opracowanie własne

Średnia liczba połączeń uzyskiwanych przez wierzchołek jest proporcjonalna do stosunku obecnego czasu istnienia sieci do czasu dołączenia węzła $\frac{t}{t_i}$. Dzieje się tak zarówno w modelu BA (równanie (2.13)), jak też w modelu BB (zależność (3.4)). Z tego powodu w przypadku gdy dwa węzły – i oraz j – posiadają taki sam parametr różnorodności adaptacyjnej $\eta_i = \eta_j$, ale dołączyły do sieci w różnych momentach czasowych $t_i < t_j$, modele te zakładają, że szybciej będzie rósł stopień tego wierzchołka, który wcześniej dołączył do sieci, ponieważ wówczas $\frac{t}{t_i} > \frac{t}{t_j}$.

Załóżmy, że dwa wierzchołki sieci posiadają różne czasy dołączenia $t_i < t_j$ oraz różne parametry adaptacyjne $\eta_i < \eta_j$. Aby dowiedzieć się, od którego momentu czasowego stopień lepiej „dopasowanego” węzła dołączonego później do sieci $k(t,t_j,\eta_j)$ będzie większy niż stopień węzła dłużej występującego w sieci o mniejszym parametrze adaptacyjnym $k(t,t_i,\eta_i)$, przekształcamy utworzoną nierówność

$$
\begin{aligned}
k(t,t_i,\eta_i) &< k(t,t_j,\eta_j),\\
m\cdot\left(\frac{t}{t_i}\right)^{\beta(\eta_i)} &< m\cdot\left(\frac{t}{t_j}\right)^{\beta(\eta_j)},\\
\frac{t^{\beta(\eta_i)}}{t_i^{\beta(\eta_i)}} &< \frac{t^{\beta(\eta_j)}}{t_j^{\beta(\eta_j)}},\\
t^{\beta(\eta_i)-\beta(\eta_j)} &< \frac{t_i^{\beta(\eta_i)}}{t_j^{\beta(\eta_j)}},\\
t^{\frac{\eta_i-\eta_j}{C}} &< \left(\frac{t_i^{\eta_i}}{t_j^{\eta_j}}\right)^{\frac{1}{C}},\ C>0,\ \eta_i-\eta_j<0,\\
t &> \left(\frac{t_i^{\eta_i}}{t_j^{\eta_j}}\right)^{\frac{1}{\eta_i-\eta_j}}.
\end{aligned}
\tag{3.34}
$$

Przykładowo, jeżeli czas dołączenia wierzchołka j byłby o rząd wielkości większy od czasu przyłączenia węzła i, tzn. $t_j = 10\cdot t_i$, natomiast parametr adaptacyjny węzła j byłby dwukrotnie większy od tego występującego dla wierzchołka i, tj. $\eta_j = 2\eta_i$, wówczas, podstawiając dane do zależności uzyskanej w (3.34), otrzymalibyśmy, że

$$
t > \left(\frac{t_i^{\eta_i}}{(10t_i)^{2\eta_i}}\right)^{\frac{1}{\eta_i-2\eta_i}} = \frac{t_i^{-1}}{(10t_i)^{-2}} = 100t_i. \tag{3.35}
$$

Oznacza to, że aby j-ty wierzchołek osiągnął stopień większy niż i-ty, czas ewolucji sieci musi być o dwa rzędy wielkości większy niż czas przyłączenia do sieci i-tego węzła. Tłumaczy to, dlaczego podczas obserwacji sieci zauważyć można, że węzły o stosunkowo dużej wartości parametru η nie zawsze osiągają stopnie większe od tych uzyskiwanych przez wierzchołki o mniejszych wartościach parametrów adaptacyjnych.

Gdybyśmy natomiast przyjęli, że $\beta(\eta_i) < \beta(\eta_j)$, a parametr m jest estymowany dla każdego z wierzchołków, przez co we wzorze występują parametry $m_i, m_j > 0$, wówczas analogicznie średni stopień węzła j-tego będzie większy od średniego stopnia węzła i-tego od momentu czasowego t spełniającego nierówność

$$\begin{aligned} k(t,t_i,\beta(\eta_i)) &< k(t,t_j,\beta(\eta_j)),\\ m_i\cdot\left(\frac{t}{t_i}\right)^{\beta(\eta_i)} &< m_j\cdot\left(\frac{t}{t_j}\right)^{\beta(\eta_j)},\\ t &> \left(\frac{m_j}{m_i}\cdot\frac{t_i^{\beta(\eta_i)}}{t_j^{\beta(\eta_j)}}\right)^{\frac{1}{\beta(\eta_i)-\beta(\eta_j)}}. \end{aligned} \tag{3.36}$$

3.2. Rozkład stopni wierzchołków

W modelu Bianconi–Barabásiego rozkład stopni wierzchołków (ang. *connectivity distribution, degree distribution*) $P(k)$, określający prawdopodobieństwo, z jakim losowo wybrany z sieci wierzchołek będzie miał stopień k, zależy m.in. od gęstości rozkładu parametru różnorodności adaptacyjnej $\rho(\eta)$. Do określania rozkładu stopni węzłów sieci utworzonej przy pomocy modelu BB będzie nam służyło następujące spostrzeżenie.

Niech m będzie liczbą krawędzi tworzonych z nowym węzłem sieci w każdym kroku czasowym, C oznacza stałą daną wzorem (3.6), a η_{max} symbolizuje największą wartość parametru adaptacyjnego w układzie. Rozkład stopni wierzchołków w modelu Bianconi–Barabásiego zależy od rozkładu prawdopodobieństwa $\rho(\eta)$ parametrów różnorodności adaptacyjnej η, a jego przybliżoną postać, otrzymaną przy pomocy metody czasu ciągłego można zapisać jako [2]

$$P(k) \propto C\int_0^{\eta_{max}} \frac{\rho(\eta_i)}{\eta_i}\left(\frac{m}{k}\right)^{\frac{C}{\eta_i}+1} d\eta_i. \tag{3.37}$$

Dowód. [2] Do uzyskania przybliżonego wzoru na rozkład stopni wierzchołków $P(k)$ w modelu BB możemy skorzystać z opisanej w podrozdziale 2.2 metody czasu ciągłego oraz z dynamiki stopni węzłów przedstawionej w podrozdziale 3.1.

Z uzyskanego w (2.19) wyniku wiemy, że jeżeli dynamiczny wykładnik β jest stały (tzn. jednakowy dla wszystkich wierzchołków w sieci), wówczas mamy do czynienia z potęgowym rozkładem stopni węzłów o wykładniku charakterystycznym $\alpha = \frac{1}{\beta}+1$, którego skumulowany rozkład stopni wierzchołków $P^c(k)$ wyprowadzaliśmy w (2.18). W modelu BB β zależy od zmiennej losowej η, dlatego aby określić skumulowany rozkład prawdopodobieństwa $P^c(k)$ w modelu BB na podobieństwo (2.15), będziemy chcieli obliczyć

$$P^c(k) = P(k_i(t,t_i,\eta_i) \leqslant k) = 1 - P(k_i(t,t_i,\eta_i) > k). \tag{3.38}$$

Aby otrzymać $P(k_i(t,t_i,\eta_i) > k)$, określamy prawdopodobieństwo tego, że losowo wybrany z sieci wierzchołek będzie posiadał średni stopień większy od wybranego stopnia k. Podstawiając $\beta = \beta(\eta_i) = \frac{\eta_i}{C}$ z zależności (3.5) do (2.17), otrzymujemy

$$t_i < t \cdot \left(\frac{m}{k}\right)^{\frac{1}{\beta(\eta_i)}} = t \cdot \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}}. \tag{3.39}$$

W procedurze konstrukcyjnej w każdym kroku czasowym do sieci dodawany jest jeden wierzchołek, stąd liczba węzłów o średnim stopniu $k_i(t,t_i,\eta_i)$ większym niż k będzie liczbą wierzchołków o momencie t_i dołączenia do sieci. Wartość η_i jest przyporządkowywana każdemu węzłowi podczas jego włączania do sieci z prawdopodobieństwem określonym gęstością $\rho(\eta_i)$. Ze względu na to, że parametr ten jest ciągłą zmienną losową, we wzorze na liczbę wierzchołków o $k_i(t,t_i,\eta_i) > k$ uwzględniamy prawdopodobieństwo $\rho(\eta_i)$ z jakim i-ty wierzchołek przyjmuje parametr różnorodności η_i [2]. Dlatego też liczbę wierzchołków o średnim stopniu $k_i(t,t_i,\eta_i) > k$ możemy określić jako sumę po zmieniającym się parametrze η_i z uwzględnieniem prawdopodobieństwa $\rho(\eta_i)$, a ze względu na ciągłość zmiennej losowej η_i zapisać ją jako całkę

$$\int_0^{\eta_{max}} t \cdot \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}} \cdot \rho(\eta_i) d\eta_i. \tag{3.40}$$

Wówczas, analogicznie jak w (2.18), wyprowadzamy wzór na asymptotyczne przybliżenie $P^c(k)$ w granicach dużych t dążących do nieskończoności, przez co liczba n_0 wierzchołków grafu początkowego jest uznawana za zaniedbywalną

$$\begin{aligned} P^c(k) &= P(k_i(t,t_i,\eta_i) \leqslant k) = 1 - P(k_i(t,t_i,\eta_i) > k) \approx \\ &\approx 1 - \frac{\int_0^{\eta_{max}} t \cdot \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}} \cdot \rho(\eta_i) d\eta_i}{N} = 1 - \frac{\int_0^{\eta_{max}} t \cdot \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}} \cdot \rho(\eta_i) d\eta_i}{t + n_0} \approx \\ &\approx 1 - \frac{\int_0^{\eta_{max}} t \cdot \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}} \cdot \rho(\eta_i) d\eta_i}{t} = 1 - \int_0^{\eta_{max}} \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}} \cdot \rho(\eta_i) d\eta_i. \end{aligned} \tag{3.41}$$

Ostatecznie po wyprowadzeniu pochodnej, aby dowiedzieć się jak przebiega zmiana rozkładu skumulowanego w stopniach, otrzymujemy rozkład stopni wierzchołków

w modelu BB

$$
\begin{aligned}
P(k) &= \frac{\partial P^c(k)}{\partial k} = \frac{\partial}{\partial k}\left(1 - \int_0^{\eta_{max}} \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}} \cdot \rho(\eta_i) d\eta_i\right) = \\
&= -\int_0^{\eta_{max}} \frac{\partial}{\partial k}\left(\left(\frac{m}{k}\right)^{\frac{C}{\eta_i}}\right) \cdot \rho(\eta_i) d\eta_i = \\
&= -\int_0^{\eta_{max}} m^{\frac{C}{\eta_i}} \cdot \left(-\frac{C}{\eta_i}\right) \cdot k^{-\frac{C}{\eta_i}-1} \cdot \rho(\eta_i) d\eta_i = \int_0^{\eta_{max}} m^{\frac{C}{\eta_i}} \cdot \frac{C}{\eta_i} \cdot k^{-\left(\frac{C}{\eta_i}+1\right)} \cdot \rho(\eta_i) d\eta_i = \\
&= C \cdot \int_0^{\eta_{max}} \frac{\rho(\eta_i)}{\eta_i} \cdot \frac{m^{\frac{C}{\eta_i}}}{k^{\frac{C}{\eta_i}+1}} d\eta_i \propto C \cdot \int_0^{\eta_{max}} \frac{\rho(\eta_i)}{\eta_i} \cdot \left(\frac{m}{k}\right)^{\frac{C}{\eta_i}+1} d\eta_i,
\end{aligned}
\tag{3.42}
$$

gdzie

- $P(k)$ – prawdopodobieństwo, że wylosowany z sieci węzeł będzie stopnia k,
- η_{max} – największa wartość parametru różnorodności adaptacyjnej η w sieci,
- m – liczba krawędzi przyłączanych w każdym kroku czasowym,
- C – stała określona za pomocą wzoru (3.6), o wartości z zakresu (3.27): $\eta_{max} \leqslant C \leqslant 2\eta_{max}$,
- η_i – ciągła zmienna losowa będąca parametrem różnorodności adaptacyjnej i-tego wierzchołka,
- $\rho(\eta_i)$ – rozkład (gęstość) prawdopodobieństwa parametru różnorodności adaptacyjnej η_i i-tego węzła (ze względu na to, że wszystkie wierzchołki w sieci otrzymują parametr η z tego samego rozkładu, można również zapisać $\rho(\eta)$).

□

Z obecności parametru różnorodności adaptacyjnej η każdego z węzłów i ich rozkładu $\rho(\eta)$ we wzorach na rozkład stopni wierzchołków (3.37) oraz dynamicznego wykładnika (3.5) wynika, że jakość każdego z węzłów wpływa na strukturę i dynamikę tworzącej się sieci.

Według książki [12], ze względu na mapowanie dokonane pomiędzy modelem BB a gazem Bosego, w układzie takim występować mogą dwa zjawiska zdeterminowane przez rozkład parametru różnorodności adaptacyjnej $\rho(\eta)$. Pierwsze z nich dotyczy sytuacji, w której liczba krawędzi zdobywanych przez i-ty wierzchołek różni się w zależności od przyjętego parametru adaptacyjnego η_i, ale żaden z węzłów nie osiąga liczbą uzyskanych połączeń ekstremalnej przewagi nad pozostałymi. W przypadku sieci o ogromnych rozmiarach stosunek stopnia danego wierzchołka do wszystkich krawędzi występujących w sieci dąży wówczas do zera. Uznawane

jest, że stan taki dotyczy sieci o potęgowych rozkładach stopni wierzchołków $P(k)$, a więc sieci bezskalowych. Drugie zjawisko obrazuje sytuację, w której jeden najbardziej „dopasowany" do sieci wierzchołek zdobywa większość połączeń, tj. skończony ułamek wszystkich krawędzi w sieci, i utrzymuje zdobytą przewagę stopnia nawet wówczas, gdy sieć rozrasta się do ogromnych rozmiarów. Zjawisko takie nazywane jest stanem kondensacji Bosego–Einsteina i nawiązuje do gromadzenia się elektronów na najniższej powłoce elektronowej w gazie Bosego, gdy gaz ten schładzany jest do temperatury bliskiej zera absolutnego. Układ przypomina wspomniany kondensat Bosego–Einsteina wówczas, gdy jest zdominowany przez jeden występujący „hub". Sieć taka nie jest wówczas uznawana za bezskalową.

W artykule [6] autorzy: C. Borgs, C. Daskalakis, J. Chayes i S. Roch wskazują, że w zależności od przyjętego rozkładu parametru różnorodności adaptacyjnej rozpatrywany układ może osiągnąć jeden z trzech stanów, jakimi są:

— stan przewagi pierwszego gracza (ang. *first-mover-advantage*),
— stan „wzbogacania się najlepiej dopasowanego gracza" (ang. *fit-get-richer*),
— stan „opłacalnej inwestycji" (ang. *innovation-pays-off*) znany jako stan kondensacji Bosego–Einsteina (ang. *Bose-Einstein condensation*).

Stan przewagi pierwszego gracza występuje, gdy rozkład $\rho(\eta)$ jest rozkładem jednostajnym. Zjawisko to związane jest z potęgowym (bezskalowym) charakterem rozkładu stopni wierzchołków i nawiązuje do ewolucji sieci opisanej za pomocą klasycznej reguły preferencyjnego przyłączania węzłów (2.6). W tym stanie „hubami" zostają węzły, które pojawiły się w sieci jako pierwsze. Stan „wzbogacania się najlepiej dopasowanego gracza" opisuje zjawisko przyłączania przez wierzchołki o największych parametrach adaptacyjnych η największej liczby krawędzi w miarę upływu czasu. Węzły takie zostają „hubami" dzięki występowaniu zmodyfikowanej reguły preferencyjnego dołączania wierzchołków (3.1), uwzględniającej skalowanie stopnia węzła przez jego parametr adaptacyjny. Stan „opłacalnej inwestycji", znany też jako kondensacja Bosego–Einsteina, nawiązuje natomiast do sytuacji, w której stała liczba połączeń z sieci przełącza się do węzła o większym parametrze adaptacyjnym. Na zjawisko to wpływa rozkład parametrów adaptacyjnych $\rho(\eta)$. Nie występuje ono w fazie „wzbogacania się najlepiej dopasowanego gracza" ani nie zależy od wielkości sieci.

Warto zauważyć, że jeżeli $\rho(\eta)$ posiada ograniczoną dziedzinę (ang. *finite domain*, *finite support*), wówczas rozkład stopni wierzchołków $P(k)$ będzie potęgowy [7]. Z tego powodu w sieciach modelu BB o takich rozkładach będziemy mieli do czynienia ze zjawiskiem bezskalowości opisanym szerzej w podrozdziale 1. Jednym z rozkładów o ograniczonej dziedzinie jest rozkład $\rho(\eta)$, opisany za pomocą wzoru (3.3), w którym wszystkie wartości parametru adaptacyjnego η wynoszą 1. Jest to najprostsza forma rozkładu $\rho(\eta)$ zakładająca, że jakość węzłów z punktu widzenia

sieci nie różni się między sobą. Wstawiając $\eta_j = 1$ do (3.21), uzyskujemy

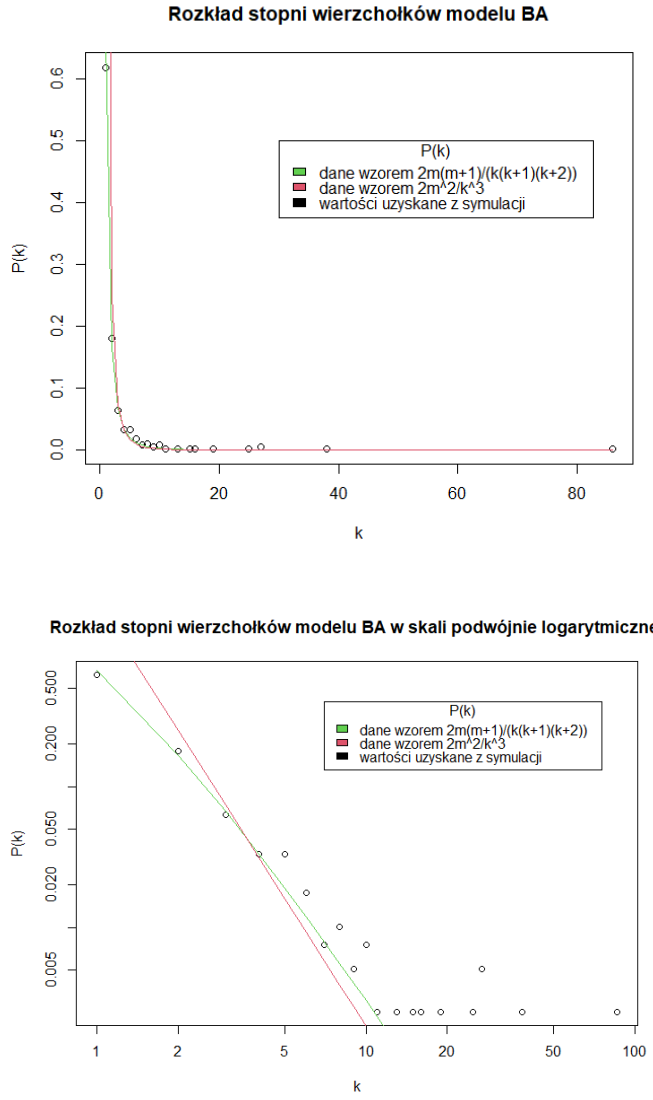
$$\begin{aligned} \int_0^{\eta_{max}} \rho(\eta_j) \frac{1}{C-1} d\eta_j = 1, \\ \int_0^{\eta_{max}} \rho(\eta_j) d\eta_j = C-1, \end{aligned} \tag{3.43}$$

a uwzględniając, że całka z gęstości rozkładu prawdopodobieństwa po całej przestrzeni wynosi 1, otrzymujemy, że $C = 2$. Podstawiając określone w ten sposób η_j i C do wzoru (3.5), uzyskujemy, że dynamiczny wykładnik $\beta(\eta_j) = \frac{1}{2}$ dla każdego j. Oznacza to, że β przyjmuje stałą wartość, a wówczas zgodnie ze wzorem (2.20) charakterystyczny wykładnik rozkładu potęgowego będzie wynosił $\alpha = 3$, stąd rozkład stopni wierzchołków

$$P(k) \propto k^{-3}. \tag{3.44}$$

Powoduje to, że model Bianconi–Barabásiego staje się modelem Barabásiego–Albert. Warto zauważyć, że C oraz α osiągają w tym przypadku największe możliwe wartości uzyskiwane dla sieci bezskalowych [5] (w sieciach rzeczywistych α najczęściej należy do zakresu $(2,3)$, co zostało podane w podrozdziale 2.3). Uzyskany w ten sposób rozkład stopni wierzchołków, widoczny na rysunku 3, podlega prawom potęgowym, dlatego jego kształt jest zbliżony do kształtu rozkładu potęgowego z rysunku 1. W prawym dolnym rogu wykresu widoczne są punkty odbiegające od linii aproksymujących rozkład. Ich obecność wskazuje na występowanie w sieci niewielu wierzchołków o stosunkowo dużym stopniu.

Należy jednak nadmienić, że przedstawiając wykresy rozkładu stopni węzłów w skali podwójnie logarytmicznej, zdarza się, że rozkłady niepodlegające prawom potęgowym również są prezentowane jak prostoliniowe zależności. Nie oznacza to jednak, że faktycznie będzie występować w nich bezskalowość, ponieważ z nią mamy do czynienia jedynie w rozkładach potęgowych. Najczęściej tego typu pomyłki w interpretacjach zdarzają się w przypadku rozkładów log-normalnych (ang. *log-normal distribution*) – odznaczających się prawostronną asymetrią oraz, w przeciwieństwie do rozkładu potęgowego, możliwością występowania maksymalnej wartości funkcji gęstości – a także rozkładów potęgowych z wykładniczym obcięciem (ang. *power law with an exponential cutoff*) [9]. Ostatnie z nich charakteryzowane są przez zachowywanie cech funkcji potęgowej w przypadku małych wartości argumentów oraz płynne przechodzenie w funkcję gęstości prawdopodobieństwa malejącą wykładniczo, tj. szybciej od funkcji potęgowej, gdy argumenty osiągają duże wartości [11].



Rysunek 3. Wykresy przedstawiające rozkład stopni wierzchołków, uzyskany na podstawie symulacji sieci modelu BA o $N = 1000$ **węzłów, gdy graf początkowy składał się z** `num_seed`$= 3$ **wierzchołków, w każdym kroku czasowym dodawany był** `multiple_node`$= 1$ **wierzchołek i** `m`$= 1$ **krawędź, a wykładnik charakterystyczny** α**=1 zapewnił liniową regułę preferencyjnego dołączania wierzchołków**

Źródło: Opracowanie własne

Innym przykładem $\rho(\eta)$ o ograniczonej dziedzinie jest rozkład jednostajny na przedziale $[a,b]$. Funkcja gęstości rozkładu jednostajnego dana jest wzorem

$$\rho(\eta) = \frac{1}{b-a}, \tag{3.45}$$

gdzie a i b to odpowiednio początek i koniec przedziału. Jest to jeden z najprostszych rozkładów wykorzystywanych do modelowania zachowania sieci, w której węzły różnią się między sobą jakością. Aby obliczyć stałą C, którą dla tego rozkładu oznaczymy jako $\widetilde{C}$, do całki z zależności (3.21) podstawiamy gęstość prawdopodobieństwa (3.45) i otrzymujemy

$$\begin{aligned}
&\int_0^1 \frac{1}{b-a}\cdot\frac{1}{\frac{\widetilde{C}}{\eta_j}-1}d\eta_j = \frac{1}{b-a}\cdot\int_0^1 \frac{\eta_j}{\widetilde{C}-\eta_j}d\eta_j = \frac{1}{b-a}\cdot\int_0^1 \frac{\widetilde{C}+\eta_j-\widetilde{C}}{\widetilde{C}-\eta_j}d\eta_j = \\
&= \frac{1}{b-a}\cdot\left(\widetilde{C}\int_0^1 \frac{1}{\widetilde{C}-\eta_j}d\eta_j + \int_0^1 \frac{\eta_j-\widetilde{C}}{\widetilde{C}-\eta_j}d\eta_j\right) = \\
&= \frac{1}{b-a}\cdot\left(\widetilde{C}\cdot\left[-\ln|\widetilde{C}-\eta_j|\right]_0^1 - \int_0^1 \frac{\widetilde{C}-\eta_j}{\widetilde{C}-\eta_j}d\eta_j\right) = \\
&= \frac{1}{b-a}\cdot\left[-\widetilde{C}\cdot(\ln|\widetilde{C}-1|-\ln|\widetilde{C}|)-1\right] = \\
&= \frac{1}{b-a}\left(-\widetilde{C}\cdot\ln\left|\frac{\widetilde{C}-1}{\widetilde{C}}\right|-1\right) = \frac{1}{b-a}\left(-\widetilde{C}\cdot\ln\left|1-\frac{1}{\widetilde{C}}\right|-1\right),
\end{aligned} \tag{3.46}$$

gdzie, według wspomnianej zależności, uzyskany wynik ma być równy 1. Stąd

$$\begin{aligned}
&\frac{1}{b-a}\left(-\widetilde{C}\cdot\ln\left|1-\frac{1}{\widetilde{C}}\right|-1\right) = 1 \\
&\ln\left|1-\frac{1}{\widetilde{C}}\right| = \frac{(b-a)+1}{-\widetilde{C}} \\
&1-\frac{1}{\widetilde{C}} = \exp\left(-\frac{b-a+1}{\widetilde{C}}\right).
\end{aligned} \tag{3.47}$$

Z powyższego równania numerycznie wyznaczyć można wartość $\widetilde{C}$. Dla $\rho(\eta)$, będącego rozkładem jednostajnym na przedziale $[0,1]$, stała ta wynosi 1.255.

W efekcie, na podstawie wzoru (3.37) i przy uwzględnieniu gęstości (3.45), uzyskujemy zależność, jaką charakteryzuje się rozkład stopni wierzchołków

$$P(k) \propto \widetilde{C} \int_0^{\eta_{max}} \frac{1}{b-a} \cdot \frac{1}{\eta_j} \left(\frac{m}{k_{j,t}} \right)^{\frac{\widetilde{C}}{\eta_j}+1} d\eta_j = \frac{\widetilde{C}}{b-a} \int_0^{\eta_{max}} \frac{m^{\frac{\widetilde{C}}{\eta_j}+1}}{\eta_j} \cdot k_{j,t}^{-(\frac{\widetilde{C}}{\eta_j}+1)} d\eta_j. \quad (3.48)$$

Gdy rozkład $\rho(\eta)$ w sieci BB jest jednostajny na przedziale [0,1], uznawane jest, że rozkład stopni wierzchołków (3.48) można przybliżyć poprzez

$$P(k) \sim \int_0^1 \frac{C}{\eta} \frac{1}{k^{C+1}}. \quad (3.49)$$

Ze względu na to, że w rzeczywistych układach nie spodziewamy się występowania idealnego rozkładu potęgowego, aby móc lepiej dopasować wykres do wzoru, wprowadzana jest logarytmiczna korekta. W literaturze spotyka się stwierdzenie, że gdy $\rho(\eta) = const$, wówczas rozkład stopni wierzchołków można aproksymować uogólnionym rozkładem potęgowym posiadającym logarytmiczną korektę

$$P(k) \sim \frac{k^{-(1+\widetilde{C})}}{\ln k} \quad (3.50)$$

o wyznaczonej stałej $\widetilde{C}$ równej 1.255. Sieci generowane za pomocą takiego modelu wykazują własności dysasortatywne (ang. *disassortative*), co oznacza, że w ich strukturze węzły o dużym stopniu z największym prawdopodobieństwem będą łączyły się z węzłami posiadającymi niewielki stopień, a węzły o małym stopniu z największym prawdopodobieństwem będą tworzyły krawędzie z wierzchołkami o dużych stopniach. Zjawisko to jest dostrzegalne w układach rzeczywistych takich jak rynki finansowe oraz internet [9].

W większości rozkładów $\rho(\eta)$ rozkład stopni wierzchołków $P(k)$ wciąż posiada cechy charakterystyczne dla rozkładów sieci bezskalowych, mimo że nie jest idealnym rozkładem potęgowym [14]. Dopiero w przypadku dobrania odpowiedniej funkcji gęstości $\rho(\eta)$ może dojść w $P(k)$ oraz dynamicznym wykładniku $\beta(\eta)$ do faktycznego zaniku potęgowego charakteru rozkładów na rzecz zależności opartych na np. „rozciągniętej" funkcji wykładniczej (opisanej za pomocą wzoru (3.52)) oraz złożonej kombinacji logarytmów i prawa potęgowego, jak dzieje się to w przypadku $\rho(\eta)$ będącego funkcją wykładniczą [5].

Gdy rozkład $\rho(\eta)$ posiada dziedzinę nieograniczoną (ang. *infinite domain*, *infinite support*), wówczas nie istnieje η_{max}, dla którego prawdopodobieństwo wystąpienia $\eta > \eta_{max}$ wynosiłoby 0. W takich układach istnieje skończone

prawdopodobieństwo, że uzyskamy dowolną $\eta > 0$, ponieważ $\eta_{max} = \infty$. Z tego powodu całka znana ze wzoru (3.21) jest osobliwa, gdy $\eta = C$, dlatego do przeprowadzania obliczeń ją uwzględniających wykorzystywane są numeryczne symulacje. Dla przypadku $\rho(\eta) = e^{-\eta}$ na ich podstawie otrzymano wniosek, że średni czas t potrzebny na pojawienie się w sieci węzła o bardzo dużej wartości η jest porównywalny z $\frac{1}{\rho(\eta)} = e^{\eta}$, stąd przyjmuje się, że $\eta_{max} \sim \ln(t)$. Dlatego też

$$\sum_j k_j \cdot \eta_j = \sum_j k_j(t,t_j,\eta_j) \cdot \eta_j \leqslant \eta_{max} \cdot \sum_j k_j(t,t_j,\eta_j) \sim \ln(t) \cdot \sum_j k_j(t,t_j,\eta_j).$$

Przy uwzględnieniu założenia, że $\beta(\eta) < 1$, czyli tego, że średni stopień $k_j(t,t_j,\eta_j)$ nie rośnie szybciej niż czas t otrzymujemy

$$\sum_j k_j(t,t_j,\eta_j) \cdot \eta_j \leqslant \ln(t) \cdot Dt,$$

gdzie D jest pewną stałą. Według artykułu [5] autorstwa Bianconi i Barabásiego, jeśli przyjmiemy, że gdy $t \to \infty$

$$\frac{\sum_j k_j(t,t_j,\eta_j) \cdot \eta_j}{D\ln(t)t} \to Dm$$

oraz wykorzystamy wzór (3.2), gdzie stopień początkowy węzła $k_i(t_i)$ zastąpi liczbę tworzonych połączeń m, to otrzymamy

$$k_i(t,t_i,\eta_i) = k(t_i) \cdot \left(\frac{\ln(t)}{\ln(t_i)}\right)^{\frac{D}{\eta_i}}. \tag{3.51}$$

Za pomocą symulacji numerycznych zbadane zostało, że w takim przypadku rozkład $P(k)$ przyjmuje formę funkcji „rozciągniętej wykładniczo" (ang. *stretched exponential*) [5]. Funkcja ta znana jest również jako „komplementarna" do dystrybuanty $F(x)$ rozkładu Weibulla (ang. *complementary cumulative distribution function (CCDF) of Weibull distribution*), a jej wzór ogólny ma postać

$$\phi(x) = 1 - F(x) = \begin{cases} \exp\left(-\left(\frac{x}{\lambda}\right)^{\xi}\right), & x \geqslant 0 \\ 0, & x < 0 \end{cases}, \tag{3.52}$$

gdzie $\xi > 0$ jest parametrem kształtu, a $\lambda > 0$ parametrem skali. Klasyczny rozciągnięty zanik wykładniczy (ang. *classical stretched exponential decay*) jest obserwowany, gdy $\xi \in (0,1)$ [10].

Zatem gdy $\rho(\eta)$ ma nieograniczoną dziedzinę, wówczas węzły posiadające największą wartość parametru różnorodności adaptacyjnej η zdobędą bardzo dużą – w porównaniu do pozostałych węzłów – liczbę połączeń, co w konsekwencji obrazuje zjawisko znane pod nazwą „zwycięzca bierze wszystko" (ang. *winner-takes-all phenomenon*) [7]. Rozkłady gamma, log-normalny i potęgowy są rozkładami o nieograniczonej dziedzinie, dlatego może być w nich zauważalny efekt gromadzenia większości połączeń przez jeden bądź kilka wierzchołków.

Literatura

[1] Albert A., Barabási A.L., *Topology of Evolving Networks: Local Events and Universality*, „Physical Review Letters", 1999, vol. 85(24), p. 5234–5237.

[2] Barabási A.L., *Network Science. Evolving Networks*, http://networksciencebook.com/chapter/6, [dostęp: 30.10.2023 r.].

[3] Barabási A.L., *Network Science. The Barabási-Albert Model*, http://networksciencebook.com/chapter/5, [dostęp: 30.10.2023 r.].

[4] Barabási A.L., Albert R., *Emergence of Scaling in Random Networks*, „Science", 1999, vol. 286, p. 509–512.

[5] Bianconi G., Barabási A.L., *Competition and Multiscaling in Evolving Networks*, „Europhysics Letters", 2001, vol. 54(4), p. 436–442.

[6] Borgs C., Chayes J., Daskalakis C., Roch S., *First to Market is not Everything: an Analysis of Preferential Attachment with Fitness*, „STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing", 2007, p. 135–144.

[7] Chen G., Wang X., Li X., *Fundamentals of Complex Networks: Models, Structures, and Dynamics*, Newark: Wiley, 2014, p. 126.

[8] Erdős P., Rényi A., *On Random Graphs I*, „Publicationes Mathematicae Debrecen", 1959, vol. 6, p. 290–297.

[9] Fronczak A., Fronczak P., *Świat sieci złożonych. Od fizyki do Internetu*, Wydawnictwo Naukowe PWN SA, Warszawa 2021.

[10] Kim Y., Weon B.M., *Stretched Exponential Dynamics in Online Article Views*, „Frontiers in Physics", 2021, vol. 8:619729.

[11] Milojević S., *Power Law Distributions in Information Science: Making the case for logarithmic binning*, „Journal of the American Society for Information Science and Technology", 2010, vol. 61(12), p. 2417–2425.

[12] Newman M., Barabási A.L., Watts D.J., *The Structure and Dynamics of Networks*, Princeton: Princeton University Press, 2006, p. 340–342.

[13] Sayama H., *17.5: Degree distribution*, [W:] *Introduction to the Modeling and Analysis of Complex Systems*, Binghampton University, State Univ. of New York, 2024.

[14] Sun J., Staab S., Karimi F., *Decay of Relevance in Exponentially Growing Networks* WebSci '18: Proceedings of the 10th ACM Conference on Web Science, p. 343–351, 2018.

[15] *Total number of Websites*, https://www.internetlivestats.com/total-number-of-websites/, [dostęp 30.10.2023 r.]