

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI I INŻYNIERII
BIOMEDYCZNEJ

KATEDRA AUTOMATYKI I INŻYNIERII BIOMEDYCZNEJ



ROZPRAWA DOKTORSKA

***MODELOWANIE WIELOAGENTOWE WYBRANYCH ZACHOWAŃ
W SYSTEMACH SOCJO-EKONOMICZNYCH***

CEZARY PISKOR-IGNATOWICZ

PROMOTOR:

prof. dr hab. inż. Ewa Dudek-Dyduch

KRAKÓW, 2019

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI I INŻYNIERII
BIOMEDYCZNEJ

KATEDRA AUTOMATYKI I INŻYNIERII BIOMEDYCZNEJ



PHD THESIS

MULTI-AGENT MODELLING FOR SELECTED BEHAVIOUR IN
SOCIO-ECONOMICAL SYSTEM

CEZARY PISKOR-IGNATOWICZ

SUPERVISOR:

prof. dr hab. inż. Ewa Dudek-Dyduch

KRAKÓW, 2019

Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystycznego wykonania albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.): „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej «sądem koleżeńskim».”, oświadczam, że niniejszą pracę dyplomową wykonałem osobiście i samodzielnie i że nie korzystałem ze źródeł innych niż wymienione w pracy.

*Serdecznie dziękuję prof. dr hab. inż.
Ewie Dudek-Dyduch za poświęcony
czas i wielką życzliwość, dr Markowi
Zacharze za wspólną pracę i wysiłek
włożony w kodowanie środowiska do
symulacji.*

Streszczenie

Praca dotyczy tworzenia nowych modeli procesów socjo-ekonomicznych, ich analizy i badań eksperymentalnych. Jest poświęcona tworzeniu nowego modelu sieci społecznej i w oparciu o niego nowego modelu procesów kształtowania się cen.

Praca ma dwa zasadnicze, wzajemnie powiązane cele. Pierwszym jest stworzenie i zbadanie dynamicznego, parametrycznego modelu sztucznej sieci społecznej (SSS). Drugim jest skonstruowanie, na bazie opracowanego modelu SSS, nowego, bliższego rzeczywistości modelu współdziałania procesów decyzyjnych kupujących i sprzedawców i ostatecznie procesu kształtowania cen, a następnie przeprowadzenie analizy rozkładów **rzeczywistych** cen w sklepach internetowych i porównanie z otrzymanymi z modelu. **Teza pracy** składa się z dwóch wzajemnie powiązanych części i jest następująca:

Można wykazać, że obserwowane w sklepach internetowych rozkłady cen nie mają charakteru Gaussowskiego, lecz wskazują na ciężkoogonowy charakter rozkładu w kierunku wyższych cen a także można stworzyć bliski rzeczywistości, wieloagentowy model procesów zakupowych i zweryfikować go poprzez porównanie rozkładów cen rzeczywistych i uzyskanych z modelu.

Najważniejsze osiągnięcia w pracy są następujące.

1. Opracowanie nowego, parametrycznego, dynamicznego modelu sztucznej sieci społecznościowej, poprzez wprowadzenie nowych, zróżnicowanych cech osobowości agentów oraz nowych mechanizmów modelujących procesy interakcji pomiędzy agentami, co istotnie wpływa na tworzenie i dynamikę relacji znajomości (węzłów pomiędzy agentami)
2. Opracowanie parametrycznego, dynamicznego modelu współdziałania procesów decyzyjnych agentów kupujących i sprzedających (zawierającego w sobie model sztucznej sieci społecznościowej), a szczególności:
 - a) opracowanie modelu zindywidualizowanych cech i mechanizmów działania agentów kupujących,
 - b) zamodelowanie mechanizmu decyzyjnego sprzedawców, uzależniającego zmiany ich cen i jakości od dynamicznie zmieniającego się zysku,
 - c) opracowanie i zamodelowanie procesu wzajemnego sprzężenia informacyjno-decyzyjnego pomiędzy klasami agentów kupujących i sprzedających.
3. Wykonanie analizy rozkładów cen dla dziesięciu losowo wybranych produktów w sklepach internetowych.
4. Przeprowadzenie weryfikacji generowanych w modelu rozkładów cen, przy pomocy rzeczywistych rozkładów cen obserwowanych w sklepach internetowych dla omawianych produktów.
5. Przedstawienie swoich przemyśleń dotyczących tematyki układów złożonych.

Opracowanie obu wspomnianych modeli oraz zamieszczone w pracy wnioski z kolejnych, przeprowadzonych badań eksperymentalnych potwierdziły prawdziwość tezy.

Abstract

The dissertation concerns the creation of new models of socio-economic processes, their analysis and experimental research. It is devoted to creating a new model of social network and based on it a new model of price formation processes.

The work has two main, interrelated goals. The first is to create and explore a dynamic, parametric model of the artificial social network (ASN). The second is to construct, on the basis of the developed ASN model, a new model of interdependence between decision-making processes of buyers and sellers and finally the price formation process, and then to analyze the distribution of real prices in online stores and compare with those obtained from the model.

The thesis of the dissertation consists of two interrelated parts and is as follows. It can be shown that the price distributions observed in online stores are not Gaussian, but they indicate the heavy-tailed nature of the distribution towards higher prices. It is also possible to create a multi-agent model of purchasing processes close to reality and verify it by comparing the actual and obtained price distributions.

The most important achievements of a doctoral dissertation are as follows.

1. Development of a new, parametric, dynamic model of an artificial social network by introducing new, diverse personality attributes of agents and new mechanisms modeling the processes of interaction between agents, which significantly affects the creation and dynamics of relationships
2. Development of a parametric, dynamic model of interaction of decision-making processes of buying and selling agents (including an artificial social network model), in particular:
 - a) developing a model of individualized features and mechanisms of action of buying agents,
 - b) modeling the decision-making mechanism of sellers, making changes in their prices and quality dependent on dynamically changing profit,
 - c) developing and modeling the process of mutual information and decision coupling between classes of buying and selling agents.
3. Performing an analysis of price distributions for ten randomly selected products in online stores
4. Perform verification of price distribution generated in the model, using actual price distribution observed in online stores for the products in question.
5. Presentation of own thoughts on the subject of complex systems

The development of the both models and the conclusions of the subsequent experimental studies carried out in a frame of the dissertation confirmed the truth of the thesis.

1.	Wstęp	10
1.1.	Geneza pracy	11
1.2.	Teza i cel pracy	12
1.3.	Koncepcja badań i zakres pracy	13
2.	Układy złożone	17
2.1.	Pojęcie układu złożonego	17
2.2.	Własności układów złożonych	24
2.2.1.	Samooorganizacja	24
2.2.2.	Synergia w układach złożonych	25
2.2.3.	Nieliniowość układów złożonych	26
2.3.	Proces kształtowania cen, jako przykład socjo-ekonomicznego procesu złożonego.	27
2.4.	Modelowanie układów złożonych	30
2.5.	Sieci złożone	31
3.	Reprezentacja układów złożonych za pomocą sieci złożonych	35
3.1.	Podstawowe pojęcia	36
3.2.	Odległości w sieciach złożonych i własność „małych światów”	38
3.3.	Rozkład stopni węzłów	39
3.4.	Współczynnik gronowania	43
3.5.	Sieci losowe Reniy’go Erdos’a	44
3.6.	Podsumowanie	46
4.	Modelowanie sieci społecznych	47
4.1.	Przegląd literatury	48
4.2.	Proponowany formalny model sieci społecznej	53
4.2.1.	Model agenta	54
4.2.1.1.	Lokalizacja i położenie agentów	55
4.2.1.2.	Osobowość agenta	56
4.2.1.3.	Charakterystyka relacji między agentami	57
4.2.2.	Dynamika rozwoju sieci	59
4.2.2.1.	Spotkania agentów	59
4.2.2.2.	Oslabianie siły relacji – brak spotkań	61

4.3.	Podsumowanie	62
5.	Koncepcja badań symulacyjnych sztucznych sieci	63
5.1.	Parametry modelu odpowiadające za powstawanie więzi między agentami	70
5.2.	Parametry modelu odpowiedzialne za rozpad więzi między agentami...	74
5.3.	Podsumowanie	76
6.	Model współdziałania procesów decyzyjnych kupujących i sprzedawców	78
6.1.	Ogólna koncepcja	78
6.2.	Model agenta sprzedawcy	81
6.2.1.	Obliczanie zysku sprzedawcy	82
6.2.2.	Algorytm decyzyjny zmieniający cenę	83
6.2.3.	Decyzja o bankructwie i pojawieniu się nowych agentów sprzedawców	85
6.3.	Modelowanie procesów decyzyjnych Agentów Kupujących	89
6.3.1.	Model agenta kupującego	89
6.3.2.	Etap 1 - podjęcie decyzji o potrzebie zakupu	91
6.3.3.	Etap 2 - wybór sprzedawcy	94
6.3.3.1.	Wiedza agenta kupującego	94
6.3.3.2.	Atrakcyjność zakupu	95
6.3.4.	Etap dokonywania zakupu i uwzględnienie nieaktualnej wiedzy	95
7.	Analiza rozkładu cen produktu w sklepach internetowych	97
7.1.	Metodyka analizy statystycznej pozyskanych danych	97
7.1.1.	Badania rozkładu cen czajnika „Philips HD 4646/00 Biały”	100
7.1.2.	Analiza rozkładu cen lodówki “Samsung RSH5ZLBG”	104
7.1.3.	Badania rozkładu cen monitora „BenQ 18,5 GL955A”	106
7.1.4.	Analiza statystyczna cen lodówki „Samsung RL55VTEBG”	110
7.1.5.	Badania statystyczne rozkładu cen komputera „Lenovo ThinkCenter M83 Tower (10BE001APB)”	112

7.1.6.	Analiza statystyczna rozkładu cen pralki „Electrolux EWT1266TLW”	115
7.1.7.	Analiza rozkładu cen telewizora „LG 50LF652V”	117
7.1.8.	Badania rozkładu cen odkurzacza „Zelmer Jupiter ZVC425HT”	119
7.1.9.	Analiza statystyczna rozkładu cen ekspresu „Nivona 767 Cafe Romatica”	121
7.1.10.	Badanie rozkładu cen klimatyzatora „LG G09Wl Artcool Stylist”	124
7.1.11.	Parametryczny opis funkcji gęstości prawdopodobieństwa rozkładu cen produktu	126
7.1.12.	Analiza statystyczna rozkładu cen w ramach jednej kategorii produktu	127
7.2.	Wnioski z przeprowadzonej analizy rozkładów cen	130
8.	Rozkłady cen produktów generowane przez model	132
8.1.	Badanie zgodności dwóch rozkładów	133
8.2.	Porównanie danych i wyników symulacji dla czajnika „Philips HD 4646/00 Biały”	134
8.3.	Porównanie rozkładu cen monitora „BenQ 18,5 GL955A” z rezultatem symulacji	135
8.4.	Porównanie rozkładu cen pralki „Electrolux EWT1266TLW” z rezultatem symulacji	136
8.5.	Porównanie rozkładu cen telewizora „LG 50LF652V” z rozkładem cen z symulacji	137
8.6.	Porównanie rozkładu cen odkurzacza „Zelmer Jupiter ZVC425HT” z rozkładem cen uzyskanym z symulacji	138
8.7.	Porównanie rozkładu cen klimatyzatora „LG G09Wl Artcool Stylist” z wynikami symulacji	138
8.8.	Podsumowanie otrzymanych rezultatów	139
9.	Podsumowanie	142
10.	Spis literatury	145

1. Wstęp

Praca dotyczy tworzenia nowych modeli procesów socjo-ekonomicznych, ich analizy i badań eksperymentalnych. Procesy te, mówiąc bardzo ogólnie, są dynamicznymi procesami, na które mają wpływ dwie podstawowe składowe: ekonomiczna i socjologiczna. Przykładami takich procesów, między innymi, może być powstawanie baniek cenowych i spekulacyjnych, powstawanie mód czy też kształtowanie się cen. **Niniejsza praca będzie poświęcona tworzeniu nowego modelu sieci społecznej i w oparciu o niego nowego modelu procesów kształtowania się cen.**

Składowa ekonomiczna rozważanych procesów mająca wpływ na decyzje zakupowe kupującego, zależy od dostępu do informacji i warunków otoczenia. W szczególności dotyczy to takich własności jak: cena i bezpośrednie koszty pozyskania danego dobra, koszty dodatkowe jak np. czas dojazdu, czy ewentualnie dostępność produktu. Składową tą można by utożsamiać z racjonalnymi działaniami jednostki dążącej do optymalizacji realizowanych celów. Warto w tym miejscu podkreślić, iż różne jednostki mogą odmiennie definiować stawiane sobie cele.

Składowa socjologiczna jest zależna od oddziaływania z innymi jednostkami i może wiązać się z poglądami i wyobrażeniami jednostek, które propagują się w danej społeczności w sposób nie zawsze zależny od aktualnego stanu otoczenia. Należy, bowiem zauważyć, iż propagacja zależy nie tylko od aktualnego stanu otoczenia, lecz często od tego, co się działo wcześniej (kwestia histerezy).

Przykładami typowo socjologicznego aspektu są: “efekt snoba” (ang. snob effect), moda, wybuchy paniki (w postaci wykupywania określonego dobra lub dóbr czy też niekontrolowanej wysprzedaży akcji) czy bojkot (ang. boycott).

Zjawisko znane pod nazwą „efektu snoba” jest zjawiskiem mikroekonomicznym polegającym na gwałtownym spadku popytu na dane dobro pośród ludzi o wyższym dochodzie wraz ze wzrostem popytu na to dobro u osób o niższych dochodach.

Zjawisko bojkotu polega na celowym powstrzymaniu się przed kupnem czy używaniem określonych produktów. Najczęściej jest wyrazem protestu o charakterze moralnym, ekologicznym czy politycznym.

Impulsem do wybuchu panicznej wysprzedaży akcji może być określona plotka, ale najczęściej impuls pochodzi od „wystarczająco dużej” fluktuacji, która wyzwala reakcję łańcuchową w układzie. Należy w tym miejscu podkreślić, irracjonalność tego typu kolektywnych zachowań, gdyż na przestrzeni godzin warunki ekonomiczne na ogół nie ulegają tak drastycznym zmianom. Jedynie globalne kataklizmy czy katastrofy niszczące kilkadziesiąt procent mocy produkcyjnej i generujące wielkie straty mogłyby spowodować skokową zmianę wartości przedsiębiorstwa czy całego sektora. Tymczasem w rzeczywistości jedyne, co się zmienia w przypadku stanowczej większości przypadków panicznej wysprzedaży akcji, to sposób postrzegania przez inwestorów opłacalności inwestycji.

Trochę inaczej ma się sprawa ze zjawiskiem panicznego wykupowania produktów, gdyż w tym przypadku udział plotek czy określonych poglądów jest o wiele większy. Niemniej kolektywny charakter zachowań jest wspólny dla obu procesów. Indywidualna ocena ryzyka czy samodzielne procesy podejmowania decyzji zostają zdominowane przez procesy naśladownictwa innych osobników z danej społeczności.

Szczególnie dobitnie rolę naśladownictwa widać w przypadku mód, których powstawanie i propagacja nieodzownie wiąże się z powielaniem zachowań, wyglądu czy posiadania określonych przedmiotów innych osobników.

1.1. Geneza pracy

Badania procesów socjoekonomicznych, prowadzone są przez badaczy z wielu obszarów, między innymi z ekonomii, fizyki, matematyki oraz sztucznej inteligencji a zwłaszcza kolektywnej inteligencji. Biorąc pod uwagę metodologię badań, można stwierdzić, że jest ona w poszczególnych obszarach odmienna.

Dotyczy to również badań zachowania się uczestników procesów rynkowych ze szczególnym uwzględnieniem procesów ustalania cen.

W dotychczasowych badaniach ekonomiści postulowali racjonalność działań jednostek. Co więcej, decyzje poszczególnych uczestników rynku traktowano, jako niezależne. Naturalną konsekwencją takiego postrzegania rynku były koncepcje o gaussowskim rozkładzie cen, historycznie mające swoje korzenie w pracach Louis’a Bachelier’a. Kolejnym modnym postulatem ekonomii jest pogląd o równowagowym

charakterze procesów rynkowych. Z tą bardzo szeroką klasą teorii wiąże się kolejna koncepcja, a mianowicie teoria jednej ceny.

Od tego czasu nauka zrobiła postępy w dwóch obszernych dziedzinach: obszarze informatyki, oraz teorii układów złożonych. Jeśli chodzi o informatykę, to niesamowicie zwiększyła się moc komputerów oraz zostały rozwinięte nowe technologie informatyczne, między innymi modelowanie i symulacje wieloagentowe skutkujące dynamicznym rozwojem kolektywnej inteligencji (Szuba 2011).

Natomiast rozwinięcie teorii układów złożonych, pociągnęło za sobą całkowicie odmienne metodyki symulacji. W szczególności metodyki te pozwalają na symulacje procesów występujących w złożonych układach, zawierających dziesiątki tysięcy elementów (jednostek). W rezultacie powstała nowa dziedzina nauki o układach złożonych.

W wyniku rozwoju wspomnianych dziedzin nauki dokonała się rewolucja w postrzeganiu świata. Co więcej, dynamiczny rozwój wspomnianych dziedzin pozwala na przeprowadzenie całkowicie nowych badań i zweryfikowanie dawnych hipotez. W szczególności, pozwala na stworzenie nowych modeli pozwalających bardziej wnikliwie analizować procesy zakupowe. Umożliwia też implementacje tych złożonych modeli z wykorzystaniem nowoczesnych technologii informatycznych i narzędzi oraz pokonanie bariery czasowej w przeprowadzaniu eksperymentów. Przytoczone fakty stanowią przyczynę i inspirację podjęcia przez autora badań przedstawionych w pracy.

1.2. Teza i cel pracy

Praca ma dwa zasadnicze, wzajemnie powiązane cele.

Pierwszym jest stworzenie i zbadanie dynamicznego, parametrycznego modelu wybranych aspektów sztucznego społeczeństwa (modelu sztucznej sieci społecznej).

Drugim jest skonstruowanie, na bazie opracowanego modelu sztucznej sieci społecznej, bliższego rzeczywistości modelu współdziałania procesów decyzyjnych kupujących i sprzedawców i ostatecznie procesu kształtowania cen, a następnie przeprowadzenie analizy rozkładów rzeczywistych cen w sklepach internetowych i porównanie ich z otrzymanymi z modelu.

W wielu współczesnych modelach ekonomicznych zakładane jest, że obserwowane rozkłady cen mają charakter rozkładu Gaussowskiego. Jest to wynikiem

przyjmowanego silnie uproszczonego modelu procesów zakupowych. Zdaniem autora, rozkład cen ma charakter inny niż Gaussowski, mianowicie rozkład ten ma charakter ciężkoogonowy (ang. heavy tailed distribution), jak np. rozkład potęgowy, logarytmiczno-normalny lub rozkład Weibulla. Definicja ciężkoogonowego rozkładu prawdopodobieństwa podana jest w rozdziale 3.

Zbadanie tego charakteru stanowi, zatem istotny cel, jaki postawił sobie doktorant.

Nowy, zaproponowany w pracy model ma uwzględniać więcej aspektów interakcji zachodzących podczas procesu decyzyjnego dotyczącego zakupów. Model ten jest przeznaczony do symulacji rozkładów cen w sklepach internetowych. Dla stworzenia tego nowego modelu konieczna jest, zatem analiza czynników mających wpływ na obserwowane rozkłady cen. Tak, więc pierwszym pośrednim celem tej części pracy jest dokonanie analizy czynników mających wpływ na obserwowane rozkłady cen. Drugim, pośrednim celem jest stworzenie modelu agenta wykorzystywanego w wieloagentowym systemie modelującym procesy zakupowe. Model ten następnie został zaimplementowany przez dr. Marka Zacharę.

W konsekwencji, teza pracy składa się z dwóch wzajemnie powiązanych części i jest następująca.

Można wykazać, że obserwowane w sklepach internetowych rozkłady cen nie mają charakteru Gaussowskiego, lecz wskazują na ciężkoogonowy charakter rozkładu w kierunku wyższych cen a także można stworzyć bliski rzeczywistości, wieloagentowy model procesów zakupowych i zweryfikować go poprzez porównanie rozkładów cen rzeczywistych i uzyskanych z modelu.

1.3. Koncepcja badań i zakres pracy

Koncepcja pracy związana jest z faktem, że jej autor jest fizykiem. Zatem, zarówno terminologia jak i podejście do modelowania układów socjo-ekonomicznych przedstawiane jest w niniejszej pracy w języku oddziaływań w systemach złożonych.

Wychodząc od ogólnego wprowadzenia w układy złożone, charakteryzujemy ich cechy i mówimy, że jako jeden ze sposobów ich reprezentacji można przyjąć sieci złożone.

Następnie opracowano parametryczny, wieloagentowy model sztucznej sieci społecznej (ang. artificial social network), uwzględniający szereg własności i

aspektów, które nie były brane pod uwagę we wcześniejszych modelach sieci złożonych opisanych w literaturze. Zaproponowany model stanowi środowisko dla modelowania szerokiego spektrum sieci złożonych. Praca przedstawia wykorzystanie tego środowiska do modelowania i analizy parametrycznej wybranych zachowań agentów w systemach socjo-ekonomicznych, a w szczególności procesu kształtowania cen.

Dla dalszych badań konieczna była analiza wpływu zmian wartości parametrów modelu na postać tworzonej sieci społecznościowej. Taką analizę eksperymentalną przeprowadzono w kolejnym etapie badań i do dalszych badań wybrano zbiór parametrów modelu, dla których charakterystyka sieci wynikowej jest najbardziej zbliżona do rzeczywistej.

Następnie opracowany został formalny model współdziałania procesów decyzyjnych agentów kupujących i agentów sprzedawców. Pierwszym etapem było zaproponowanie formalnego modelu sprzedawcy i zależności leżących u podstaw jego decyzji ekonomicznych. Kolejnym, zaproponowanie formalnego modelu kupującego oraz sposobu podejmowania decyzji zakupowych uwzględniających preferencje osobowościowe.

Wykorzystując modele obu powyższych typów agentów, szczegółowo przedstawiono parametryczny model współzależności między kupującymi i sprzedawcami. Model ten uwzględnia aspekty socjologiczne.

Ostatnimi, istotnymi dla potwierdzenia tezy pracy etapem było zbadanie charakteru rozkładów rzeczywistych cen w sklepach internetowych oraz weryfikacja modelu poprzez porównanie rozkładów cen rzeczywistych i uzyskanych z modelu.

Praca składa się z dziewięciu rozdziałów, bibliografii oraz skorowidzu oznaczeń.

Zakres pracy odpowiada przedstawionej logice badań i jest następujący.

Ponieważ każdy układ socjo-ekonomiczny można traktować, jako układ złożony (Allen 2009), dlatego treść **rozdziału drugiego** stanowi bazę dla pokazania, że proces kształtowania się cen jest procesem w odpowiednim układzie złożonym. W rozdziale tym jest przybliżone pojęcie układu złożonego oraz charakterystyka jego własności, takich jak samoorganizacja, synergia, nieliniowość i szereg innych. Podane są i przeanalizowane argumenty wskazujące, że układ wzajemnego oddziaływania kupujących i sprzedających jest socjo-ekonomicznym układem złożonym.

Przeanalizowano sposoby modelowania układów złożonych ze szczególnym uwzględnieniem sieci złożonych. Rozdział ten zawiera oryginalne opracowanie autora.

W **rozdziale trzecim** omówiono pojęcie i cechy sieci złożonych, które wykorzystywane są do reprezentowania układów złożonych, w zakresie wykorzystywanym w niniejszej pracy.

Rozdział czwarty przedstawia zaproponowany przez autora wieloagentowy, parametryczny model dynamiki rozwoju powiązań w sieci społecznej. Podano tu formalny model agenta uwzględniając cechy osobowości istotne do nawiązywania i utrzymywania kontaktów/przyjaźni. Przedstawiono też sposób modelowania dynamiki rozwoju sieci społecznościowej (jednowarstwowej).

W **rozdziale piątym** opisano przeprowadzone eksperymenty symulacyjne mające na celu analizę wpływu zmian wartości parametrów modelu na postać tworzonej sieci społecznościowej.

Rozdział szósty zawiera propozycję parametrycznego dynamicznego modelu współzależności procesów decyzyjnych kupujących i sprzedawców. Podano rozszerzony model agenta kupującego uwzględniający dodatkowe cechy natury psychologicznej istotne dla decyzji o zakupie i wyborze sprzedawcy, model procesów decyzyjnych agentów sprzedawców uwzględniający decyzje dotyczące zmiany cen oraz decyzje dotyczące ewentualnych bankructw. Uwzględniono też możliwości pojawiania się nowych sprzedawców w przypadku koniunktury.

W **rozdziale siódmym** zaprezentowano wyniki analizy danych dotyczących rozkładów cen dziesięciu, losowych produktów z portalu Ceneo.pl, wybranych w ten sposób, by były sprzedawane w przynajmniej kilkunastu, (co najmniej szesnastu) różnych sklepach internetowych.

Celem analizy jest zbadanie charakterystyk rozkładów cen rzeczywistych produktów. Otrzymane wyniki są następnie wykorzystywane do weryfikacji zgodności rozkładów cen uzyskiwanych w modelu Piskor-Ignatowicza-Zachary z rozkładami cen realnych produktów (patrz rozdział dziewiąty).

Dla każdego produktu przetestowano hipotezę o gaussowskim rozkładzie cen, przy użyciu testu Shapiro-Wilka otrzymując niezgodność rozkładu rzeczywistych cen z rozkładem Gaussa. Przetestowano też hipotezy o ciężkoogonowym charakterze

rozkładu cen sprawdzając hipotezy, o rozkładach cen typu potęgowego, logarytmiczno-normalnego, wykładniczego oraz Weibulla.

Rozdział ósmy zawiera wyniki symulacji przeprowadzonych przy użyciu zaproponowanego przez autora wieloagentowego modelu kształtowania cen w procesach kupna - sprzedaży. Zawiera też porównanie otrzymanych z symulacji wyników z rzeczywistymi rozkładami cen produktów. Porównanie przeprowadzono przy pomocy testu Kołmogorowa-Smirnowa dla dwóch prób (dystrybuant empirycznych) (ang. Two-sample Kolmogorov-Smirnov test).

Rozdział dziewiąty stanowi podsumowanie otrzymanych w pracy rezultatów badań oraz zawiera plan dalszych prac badawczych.

2. Układy złożone

Można zaryzykować stwierdzenie, że każdy układ socjo-ekonomiczny jest układem złożonym (Allen 2009), dlatego treść tego rozdziału stanowi bazę dla pokazania, że proces kształtowania się cen jest procesem w odpowiednim układzie złożonym. W rozdziale jest przybliżone pojęcie układu złożonego oraz charakterystyka jego własności, takich jak samoorganizacja, synergia, nieliniowość i szereg innych. Podane są i przeanalizowane argumenty wskazujące, że układ wzajemnego oddziaływania kupujących i sprzedających jest socjo-ekonomicznym układem złożonym. Przeanalizowano sposoby modelowania układów złożonych ze szczególnym uwzględnieniem sieci złożonych.

Stworzony w ramach pracy wieloagentowy model jest przykładem modelu układu złożonego, którego właściwości będą analizowane za pomocą reprezentacji sieci złożonej.

Rozdział zawiera oryginalne opracowanie autora dotyczące układów złożonych, będące wynikiem jego przemyśleń i wiedzy z różnych obszarów.

2.1. Pojęcie układu złożonego

Podobnie jak rozwiązanie problemu mostów w Królewcu przez Eulera (Euler 1736) wymagało odejścia od myślenia „analityczno-geometrycznego” i zaowocowało stworzeniem teorii grafów, tak i obecnie zrozumienie układów złożonych wymaga wyjścia poza mechanistyczno-redukcyjny sposób postrzegania świata (Pourbohloul and Kieny 2011). Klasyczne układy opisywano poprzez redukcję ich do prostszych komponentów dających się opisać w sposób kompletny. Oddziaływanie między komponentami tak rozumianego układu klasycznego zawsze daje się przedstawić, jako superpozycja oddziaływań typu komponent-komponent. O teoriach posiadających tę własność mówimy, że są abelowe (Babusci et al. 2013). Opis całego układu klasycznego uzyskiwano w wyniku ponownego „połączenia” opisów prostszych elementów.

Dla układów złożonych takie podejście jest niemożliwe, gdyż wiele z ich własności nie wynika z własności poszczególnych elementów, lecz jest wynikiem oddziaływań między elementami układu. Właściwości elementów są modyfikowane

przez oddziaływania, a i same oddziaływania ulegają modyfikacji. Teorie opisujące modyfikowanie oddziaływań przez oddziaływania nazywane są teoriami nie-abelowymi. Przykładem takiej teorii może być chociażby chromodynamika kwantowa, a przykładem procesu, który ulega modyfikacji w wyniku oddziaływania trzech ciał - proces produkcji mezonu η w reakcji zderzenia deuteron-proton (Piskor-Ignatowicz et al. 2007). O ile z modyfikacją oddziaływań fizycy się uporali poprzez formalizm szeregów Faddeev'a (Skibiński et al. 2017), o tyle, nadal nie stworzono formalizmu pozwalającego opisać zmiany właściwości oddziałujących ze sobą elementów systemu złożonego. Warto tu wspomnieć, iż formalizm szeregów Faddeev'a jest szalenie skomplikowany, a przez to praktyczne zastosowanie go do układów złożonych na obecnym etapie jest niemożliwe. Wymusza to konieczność poszukania nowych sposobów opisu (Cilliers 1998).

Do obecnej chwili nie wypracowano ogólnie przyjmowanej definicji „układów złożonych”, pozwalającej wyróżnić je z mnogości innych układów. Trudność zasadza się w interdyscyplinarnym charakterze tego pojęcia, gdyż teoria opisuje tak różne układy jak sieci metaboliczne w żywych komórkach (Barabási and Oltvai 2004), społeczności (M. E. J. Newman and Park 2003), miasta (Crucitti, Latora, and Porta 2006), języki (Yu, Liu, and Xu 2011) czy układy ekonomiczne (Rendón de la Torre et al. 2016).

Nie można też wskazać żadnej cechy budowy, która wyróżniałaby układy złożone, a więc patrząc na budowę układu, bez znajomości oddziaływań między jego składnikami, nie jesteśmy w stanie orzec o jego złożoności. Zauważmy, iż skomplikowanie budowy układu wcale nie oznacza, że układ jest układem złożonym. Doskonałym przykładem może być komputer, którego budowa jest skomplikowana a który wcale nie jest układem złożonym! Układy złożone mogą mieć praktycznie dowolną postać. Może to być sieć złożona, jak w przypadku sieci neuronów w żywym mózgu, czy sieci połączeń między komputerami w przypadku Internetu, ale również może być to człowiek, społeczeństwo, ekosystem, układ socjo-ekonomiczny czy turbulentny przepływ cieczy (Xiao, Wang, and Ghanem 2017). Dyskusję problemów napotykaną przy próbach definicji układu złożonego można znaleźć w pracy „What is a Complex System?” (Ladyman, Lambert, and Wiesner 2011). W ocenie autora, mimo iż wymieniona powyżej praca, daleka jest od kompletności, to stanowi dobrą ilustrację zarówno trudności formalnych napotykaną przy próbach stworzenia

ściślej definicji układu złożonego, jak i mnogości i różnorodności podejść zarówno do opisu jak i rozumienia układów złożonych.

Badacze zajmujący się zagadnieniami układów złożonych są jednak zgodni w kwestii, iż układy te mieszczą się pomiędzy deterministycznym porządkiem reprezentowanym przez kryształy czy dowolne klasyczne układy mechaniczne, a deterministycznym chaosem (np. gaz, układy trzech lub więcej ciał o podobnych masach związanych grawitacyjnie itd.).

Układy złożone cechuje adaptatywność, elastyczność, autonomiczność, dynamiczna stabilność (Leng, Lin, and Kurths 2016), ale też nieprzewidywalność zmian jego stanu. Te właściwości układów złożonych mogą być postrzegane, jako efekt ich zdolności do samoorganizacji i optymalizacji. Poszczególne elementy układu złożonego nie są ani w pełni od siebie zależne, ani nie są w pełni niezależne (Heylighen 2009). Aby zilustrować kluczową rolę oddziaływań dla złożoności układu można rozważyć żywy mózg ssaka i ten sam mózg w jakiś czas po śmierci osobnika umieszczony w roztworze formaliny. O ile w przypadku żywego organizmu, mamy do czynienia z układem złożonym, o tyle ten sam mózg tylko martwy nie przejawia już cech układu złożonego – jego budowa jest skomplikowana, ale nie jest to już układ złożony. Powodem jest fakt, że wraz ze śmiercią zanikły takie cechy układu jak zdolność do przetwarzania informacji, pamięć, kontekstowość reakcji na bodźce, neuroplastyczność i wiele innych, będących wynikiem oddziaływań. Innym przykładem może być Internet. Sama sieć połączeń między komputerami pozostawiona „sama sobie” była by skomplikowana, ale nic ponad to. Dopiero ludzkie lokalne działania takie jak naprawy popsutych komponentów, wymiany elementów systemu na nowsze w odpowiedzi na zmieniające się potrzeby użytkowników czy tworzenie nowych połączeń decydują o złożoności systemu. A więc to różnorodność i często kontekstowość lokalnych oddziaływań decyduje o tym czy dany system będzie systemem złożonym, czy tylko skomplikowanym.

Należy w tym miejscu zwrócić uwagę na jeszcze jedną istotną różnicę między układami klasycznymi a złożonymi, a mianowicie jak rozumiana jest „lokalność” w każdym z tych układów. W przypadku układów klasycznych lokalność jest ściśle związana z geometryczną odległością, atomy czy cząsteczki w kryształach oddziałują na odległościach rzędu rozmiaru pojedynczych atomów, bardzo podobne skale odległości odpowiadają oddziaływaniom w cieczach czy gazach w trakcie zderzeń ich

komponentów. Tymczasem dla układów złożonych „lokalność” oddziaływania rozumiana jest w sensie istnienia jakiejś relacji między jego składnikami, które geometrycznie (w skali odległości właściwej dla danego obiektu) mogą znajdować się bardzo daleko. Na przykład: może to być połączenie między dwoma neuronami odległymi o wiele centymetrów, co w skali rozmiaru somy (ciała komórki nerwowej) przekracza cztery rzędy wielkości. Może to być przekaz informacji między znajomymi osobami żyjącymi na dwu różnych kontynentach. Tak, więc „lokalność” oddziaływania w przypadku układów złożonych nie musi mieć żadnego związku z geometrycznym położeniem obiektów a odnosi się do takich własności jak „mój bezpośredni znajomy”, „neuron, z którym mam bezpośrednie połączenie”, „sprzedawca, u którego kupiłem towar” (i może on znajdować się zarówno w Chinach jak i parę przecznic od naszego domu).

Stosunkowo niedawno pojawiło się nowe narzędzie, jakim jest entropia permutacyjna (ang. permutation entropy) – stanowiąca naturalną miarę złożoności układów dynamicznych. Znajduje ona zastosowanie dla szeregów czasowych i pozwala rozróżnić układy klasyczne od układów złożonych, a te z kolei od układów chaotycznych (Bandt and Pompe 2002). Można by, więc wykorzystać ją, do ilościowego określenia złożoności układu. Jednakże do chwili obecnej tego nie uczyniono. Współczesny przegląd prac z tematyki poświęconej zastosowaniom entropii permutacyjnej w analizie układów złożonych znajduje się w artykule pod tytułem: „Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review” (Zanin et al. 2012).

To, co wyróżnia układy złożone to ich holistyczny charakter – to jest przejawiają one właściwości, których nie posiadają poszczególne składniki tych układów. Jako przykład można podać sieć neuronów – jej właściwości nie da się wywieść z własności pojedynczego neuronu, bo właściwości sieci będą zależały od sposobu połączeń między neuronami (i ich wzajemnych oddziaływań). Budując dwie sieci neuronowe o takiej samej ilości neuronów i nawet takiej samej liczbie połączeń między nimi, ale o różnej strukturze otrzymamy dwie sieci o kompletnie innych właściwościach. Innym przykładem może być ludzka społeczność. Jej właściwości i procesy w niej zachodzące nie są wyłącznie konsekwencją struktury psychologicznej i umysłowej osobników, lecz zależą od wzajemnych oddziaływań. Wybierając dowolną osobę zaobserwujemy kompletnie inne jej zachowania w zależności od aktualnych

oddziaływań innych osób. Ta sama osoba wchodząc na spotkanie z grupą nieznaną sobie osób będzie zachowywała się zupełnie inaczej niż na spotkaniu z przyjaciółmi. Ta kontekstowość zachowań osobników sprawia, że również cała grupa będzie się zachowywać różnie w zależności od oddziaływań między poszczególnymi osobami nawet w dokładnie tych samych warunkach otoczenia.

Holistyczny charakter układów złożonych przejawia się również przez emergentne ich właściwości, to jest „pojawiania się” nowych jakościowo zjawisk, procesów czy struktur wraz ze wzrostem rozmiaru układu. Jako przykład może posłużyć klimat. Startując od układu o rozmiarze małego akwarium i mając poszczególne komponenty w odpowiednich proporcjach nie zaobserwujemy pojawiania się chmur, prądów ciepłych i zimnych, wiatru, czy burz. Gdy rozmiar naszego „laboratorium” wzrośnie do rozmiarów rzędu $3.6 \cdot 10^6 m^3$ pojawi się, możliwość tworzenia chmur (KSC 2007). Jakiego rozmiaru powinien być układ, aby mogła wystąpić burza nie wiadomo, za to wiemy, że aby obserwować klimat potrzebujemy obiektów porównywalnych z kulą ziemską. Warto przy tej okazji powiedzieć o „niedopatrzeniu”, jakie zrobiono w pierwszej fazie eksperymentu myślowego, zarówno akwarium jak i „budynek” umieszczono na Ziemi. Gdyby naprawdę zależało nam na zaobserwowaniu, co dzieje się z układem wraz ze zmianą jego rozmiaru, należałoby go umieścić w pustej przestrzeni w odległości od Słońca takiej jak znajduje się Ziemia. A wówczas nawet gdyby nasz układ miał rozmiar Księżyca to wciąż nie mógłby mieć własnej trwałej atmosfery. Ponieważ do tego by gazy znajdujące się w pobliżu ciała niebieskiego tworzyły jego atmosferę prędkość ruchów termicznych tego gazu nie może przekraczać prędkości ucieczki, (czyli minimalnej prędkości potrzebnej do trwałego opuszczenia pola grawitacyjnego ciała niebieskiego).

Należy podkreślić, że gdy rozważamy układy złożone, rozmiar nie koniecznie oznacza rozmiar geometryczny. Przykładowo społeczeństwo dowolnego państwa, w którym następuje gwałtowny przyrost liczby ludności, pomimo, iż obszar, na którym ludność ta zamieszkuje pozostaje stały. Wzrost rozmiaru układu wiąże się w tym wypadku ze wzrostem liczby oddziaływujących ze sobą osobników, co z kolei wymusza powstanie nowych sposobów komunikacji, nowych struktur organizacyjnych, nowych sposobów rozwiązania problemów transportu czy zaopatrzenia, i tak dalej. Innym przykładem może być gospodarka jakiegoś państwa,

jej gwałtowny wzrost będzie wiązał się z takimi wielkościami jak wzrost produkcji, handlu, przepływu informacji, zmianami technologicznymi czy organizacyjnymi. Rozmiar w tym wypadku będzie się wiązał ze wzrostem liczby i złożoności oddziaływań między firmami, bankami, ośrodkami badawczymi a nie z wymiarem geometrycznym czy liczebnością żyjącego w dany państwie społeczeństwa.

Szczególnym typem emergentnych własności układu złożonego jest pojawianie się hierarchii. Dotyczy ono zarówno oddziaływań jak i powstających w tym układzie struktur. Charakter relacji/oddziaływania między dużymi zbiorami komponentów układu może mieć inną naturę czy mechanizm oddziaływania niż w przypadku pojedynczych składników. I tak, czym innym są reakcje biochemiczne zachodzące w organellach komórkowych, czym innym jest komunikacja na poziomie samej komórki, jeszcze inny charakter ma komunikacja pomiędzy organami w ciele i wreszcie komunikacja między osobnikami danego gatunku. Innym przykładem mogą być relacje między pracownikami w firmie należącymi do jednego zespołu, inne formy będą przybierały oddziaływania między zespołami w danej firmie, a jeszcze, czym innym będą oddziaływania firma-firma.

Hierarchiczność ściśle wiąże się ze zmianą stopnia szczegółowości, najniższy poziom jest najbardziej szczegółowy. Na przykład pracownik wykonujący jakąś część do komponentu służącego do budowy samolotu, nie zna projektu całego samolotu. Na tym poziomie istotne jest wykonanie detalu z zadaną precyzją i z odpowiedniego materiału. Na kolejnym poziomie poszczególne części łączone są w większy komponent, spada, więc szczegółowość. Istotne są całe części, oraz jak i gdzie mają być połączone. Wreszcie na poziomie nabywcy samolotu wszystkie pośrednie informacje nie mają znaczenia. Dla kupującego istotnymi parametrami będzie cena samolotu, koszty eksploatacji, niezawodność oraz specyficzne charakterystyki zależne od celów, do jakich ma być stosowany.

Układy złożone przeważnie są układami otwartymi to znaczy wymieniają z otoczeniem energię lub informację. Z termodynamicznego punktu widzenia procesy w nich zachodzące mają charakter nierównowagowy (Ye et al. 2015). Co więcej zachowanie tych układów zależy od historii, to znaczy stan, jaki przyjmie układ złożony pod wpływem zewnętrznego bodźca zależy od wcześniejszych stanów tego układu. Poszczególne elementy systemu złożonego nie mają dostępu do informacji o stanie układu, jako całości, ich oddziaływanie ma charakter lokalny. Jednakże stan

systemu, jako całości może ulec zmianie w wyniku takich lokalnych oddziaływań. Jako przykład można podać rozprzestrzenienie się pandemii HIV, wirusa, który powstał na początku XX wieku w Afryce Północnej, a w latach 1969-1973 dotarł z Karaibów do USA (Z. Gong, Xu, and Han 2017). W tym przypadku lokalnym oddziaływaniem były kontakty seksualne między ludźmi. Początkowo ilość zakażonych osób była bardzo niewielka, jednakże dobre przystosowanie wirusa w połączeniu z ryzykownymi zachowaniami seksualnymi ludzi, umożliwiło propagację choroby. O ile rozprzestrzenianie się infekcji odbywa się przez oddziaływania lokalne, o tyle skutki mają charakter globalny (Ortblad, Lozano, and Murray 2013), taki jak zmiany demograficzne, zmiany społeczne i kulturowe, gwałtowny rozwój badań nad retrowirusami, rozwój metod diagnostycznych, tworzenie organizacji mających na celu pomoc osobom dotkniętym chorobą czy organizacja akcji medialnych i społecznych mających na celu zwiększenie świadomości społecznej związanej z chorobą jak i redukcję ryzykownych zachowań prowadzących do zakażeń.

Jedną z bardziej fascynujących cech układów złożonych jest samoorganizacja. Przejawia się ona poprzez pojawianie się struktur lub porządku wyższego rzędu z niskopoziomowych oddziaływań. Jednakże samoorganizacja występuje też w układach klasycznych – nie jest, więc cechą wyróżniającą układy złożone. Jako przykład zachowania układów klasycznych można podać proces krystalizacji, gdzie niskopoziomowe oddziaływania między cząstkami prowadzą do powstania struktury krystalicznej. Aczkolwiek o wiele ciekawsze są struktury produkowane przez niskopoziomowe oddziaływania w układach złożonych. Wytwarzają one wielkie bogactwo form, takich jak chociażby Internet, sieci pokarmowe, reakcje oscylacyjne w chemii, ławice ryb, roje owadów, tworzenie stad w biologii i mnóstwo innych. Odmienny jest też charakter struktur obserwowanych w przypadku układów złożonych. Struktury układów klasycznych mają o wiele bardziej „statyczny” charakter. Raz uszkodzony kryształ takim już pozostanie. Tymczasem tkanka żywego organizmu, sieć połączeń Internet czy struktura organizacyjna w społeczeństwie (o ile uszkodzenie nie jest zbyt masywne), będzie się odnawiać. Własność niskopoziomowych oddziaływań między składnikami układu złożonego pozwala na

reprezentowanie tych układów przy pomocy sieci złożonych, czemu poświęcony jest rozdział 3.

2.2. Własności układów złożonych

Ponieważ do chwili obecnej brak jest ogólnie przyjętej definicji układu złożonego, zwykło się układy te charakteryzować przez wymienienie posiadanych przez nie cech oraz opis różnych przykładów układów klasyfikowanych, jako układy złożone. Ponieważ jest to ogólnie przyjęte postępowanie w literaturze poświęconej tematyce układów złożonych, autor niniejszej pracy również scharakteryzuje układy złożone opisując posiadane przez nie cechy i podając przykłady takowych układów.

2.2.1. Samoorganizacja

Jak wcześniej wspomniano jedną z własności układów złożonych, jest ich zdolność do samoorganizacji. A dokładniej do spontanicznej samoorganizacji, gdyż proces nie jest w żaden sposób sterowany z zewnątrz czy z wewnątrz układu. Co więcej dla odpowiednio dużych układów usunięcie dowolnego komponentu nie wpływa na przebieg procesu. Proces samoorganizacji jest procesem rozproszonym, równoległym. Sprawia to, iż sterowanie układami złożonymi jest niezmiernie trudne (Liu, Slotine, and Barabási 2011).

Ławice ryb, stada ptaków, roje owadów, komórki tworzące określoną tkankę to liczne przykłady kolektywnych zachowań w przyrodzie ożywionej (DeLellis et al. 2014). Część z szalenie skomplikowanych zachowań otrzymuje się w wyniku prostych reguł naśladownictwa. Inne zachowania jak budowanie termitier przez termyty czy poszukiwanie przez mrówki pożywienia, odbywa się przez pośrednie przekazywanie informacji przez system feromonów. Jeszcze inny sposób realizowany jest przez pszczoły, które, za pośrednictwem “tańca” przekazują sobie informacje o położeniu miodnych łąk czy drzew. Pośród mechanizmów wspólnych dla funkcjonowania tak różnych systemów można wyróżnić:

- istnienie sprzężeń zwrotnych dodatnich (ang. positive feedback)
- istnienie progu odpowiedzi (ang. response threshold) układu powodującego, że układ reaguje zmianą stanu/zachowania dopiero wtedy, gdy dany bodziec

przekroczy graniczną wartość, przy czym dla poszczególnych składników układu złożonego te wartości mogą i najczęściej są różne

- istnienie sprzężeń zwrotnych ujemnych (ang. negative feedback) (Sumpter 2006).

Jednakże należy podkreślić, że nie każdy układ przejawiający własność samoorganizacji jest systemem złożonym. Przykładem mogą być drobiny piasku na pustyni tworzące pod wpływem porywów wiatru o określonej częstości wielkopowierzchniowe regularne wzory. Podobnym przykładem mogą być wzory tworzone na dnie przez morskie fale.

2.2.2. Synergia w układach złożonych

Inną zdumiewającą własnością układów złożonych jest synergia. Jest to własność układu przejawiająca się tym, że oddziaływujące składniki mają pewną cechę lub cechy o większym natężeniu (przypadającym na jeden składnik) niż suma natężeń tej cechy dla nieoddziaływujących składników układu. Mówiąc inaczej oddziaływanie między składnikami układu jest źródłem dodatkowego wzmocnienia wybranych cech każdego ze składników układu.

Przykładem może być wzrost efektywności zdobywania pożywienia przez mrówki wraz ze wzrostem ilości owadów w mrowisku (Sumpter 2006). W przypadku mało licznej populacji więcej czasu spędzają te owady na poszukiwanie pożywienia. Przyczyną jest fakt, że nawet, gdy zostanie ono odnalezione przez pojedynczą mrówkę, ale pożywienie to, znajduje się daleko od mrowiska, to pozostawiony przez mrówkę ślad feromonowy prowadzący do tego pożywienia zdąży zniknąć, nim inne mrówki go odnajdą. Tak, więc zamiast zająć się transportem już odkrytego przez pojedynczą mrówkę pożywienia, nadal będą prowadzić jego poszukiwania. Po przekroczeniu pewnej krytycznej liczebności owady zdążają odnaleźć ślady feromonowe pozostawione przez inne osobniki i czas tracony na odszukanie pożywienia spada, bo więcej mrówek może się zająć transportem już odnalezonego pożywienia.

Jako inne przykłady może posłużyć: minimalizowanie strat cieplnych przez wzajemne ogrzewanie się osobników u zwierząt stadnych. Na przykład surykatki szare występujące na terenach pustynnych Afryki południowej, grzeją się wzajemnie w zimne noce (strata ciepła przypadająca na jednego osobnika w ogrzewającym się

wzajemnie stadzie jest mniejsza niż dla samotnej surykatki w tych samych warunkach). Podobnie, wzrost ilości zdobytego pożywienia przypadający na jednego drapieżnika, gdy poluje on w stadzie, w stosunku do ilości pożywienia zdobywanego samodzielnie (Corning 2007).

Podobnie większe zbiory neuronów są w stanie nauczyć się rozwiązać problemy niedostępne dla ich mniejszych zbiorowości (Pais-Vieira et al. 2015). Zjawisko synergii powszechnie występuje w ludzkim społeczeństwie gdzie dzięki postępującej specjalizacji i komplementarności posiadanych przez osobniki umiejętności systematycznie wzrasta efektywność pracy. Zagadnienie to można zilustrować na przykładzie zespołu informatyków czy naukowców. Taka grupa, dzięki współpracy jest w stanie zrealizować tak skomplikowane projekty, że ich wykonanie nawet dla najzdolniejszej jednostki jest nie możliwe nie tylko w tej samej ilości roboczogodzin, co wykonuje go grupa, ale w ogóle. Jest tak, gdyż zakres wiedzy i umiejętności, potrzebny do wykonania projektu znacząco przekracza możliwości jej opanowania przez każdego pojedynczego człowieka. Do istotnych cech układów złożonych należy też nieliniowość.

2.2.3. Nieliniowość układów złożonych

Nieliniowość obok samoorganizacji stanowi kolejną istotną cechę układów złożonych.

Najprościej można ją określić, jako występowanie skutków nie proporcjonalnych do wywołujących je efektów. Powodem występowania nieliniowości jest występowanie dynamicznych więzów nałożonych na zmienne opisujące układ, jak również występowanie sprzężeń zwrotnych zarówno dodatnich jak i ujemnych. Kolejną przyczyną jest występowanie progu reakcji poszczególnych elementów układu, przy czym często wysokość progu reakcji dla danego komponentu układu złożonego zależy od wcześniej przybieranych przez składnik układu stanów. Ten sam bodziec np. akt agresji w zależności od stanu układu (tutaj też wcześniejszych stanów), może zostać szybko stłumiony (pacyfikacja) albo prowadzić do lawinowych reakcji w całym układzie, takich jak gwałtowne rozruchy czy panika. Liczne dalsze przykłady z zakresu ekonomii i socjologii (w ujęciu historycznym) zamieszczone są w artykule Jos'e A. Scheinkman'a zatytułowanym „Social Interactions” (Scheinkman 2008). Natomiast współczesne podejścia do problematyki nieliniowości

przedstawiono w pracy „Community resilience: toward an integrated approach” autorstwa Fikret’a Berkes’a i Helen Ross (Berkes and Ross 2013).

Nieliniowość oddziaływań w układach złożonych sprawia, iż ich ewolucja w znacznym stopniu jest nieprzewidywalna. Występowanie dodatnich sprzężeń zwrotnych powoduje, że nawet niewielkie zmiany lokalne mogą się rozpropagować na cały układ. Z kolei ujemne sprzężenia zwrotne odpowiadają za stabilizację układu, przez co wykazuje on pozornie przeciwstawne cechy: bardzo dużą trwałość i adaptacyjność.

2.3. Proces kształtowania cen, jako przykład socjo-ekonomicznego procesu złożonego.

W punkcie tym przedstawię argumenty uzasadniające, że układ kupujących i sprzedawców jest socjo-ekonomicznym systemem złożonym.

Historycznie, procesy kupna sprzedaży ekonomiści traktowali, jako proces zachodzący w układzie będącym w stanie równowagi. Dodatkowo w pierwotnie tworzonych modelach zarówno sprzedawcy jak i kupujący byli homogeniczni i mieli stały, natychmiastowy dostęp do kompletnej informacji o rynku. Co więcej, decyzje zakupowe były zawsze w pełni racjonalne (*homo economicus*). Nie jest, więc dziwnym, iż tak nierealistycznie zdefiniowane modele prowadziły do tak egzotycznych koncepcji jak teoria jednej ceny (*ang. law of one price*).

Nie może też zaskakiwać, że w tak budowanych modelach nie jest możliwym wytłumaczenie takich procesów jak tworzenie mód, zjawisko bojkotu, powstawanie baniek cenowych czy spekulacyjnych, czy wreszcie wybuchów paniki w postaci wykupywania określonych towarów czy wysprzedawania akcji.

Kumulowanie się obserwacji niedających się wytłumaczyć przez klasyczne koncepcje ekonomiczne zaowocowały poszukiwaniami alternatywnych sposobów postrzegania praw rynkowych. Coraz powszechniej kwestionuje się równowagowy charakter procesów zachodzących w ekonomii. Prace badawcze w tym zakresie szczególnie zintensyfikowano po kryzysie w 2008 roku. Dużo wcześniej zaczęto dostrzegać nieracjonalność zachowań kupujących w postaci chociażby zakupów impulsywnych (*ang. impulse purchase*) skutkujących często kupnem dóbr czy usług nieprzydatnych nabywcy. Stoi to w jawnej sprzeczności z założeniem o pełnej racjonalności decyzji zakupowych. Inną narzucającą się obserwacją jest

zróżnicowanie (heterogeniczność) zarówno nabywców jak i sprzedawców ze względu na ich umiejętności, możliwości czy preferencje. Znajdowało to od dawna swoje odzwierciedlenie w praktyce, w postaci reklam czy akcji marketingowych nakierowanych na konkretne grupy docelowe (takie jak: dzieci, młodzież, osoby starsze, miłośnicy SF i tak dalej). Zagadnienia związane z zachowaniem osób i podejmowaniem przez nich decyzji nabycia dobra lub usługi są przedmiotem badań ekonomii behawioralnej. Kolejnym kierunkiem rozwoju badań jest wpływ oddziaływań grupy lub grup na decyzje zakupowe jednostki. Zakupy dokonywane, „bo coś jest modne” czy też motywowane tym, że daną rzecz ma sąsiad (czy znajomi), mogą być przykładami procesu socjo-ekonomicznego. A więc takiego, gdzie nabycie określonej usługi czy dobra przestaje być niezależną decyzją jednostki, lecz posiada znaczącą składową zależną od oddziaływania z innymi osobami czy grupami osób.

W dalszym systematyzowaniu i dużo głębszym rozumieniu ekonomii, coraz istotniejszą rolę odgrywają dwie nowe dyscypliny naukowe: ekonofizyka i nauka o systemach złożonych. Dzięki ich rozwojowi stało się możliwym dostrzeżenie, iż wiele procesów rynkowych wcale nie ma charakteru równowagowego. Zaczęto rozumieć, że charakter zależności między poszczególnymi komponentami rynku jest odmienny od mechanizmu zegarka, bo poszczególne komponenty nie mają dobrze zdefiniowanych granic (są układami otwartymi). Pracownik nie jest „trybikiem” w maszynie, a więc to jak „działa” nie zależy tylko od tego, co robi w firmie, ale od całego kontekstu społecznego (rodziny, przyjaciół, znajomych). I znowu najpierw ta „wiedza” pojawiła się w formie praktycznej. Zaobserwowano, że odpowiednie pakiety socjalne podnoszą efektywność pracy. Pozornie, pracownik uprawiający sport będzie bardziej zmęczony i będzie miał mniej czasu, a tymczasem jest dokładnie odwrotnie. Proces uświadamiania sobie roli oddziaływań między ludźmi przebiegał powoli i był motywowany ekonomicznie. Mianowicie, poszukiwano takich form organizacji pracy w firmach, które obniżą koszty i zwiększą efektywność, tym samym podnosząc konkurencyjność firmy na coraz bardziej wymagającym rynku. Jednym ze skutków tej „cichej rewolucji” było tworzenie zwinnych metodyk zarządzania projektami (ang. agile project management). Metodyki te zorientowane są na tworzenie/uczenie pracowników metod pracy opartych na współpracy, a nie jak wcześniej konkurencji. Prowadzą też do przeniesienia nacisku z mierzenia indywidualnej efektywności na pomiar efektywności całego zespołu. Odchodzenie od tworzenia grup projektowych

czy badawczych na korzyść tworzenia zespołów, chyba najlepiej ilustruje zmianę sposobu postrzegania ze zorientowanego na pojedynczy składnik na taki, gdzie kluczową rolę odgrywają oddziaływania między komponentami. Nasuwa się tu analogia z efektem szybciej jest wolniej (ang. „faster is slower” effect) znanym z dynamiki ruch pieszych, polegającym na tym, że przykładowo w procesie ewakuacji „przepychanie się” jednostek dążących do jak najszybszego opuszczenia rejonu zagrożenia prowadzi do tworzenia zatorów i ostatecznego spowolnienia średniej grupowej prędkości opuszczania ewakuowanego obszaru (Garcimartín et al. 2014). Podobnie, w przypadku projektu „jak najszybsze” wykonanie zadania, bez komunikacji z pozostałymi twórcami projektu często prowadzi do konieczności ponownego przepisania fragmentu zadania tak by mógł on właściwie działać z kodem tworzonym później lub wymusza nieoptymalne dostosowywanie kodu pozostałych komponentów do już stworzonego fragmentu.

Nie dziwi, więc fakt, iż największa zmiana jakościowa dokonująca się współcześnie w rozumieniu ekonomii czy socjologii polega na odchodzeniu od postrzegania firm, społeczeństw czy rynków w kategoriach oddzielnych „trybików”. Na korzyść układów, o których właściwościach decyduje zarówno ich struktura jak i wzajemne oddziaływania między komponentami - a więc układów złożonych.

Przeanalizuję teraz własności, które pozwalają zaliczyć układ sprzedawców kupujących, jako układ złożony.

Cechy te są następujące:

- a) Układ jest wielokomponentowy - ilość sklepów oferujących dany towar mieści się w szerokim przedziale od nastu do kilkudziesięciu, a ilość kupujących od dziesiątek do tysięcy.
- b) Składniki układu mają wysoki stopień autonomii - każdy człowiek może sam decydować o zakupie, podobnie sprzedawcy podejmują samodzielnie decyzję o cenie sprzedaży.
- c) W układzie istnieją sprzężenia zarówno dodatnie (spadek ceny może powodować wzrost sprzedaży podobnie jak pozytywne opinie kupujących) jak i ujemne (wzrost ceny, negatywne opinie).
- d) Układ jest nieliniowy i jego ewolucja jest zależna od wcześniejszych stanów przyjmowanych przez składniki układu. Cecha ta wynika z wzajemnych, dynamicznych oddziaływań kupujących. Znajomi polecają sobie produkty,

wymieniają się opiniami, informacją, o cenie i miejscu zakupu, jednak nie spotykają się regularnie, a informacja, którą posiadają może nie być już aktualna. Nieliniowość wynika również, z pośredniego przekazu informacji od kupującego do sprzedawcy. Pieniądze działają tu podobnie jak ślad feromonowy w przypadku mrówek. Wielkość strumienia pieniędzy płynąca od nabywców do dostawców, nie zmienia się natychmiastowo i podlega sporym fluktuacjom. A zatem sprzedawca, ustalając politykę cenową swojego sklepu i kierując się dążeniem do maksymalizacji zysku, otrzymuje informację o reakcji klientów dopiero po pewnym czasie. Im dłuższy czas, tym większa precyzja określenia tendencji zmian w zysku, ale tym mniejsza aktualność tej informacji. Co więcej, to, jaki wpływ na klientów będzie miała określona zmiana ceny nie zależy tylko od szybkości propagacji informacji wśród kupujących, ale zależy również od tego, co zrobili inni sprzedawcy w tym samym czasie.

- e) Układ przejawia cechę samoorganizacji, klienci przepływają od droższych i gorszych sprzedawców do tańszych i lepszych, a sprzedawcy starają się tak prowadzić politykę cenową by zmaksymalizować osiąganą zysk.
- f) W układzie wstępuje zjawisko synergii. Przykładowo osoba mająca więcej znajomych ma dostęp do większej liczby informacji i potencjalnie znacznie większe możliwości jej weryfikacji.

2.4. Modelowanie układów złożonych

Nieliniowość układów złożonych powoduje, że modelowanie ich przy pomocy modeli z tak zwanym polem średnim może prowadzić do poważnych błędów. Przyczyną tego jest fakt, że małe zaburzenia oddziaływań pomiędzy dwoma składnikami w takim podejściu najprawdopodobniej zostaną wygaszone, podczas gdy w rzeczywistości, o ile działa dodatnie sprzężenie zwrotne, mogą ulec one rozpropagowaniu na cały układ oraz wzmocnieniu. Szczególnie mocno efekt ten może objawiać się w przypadku cen. Obniżanie ceny przez jednego sprzedawcę (producenta) może w skrajnym przypadku prowadzić do tak zwanej “wojny cenowej” (Krämer, Jung, and Burgartz 2016).

Do modelowania układów złożonych szczególnie dobrze nadają się systemy wieloagentowe. Oddziaływania pomiędzy elementami układu nie są tu uśredniane jak

w przypadku modeli średniego pola. Zróżnicowanie agentów wprowadza się szalenie prosto. Agenci mogą posiadać indywidualną pamięć opartą na ich wcześniejszych doświadczeniach, a ich reakcje mogą mieć charakter kontekstowy. Nie potrzebne są również żadne założenia o równowagowym charakterze procesów zachodzących w układzie (Parker et al. 2003). Wszystkie te własności systemów wieloagentowych czynią je szczególnie atrakcyjnymi do symulacji układów złożonych takich jak układy socjo-ekonomiczne, układy biologiczne, ekologiczne (Bousquet and Le Page 2004) i wiele innych.

Warto przywołać tu jeszcze inne metody matematyczne, stosowane do badania czy modelowania układów złożonych lub wybranych procesów w nich zachodzących. Należą do nich: modelowanie przy pomocy układów równań różniczkowych (Jirsa and Ding 2004), równań różniczkowo-całkowych (Arlotti et al. 2012) czy równania master (Breuer and Petruccione 1995).

Jednakże, pomimo ciągłego rozwoju aparatu matematycznego koniecznego do modelowania układów złożonych, podejścia analityczne napotykają wciąż liczne problemy. Praktyczny opis metodologii tworzenia równań kinetycznych opisujących układy dynamiczne jest przedstawiony w podręcznikowej pracy Roberts'a „Model emergent dynamics in complex systems” (Roberts 2014). Natomiast stosunkowo obszerny przegląd literatury dotyczący modelowania i analizy układów złożonych czytelnik znajdzie w pracy Marka Newmana „Complex Systems: A Survey” (M. E. J. Newman 2011). Dokładny opis poszczególnych typów modelowania wykracza poza ramy niniejszej pracy.

Współcześnie najbardziej dynamicznie rozwijają się metody badania układów złożonych oparte na reprezentacji ich przy użyciu sieci złożonych.

2.5. Sieci złożone

Podobnie jak w przypadku układów złożonych brak jest jednoznacznej definicji takiego układu, tak i w przypadku sieci złożonych wciąż brakuje ścisłych kryteriów definiujących te obiekty. Taki stan rzeczy musi tworzyć dyskomfort dla ścisłego umysłu. Z drugiej strony ciężko się dziwić brakowi precyzyjnych sformułowań w stosunku do obiektów będących modelami (reprezentacjami) układów złożonych, które same wymykają się jak dotychczas naszym możliwościom poznawczym. Dlatego też, badacze uciekają się do wyliczania właściwości charakteryzujące sieci

złożone. Można przypuszczać, iż wraz z postępem badań nad układami złożonymi, będą powstawać podejścia pozwalające coraz jednoznaczniej definiować i klasyfikować obiekty matematyczne, które obecnie określa się wspólnym mianem sieci złożonych. Jednakże na obecnym etapie badań, wygodniej jest używać intuicyjnych sformułowań zwiększających czytelność rozumowania, niż silić się na ścisłość kosztem zrozumiałości. Dlatego też, autor niniejszej pracy ograniczy się do podania przykładów sieci złożonych i wyliczenia w kolejnym rozdziale pewnej ilości wielkości charakteryzujących te obiekty, bez podejmowania prób ścisłego ich definiowania.

Jak wcześniej wspomniano niektóre układy złożone **mają strukturę sieci złożonej**, jak choćby: Internet, czy sieć neuronów w mózgu. Inne natomiast, takie jak społeczeństwo czy ekosystem, **mogą być reprezentowane przy pomocy sieci złożonych**, statycznych bądź dynamicznych, jedno- lub wielowarstwowych.

Dynamiczne sieci złożone pojawiają się, jako obraz oddziaływań układu złożonego w wybranym momencie t_i lub przedziale czasu $[t_i, t_j]$ gdzie $t_i < t_j$. Oddziałujące obiekty reprezentowane są, jako wierzchołki sieci (przykładowo białka, komórki czy ludzie). Oddziaływanie jest reprezentowane, jako krawędź i może przedstawiać tak różne relacje jak wchodzenie reagentów w reakcję, przepływ prądu między neuronami czy lubienie się ludzi. Wybrany sposób reprezentacji uwzględnia różne aspekty, zarówno samych obiektów jak i relacji ich łączących.

Przed badaczami stoi zadanie o dwojakim charakterze. Po pierwsze zrozumienie, które z obserwowanych fenomenów są wynikiem przyjętej reprezentacji, a które od reprezentacji nie zależą. Po drugie znalezienie takiej reprezentacji, w której opis badanego zjawiska jest najprostszy.

Sieci złożone są jedną z możliwych reprezentacji układów złożonych uwzględniającą relacje pomiędzy komponentami. Zasadniczy podział, jaki można przeprowadzić oparty jest na pochodzeniu układu złożonego. Sieci reprezentujące układy stworzone przez człowieka nazywane są sztucznymi (ang. artificial network (Goh et al. 2002) lub man-made network (Reis et al. 2014)). Przykładem sieci stworzonych przez człowieka jest Internet, sieć energetyczna czy sieć połączeń lotniczych. Natomiast przykładami sieci powstałych naturalnie mogą być, sieć połączeń między neuronami w mózgu czy sieć pokarmowa. W pracy, oba typy sieci (zarówno sieci tworzone przez człowieka jak i naturalne) będę nazywał sieciami

rzeczywistymi, aby odróżnić je od modeli matematycznych takich jak na przykład sieci Erdős-Renyi-ego (ER) czy Barabási'ego-Albert'a (BA).

Wśród sieci rzeczywistych wyróżnia się sieci jednowarstwowe, takie jak sieć pokarmowa oraz sieci wielowarstwowe (ang. multilayer network) (S. Boccaletti et al. 2014) jak przykładowo sieć połączeń komunikacyjnych między ludźmi. Sieć ta zawiera takie warstwy jak sieć bezpośrednich rozmów człowiek-człowiek, połączenia telefoniczne, połączenia przez komunikatory, fora internetowe, media społecznościowe typu Facebook i wiele innych. Każda warstwa ma właściwą sobie strukturę i specyficzne dla siebie własności. Przykładowo, sieć spotkań człowiek-człowiek charakteryzuje możliwość natychmiastowych przekazów niewerbalnych (przekaz emocjonalny) przez „mowę ciała” i mimikę. W trakcie spotkania osoby mogą się zarażać (na przykład gripą). Przepływ informacji w trakcie spotkania może być dwukierunkowy. Jest też natychmiastowy - informacja jest przekazywana w momencie mówienia. Ma charakter czasowy – nie można w dowolnym późniejszym momencie powrócić do rozmowy, która się odbyła (chyba, że ją nagraliśmy). Odmienne własności ma na przykład sieć połączeń e-mail. Przekaz informacji jest jednokierunkowy (od nadawcy do odbiorcy). Nieznany jest czas, kiedy zostanie odczytana wiadomość. Dostęp do wiadomości jest stały (w dowolnym momencie można powrócić do jej treści). Przekazywana informacja nie zawiera treści niewerbalnych tożsamyh z informacją niesioną przez mimikę czy „mowę ciała”. Można tą drogą wprawdzie, przekazywać informacje o swoich emocjach (poprzez użycie emotikonów lub opis odczuwanych uczuć), ale dotyczy ona stanu osoby piszącej w momencie pisania (ewentualnie emocji związanych z przekazywaną treścią). Nie można się zarażać tą drogą, aczkolwiek mogą się w ten sposób rozprzestrzeniać wirusy komputerowe.

Liczebność kontaktów, ich częstość, czas trwania mogą różnić się bardzo znacznie dla poszczególnych warstw sieci wielowarstwowej. Na przykład osoba mająca bardzo niewielu znajomyh rzeczywistych i spotykająca się z nimi rzadko, może mieć bardzo rozbudowaną sieć kontaktów społecznych na Facebooku a częstość jej kontaktów wirtualnych może być bardzo duża. W dalszej części pracy omawiane będą tylko sieci jednowarstwowe gdyż zarówno dostępne dane rzeczywiste jak i względnie dobrze ugruntowane metody badawcze dotyczą właśnie takich sieci. Osoby

zainteresowane sieciami wielowarstwowymi odsyłam do przeglądowej pracy: „The structure and dynamics of multilayer networks” (S. Boccaletti et al. 2014).

W celu zrozumienia procesów zachodzących w układzie złożonym odpowiadającym za proces formowania cen w sytuacji wolnorynkowej stworzono model wieloagentowy zwany dalej modelem Piskor-Ignatowicza-Zachary. Konstrukcję modelu opisano w rozdziale 4. Do badania stworzonego modelu układu złożonego kupujący-sprzedawcy wykorzystano reprezentację przy pomocy sieci złożonej jednowarstwowej. Charakterystykę sieci złożonych i ich własności omówiono w rozdziale 3.

3. Reprezentacja układów złożonych za pomocą sieci złożonych

Rozdział ten poświęcono omówieniu pojęcia „sieci złożonej”, jako matematycznego reprezentanta rzeczywistych układów złożonych. Stanowi on krótkie wprowadzenie niezbędnych pojęć pozwalających charakteryzować sieci złożone i używanych w dalszej części pracy. Scharakteryzowano też sieci losowe Renyi’ego Erdos’a wskazując na odmiennność ich charakterystyk od rzeczywistych układów złożonych.

Pojęcie „sieci złożonej”, szeroko wykorzystywane w literaturze, nie jest całkowicie jednoznaczne (jak wspomniano wcześniej). Z jednej strony oznacza klasę układów złożonych o strukturze sieci (czyli złożonych **sieci rzeczywistych**), z drugiej zaś strony, stanowi sposób reprezentowania (modelowania) układów złożonych. Ponieważ własności reprezentowanych układów są odzwierciedlone w modelach, zatem bez nieporozumienia w niniejszej pracy (podobnie jak w literaturze), nazwa ta będzie używana często bez podkreślania dokładniejszego znaczenia.

Badania sieci złożonych mają charakter interdyscyplinarny i obejmują swoim zasięgiem tak różne dyscypliny nauki jak biologia, socjologia, fizyka, informatyka czy ekonomia. U podstaw tak rozległych zainteresowań badawczych tej nowej dziedziny naukowej leżą obserwacje, iż różne układy złożone przejawiają pewne cech wspólne, nie- zależne od tworzących je komponentów. Do badania tych układów wykorzystuje się metody mechaniki statystycznej, teorii grafów, statystyki oraz różne metody symulacyjne w tym systemy wieloagentowe.

3.1. Podstawowe pojęcia

Sieci złożone, jako reprezentacja układów złożonych, to grafy lub multigrafy, cechujące się nietrywialną budową topologiczną, mającą znaczący wpływ na dynamikę procesów zachodzących w tych sieciach (S. Boccaletti et al. 2014). Ta „nietrywialność” odzwierciedla między innymi fakt, iż połączenia między komponentami układu złożonego nie są ani w pełni regularne (jak w kryształach) ani w pełni przypadkowe jak w układach reprezentowanych grafami losowymi. Skutkuje to pojawianiem się wielkiego bogactwa struktur. Nasza zdolność do analizy i rozumienia funkcji sieci złożonych zależy między innymi od umiejętności wyodrębnienia i scharakteryzowania pojawiających się w nich konfiguracji. Na przestrzeni ostatnich nastu lat nastąpił burzliwy rozwój metod badawczych dedykowanych pomiarom i charakteryzacji tychże struktur, jak i metod analizy związków między topologią sieci złożonych a dynamiką procesów zachodzących w tych sieciach (Costa et al. 2007)

Historycznie początki badań nad sieciami, rozumianymi, jako paradygmat modelowania (reprezentowania), wiążą się z rozwojem teorii grafów stworzonej przez Eulera w 1736 (Euler 1736). Początkowo zainteresowania badaczy koncentrowały się na grafach regularnych i ich własnościach. Jednakże już w latach 30-tych ubiegłego wieku, Jacob Levy Moreno zastosował grafy do ilustracji relacji społecznych (Borgatta and Moreno 1954). Pod koniec lat pięćdziesiątych prace dwóch węgierskich matematyków, Paula Erdős’a i Alfréda Rényi’ego zapoczątkowały zainteresowania grafami przypadkowymi (Erdős and Rényi 1959). Dalszy rozwój badań nad sieciami złożonymi wiąże się z kumulacją obserwacji, iż rzeczywiste sieci przejawiają cechy nieopisywane przez klasyczne grafy przypadkowe. Współczesny bardzo dynamiczny rozwój prac poświęconych sieciom złożonym rozpoczął się od dwóch prac: Watts’a i Strogatz’a poświęconej zagadnieniu „małości świata” (Watts and Strogatz 1998) oraz Barabási’ego i Albert’a wprowadzającej model sieci bezskalowej (Barabási and Albert 1999).

Poniżej przytoczone zostały niezbędne pojęcia (Stefano Boccaletti et al. 2006) istotne dla śledzenia rozważań przedstawionych w pracy.

Jak wspomniano, relacje pomiędzy agentami najwygodniej modelować jest za pomocą teorii grafów i charakteryzować za pomocą wprowadzonej tam terminologii. Należy jednak zwrócić uwagę, że w literaturze można spotkać różne definicje grafu. W rozważaniach tej pracy będzie wykorzystywana najprostsza postać grafu, zatem przyjmujemy poniższą definicję.

Graf $G = (A, R)$ gdzie A jest zbiorem nazywanym zbiorem wierzchołków grafu, zaś $R \subseteq A \times A$ jest relacją definiującą połączenia wierzchołków.

Dla celów modelowania sieci społecznościowej w tej pracy przyjmujemy, klasę grafów niezorientowanych bez pętli.

Podgrafem $G' = (A', R')$ grafu $G = (A, R)$ jest taki graf, że $A' \subseteq A$ i $R' \subseteq R$. Podgrafem indukowanym wierzchołkowo nazywamy taki podgraf G' , który zawiera wszystkie krawędzie grafu G łączące wierzchołki zbioru A' i oznaczamy go $G' = G[A']$. W szczególności, podgraf wszystkich bezpośrednich sąsiadów (sąsiadów pierwszego rzędu) wierzchołka a_j jest definiowany, jako podgraf indukowany przez zbiór A' stanowiący zbiór wierzchołków połączonych krawędzią z wierzchołkiem a_j .

Ilość wierzchołków grafu będzie oznaczana jako N , zaś ilość krawędzi jako E .

Stopień wężła a_i oznaczany jako k_i (jak wiadomo, jest to ilość krawędzi incydentnych do tego wężła).

Średni stopień wierzchołków grafu (ang. average degree) $\langle k \rangle$ jest średnią z k_i po wszystkich N wężłach sieci:

$$\langle k \rangle = \frac{1}{N} \sum_i k_i \quad (3-1)$$

W literaturze dotyczącej układów złożonych często spotykane jest nazywanie agentami wierzchołków grafu (czy multigrafu) reprezentującego elementy układu złożonego (M. E. J. Newman 2011), natomiast krawędzie reprezentują albo fizyczne połączenia jak w przypadku Internetu czy systemu nerwowego, albo istnienie określonej relacji (więzi), oddziaływania czy przepływu informacji między komponentami. Dla części układów złożonych, takich jak Internet, system nerwowy czy drogi, sieci złożone stanowią naturalny sposób przedstawienia pewnych cech budowy tych układów. Dla innych, jak na przykład społeczeństw, języków, interakcji

ekonomicznych czy rozmów telefonicznych, sieci złożone stanowią jedną z metod reprezentowania relacji między poszczególnymi składnikami układu złożonego.

W literaturze wprowadzono szereg pojęć pozwalających charakteryzować sieci złożone, jednakże omówione tu zostaną, jedynie te z nich, które wykorzystywane są w niniejszej pracy. Należą do nich:

- odległości w sieciach złożonych,
- rozkład stopni węzłów,
- współczynnik gronowania.

Wymienione powyżej własności zostaną bardziej szczegółowo omówione w dalszej części tego rozdziału.

3.2. Odległości w sieciach złożonych i własność „małych światów”

W teorii grafów odległość między wierzchołkiem i , a wierzchołkiem j , jest definiowana, jako najkrótsza droga łącząca te wierzchołki (dla grafów nieskierowanych -minimalna ilość krawędzi łączącej te wierzchołki). Również dla klasycznych sieci fizycznych takich jak d – wymiarowa regularna sieć krystaliczna, obserwuje się następującą własność: odległość między losowo wybranymi węzłami sieci o rozmiarze N węzłów jest proporcjonalna do $N^{\frac{1}{d}}$ (Albert and Barabasi 2002).

Gdyby podobne zależności obowiązywały w sieciach społecznych, to tak zwana “poczta pantoflowa” działałaby szalenie wolno i nieefektywnie. Tymczasem wiadomo z obserwacji, iż sprawność przekazu informacji przez sieć połączeń między znajomymi jest duża i efektywna. Używa się nawet określenia, że “świat to jedna wioska”. Ustalenie, kiedy ukuto to określenie w języku polskim wydaje się niemożliwe, jednakże oddaje ono dobrze ogólną intuicję dotyczącą sieci społecznych. Wiadomo natomiast, iż anglojęzyczny termin „global village” został wprowadzony w 1962 przez Herberta Marshalla McLuhana w książce „The Gutenberg Galaxy” (McLuhan 1962). Eksperymentalne dowody poprawności tej intuicji zaczęły pojawiać się od późnych lat sześćdziesiątych wraz z pionierskim eksperymentem Milgram’a opisanym w pracy pod tytułem „The small-world problem” (Milgram 1967). Doświadczenie polegało na przekazaniu paczki dokumentów między dwoma losowo wybranymi osobami (za pomocą innych nieznanych im osób) na terenie USA. W

wyniku tego eksperymentu Milgram ustalił, że potrzeba średnio 5 przeskoków, aby przekazać wiadomość między dwoma losowo wybranymi osobami dla populacji 200 milionów osób.

Teoretyczne konstrukcje sieci posiadających własność „małości świata” zaczęto tworzyć po przełomowej pracy Watts’a i Strogatz’a pod tytułem: „Collective dynamics of ‘small-world’ networks” (Watts and Strogatz 1998). Wówczas okazało się, że zaobserwowana przez Milgrama własność „małości świata” nie jest czymś wyjątkowym tylko dla sieci społecznych, lecz występuje powszechnie w wypadku stanowiącej większość rzeczywistych sieci złożonych. Odległość między dwoma losowo wybranymi węzłami sieci posiadającej własność „małości świata” jest proporcjonalna do logarytmu z ilości węzłów w sieci.

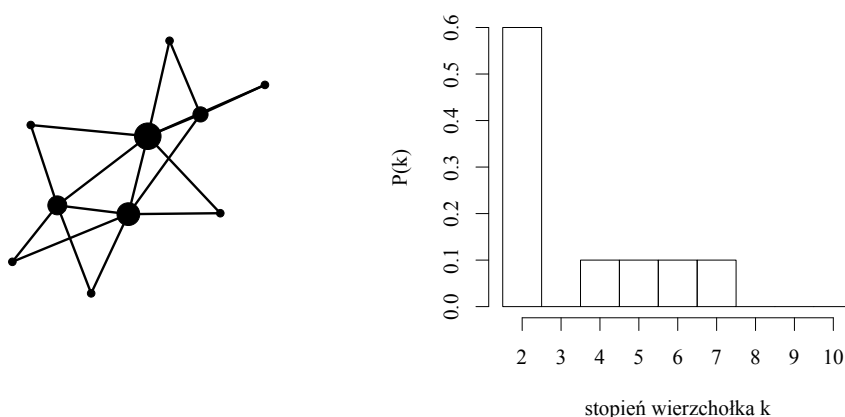
Istotną cechą, wyjaśniającą własność ‘małości świata’: jest fakt występowania „daleko zasięgowych” połączeń w rzeczywistych sieciach złożonych.

3.3. Rozkład stopni węzłów

Dziewiętnastowieczny mechanistyczny obraz świata wraz z jego „geometrycznymi” wyobrażeniami tkwi bardzo głęboko w naszej świadomości. Takie pojęcia jak średnia czy rozrzut bardzo dobrze wpisują się w nasze własne doświadczenia życiowe dotyczące wymiarów otaczających nas przedmiotów czy nas samych. Przykładowo, ilość sąsiadów w idealnej sieci krystalicznej jest stała, a w rzeczywistej mieści się w pewnym dobrze zdefiniowanym przedziale. Podobnie szerokość miejsc parkingowych oscyluje wokół pewnej średniej wartości. Wynika to z faktu, iż nie produkuje się samochodów osobowych o szerokości 10 czy 100 metrów. Tymczasem, w przypadkach socjologicznych i ekonomicznych, gdy popatrzymy na wielkość miast czy wysokość zarobków, obserwowane rozrzuty przekraczają wiele rzędów wielkości. Podobnie rzecz się ma z ilością znajomych. Mimo iż, stanowiąca większość osób posiada kilku znajomych, to nieliczne osoby posiadają ich setki lub nawet tysiące. Tę własność zaobserwował włoski inżynier, socjolog i ekonomista Vilfredo Pareto ponad sto lat temu. Jednakże świadomość konsekwencji występowania takich rozkładów zaczęła pojawiać się dopiero współcześnie wraz z narastającą ilością niedających się opisać przypadków, takich jak nagłe przeciążenia serwerów, czy choroby propagujące się, mimo iż powinny wygasnąć w myśl przewidywań modeli matematycznych opartych na poissonowskim rozkładzie stopni

węzłów. Przyczyną tego stanu rzeczy jest występowanie ciężkoogonowych rozkładów (ang. heavy-tailed distributions) stopni węzłów w sieciach rzeczywistych (Marković and Gros 2014).

Jedną z podstawowych charakterystyk grafów stanowi rozkład prawdopodobieństwa $P(k)$ wystąpienia określonych stopni k wierzchołków w grafie. Poniżej na rysunku Rys. 3-1 przedstawiono graf i odpowiadający mu rozkład występowania określonego stopnia wierzchołka.



Rys. 3-1 Po lewej stronie przedstawiono graf, a po prawej odpowiadający mu rozkład $P(k)$ wystąpienia wierzchołków o stopniu k

Dla rzeczywistych sieci złożonych najczęściej obserwowanym typem rozkładu prawdopodobieństwa $P(k)$ wystąpienia wierzchołka stopnia k jest rozkład ciężkoogonowy (ang. heavy-tailed distributions).

Mówimy, że zmienna losowa X ma rozkład prawdopodobieństwa $P(X)$ ciężkoogonowy, jeśli spełnia następującą zależność:

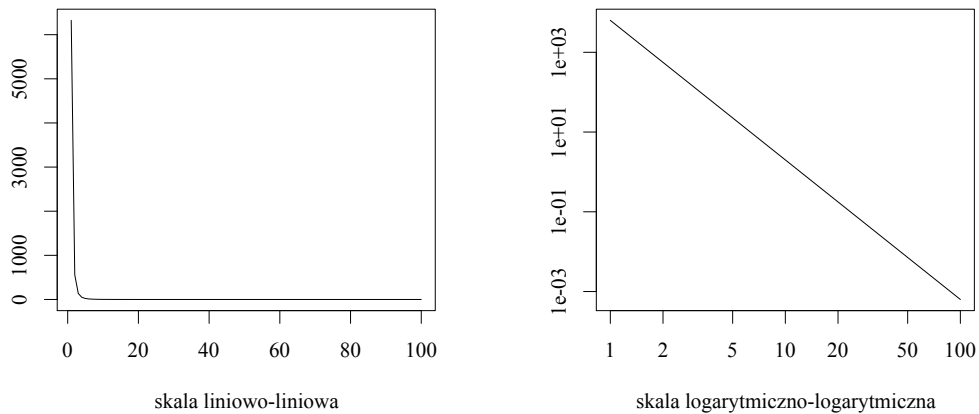
$$\lim_{x \rightarrow \infty} \exp(\lambda x) P(X > x) = \infty \text{ dla każdego parametru } \lambda > 0$$

Najczęściej rozważanymi przedstawicielami tej rodziny jest rozkład potęgowy (Muchnik et al. 2013) i rozkład logarytmiczno-normalny (N. Z. Gong et al. 2012).

Rozkład potęgowy gęstości prawdopodobieństwa $f(x)$, zmiennej losowej ciągłej X podany jest wzorem (3-2) i przedstawiony na rysunku: Rys. 3-2.

$$P(x < X < x + dx) = f(x) = \frac{\alpha - 1}{x_{min}} \left[\frac{x}{x_{min}} \right]^{-\alpha} \quad (3-2)$$

gdzie parametr skalowania $\alpha > 0$, x_{min} - parametr obciążenia odpowiadający za miejsce gdzie zaczyna się rozkład.



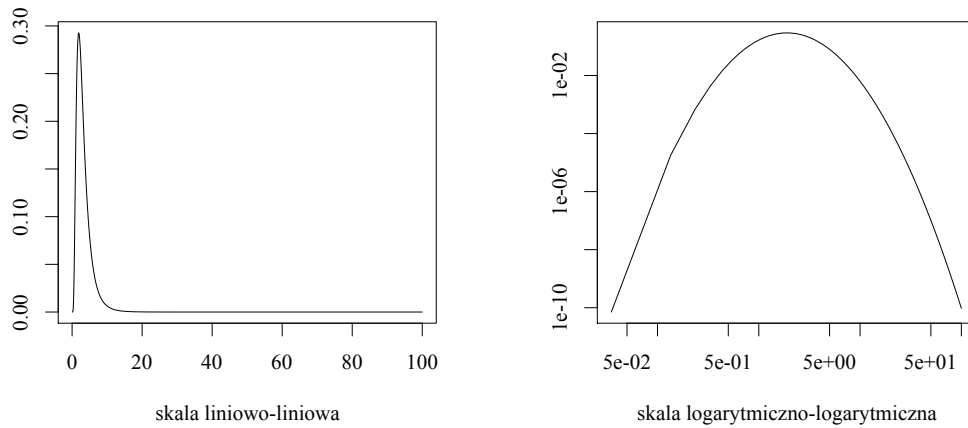
Rys. 3-2 Rozkład potęgowy ze wzoru (3-2) dla $x_{min} = 1$, $\alpha = -3.5$ przedstawiony na skali liniowej po lewej stronie oraz na skali logarytmiczno-logarytmicznej po prawej stronie ilustracji.

Rozkład potęgowy na wykresie logarytmiczno-logarytmicznym tworzy linię prostą.

Rozkład gęstości prawdopodobieństwa logarytmiczno-normalny (N. Z. Gong et al. 2012) podany jest wzorem (3-3) i przedstawiony na rysunku: Rys. 3-3.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \quad (3-3)$$

Gdzie μ – wartość oczekiwana, σ – odchylenie standardowe.



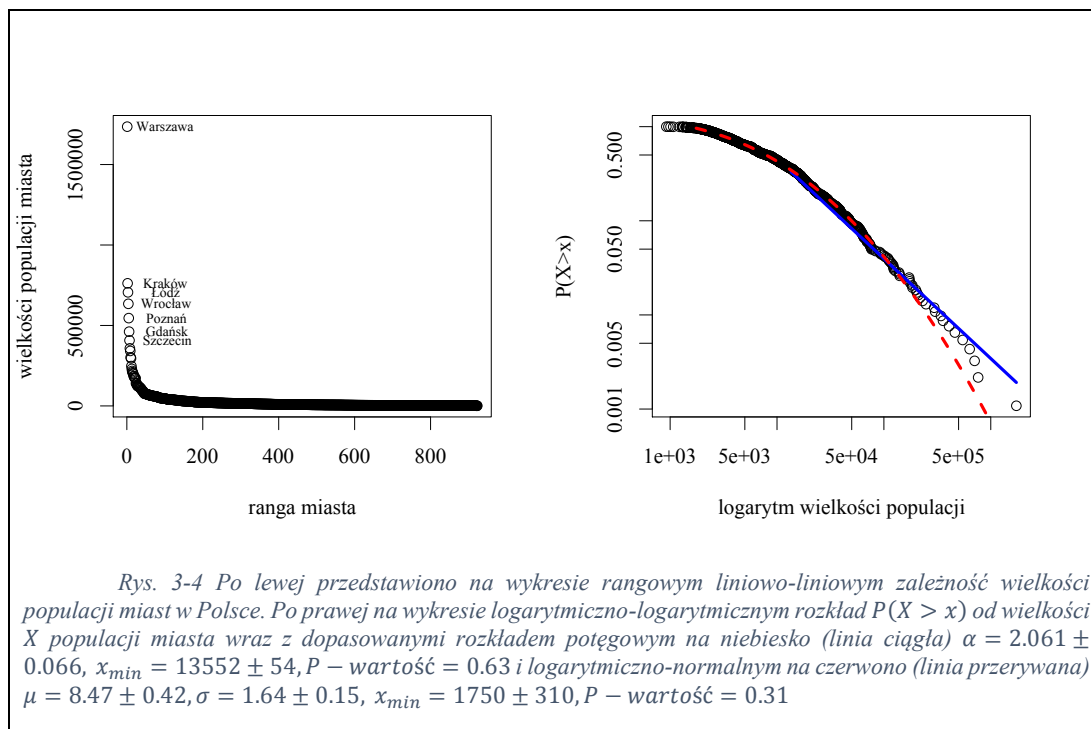
Rys. 3-3 Rozkład logarytmiczno-normalny ze wzoru (3-3) na wykresie liniowo-liniowym po lewej stronie i logarytmiczno-logarytmicznym po prawej stronie, dla wartości parametrów: $\mu = 1$, $\sigma = 0.6$

Poniżej na ilustracji Rys. 3-4 po lewej stronie przedstawiono w porządku nierosnącym (rangi malejące) populację miast Polski (ang. decreasing order of rank).

(Rangowanie - operacja stosowana w statystyce polegająca na wprowadzeniu porządku niemalejącego (rzadziej nierosnącego) w zbiorze obserwacji doświadczalnych według wartości (niekoniecznie liczbowych) jednej ze zmiennych)

Na Rys. 3-4 po prawej stronie przedstawiono $P(X > x)$ w funkcji wielkości populacji X danego miasta. Każde otwarte kółko na wykresie odpowiada jednemu polskiemu miastu, linia ciągła odpowiada dopasowanemu do danych rozkładowi

potęgowemu, a linia przerywana dopasowanemu do danych rozkładowi logarytmiczno-normalnemu (opracowanie własne).



Jak widać na zamieszczonym powyżej rysunku, różnica między oboma rozkładami na sporym kawałku wykresu jest minimalna. Dla danych z zakresu $x \ll \sigma$ w praktyce nie istnieje możliwość rozróżnienia pomiędzy tymi dwoma rozkładami. Powody takiego stanu rzeczy są dwojakie: trudność rozróżnienia obu rozkładów (Clauset, Shalizi, and Newman 2009) jak i bliski związek pomiędzy nimi - pięknie pokazany w historycznej perspektywie w pracy (Mitzenmacher 2004).

3.4. Współczynnik gronowania

Współczynnik klasteryzacji (gronowania) został pierwotnie wprowadzony w socjologii, jako parametr charakteryzujący tworzenie się grup. Z czasem zaczęto go stosować, jako jedną z ważniejszych wielkości opisujących budowę sieci (M.E.J. Newman 2003), (Albert and Barabasi 2002).

Współczynnik gronowania C (ang. clustering coefficient lub transitivity) jest miarą zdolności do tworzenia trójkątów w danej sieci.

W literaturze można znaleźć różne wzory definiujące ten współczynnik. Zgodnie z (M.E.J. Newman 2003) można go zapisać w następującej postaci:

$$C = \frac{3 * \text{ilość trójkątów w sieci}}{\text{ilość trójek połączonych ze sobą wierzchołków}} \quad (3-4)$$

Empirycznie zaobserwowano, że współczynnik gronowania najlepiej pozwala rozróżniać między sobą poszczególne typy sieci. Przyjmuje on wartości zauważalnie wyższe w naturalnych sieciach złożonych (M.E.J. Newman 2003) niż ma to miejsce w sieciach losowych Reniy’go Erdos’a (Erdős and Rényi 1959) i znacząco wyższe dla sieci społecznych (M. E. J. Newman and Park 2003). Możliwe uogólnienia współczynnika gronowania na złożone sieci ważone zaprezentowano w pracy (Saramäki et al. 2007).

3.5. Sieci losowe Reniy’go Erdos’a

Sieci losowe Reniy’go Erdos’a zwane też klasycznymi grafami losowymi Erdos’a Reniy’go (ER) (Erdős and Rényi 1959) należą do najlepiej przebadanych grafów losowych. (M. E.J. Newman 2004). Jako pierwsze znalazły zastosowanie do modelowania sieci połączeń społecznych w procesach rozprzestrzeniania chorób (House 2012), czy modelowaniu sieci komputerowych (Alderson et al. 2005).

Istnieją dwa blisko ze sobą powiązane modele sieci losowych Erdos’a Reniy’go. Pierwszy oparty jest na wylosowaniu z rozkładem jednorodnym, ze zbioru wszystkich sieci o N wierzchołkach i E krawędziach jednej sieci. Drugie bardziej ogólne podejście oparte jest na połączeniu N wierzchołków E krawędziami z zadaniem prawdopodobieństwem, tak, że dana para wierzchołków może zostać wybrana tylko jeden raz. Współcześnie właśnie to podejście jest stosowane do tworzenia sieci losowych ER.

Sieci losowe Erdosa Reniy’go posiadają własność „małych światów” to jest: średnia odległość $\langle l \rangle$ między dowolnymi dwoma węzłami tej sieci jest proporcjonalna do $\ln(N)$ (Mark E. J. Newman 2002):

$$\langle l \rangle \sim \frac{\ln(N)}{\ln(E)} \sim \ln(N) \quad (3-5)$$

Niestety jest to jedyna własność obok występowania klastra perkolacyjnego (ang. giant component) upodabniająca je do sieci rzeczywistych. Niemniej sieci losowe Erdos’a Reniy’go są stosowane do dziś, jako model referencyjny pozwalający

odróżnić zjawiska powodowane przez procesy losowe od tych, za których powstanie odpowiadają jakieś inne mechanizmy.

Wielkością odróżniającą grafy losowe Erdos'a Reniy'go od sieci rzeczywistych jest na przykład rozkład stopni wierzchołków w tym grafie opisywany rozkładem dwumianowym (ang. binomial distribution) (Mark E. J. Newman 2002).

Prawdopodobieństwo $P(k_i = k)$, że dany węzeł sieci będzie miał dokładnie stopień k , (często w literaturze oznaczane p_k), wynosi:

$$P(k_i = k) = p_k = \binom{N-1}{k} p^k (1-p)^{N-k-1} \quad (3-6)$$

$$\text{gdzie } p = \frac{\langle k \rangle}{N}$$

Gdy liczba krawędzi w grafie jest mała: to jest, gdy $p = \frac{\langle k \rangle}{N} \ll 1$ to rozkład dwumianowy można przybliżyć rozkładem Poissona (Albert and Barabasi 2002):

$$p_k = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!} \quad (3-7)$$

Dlatego też klasyczne grafy przypadkowe Erdos'a Reniy'go bywają nazywane „grafami poissonowskimi” (ang. Poisson random graph). Rozrzut stopni wierzchołków $\sigma(k)$ jest w związku z tym bardzo mały i wynosi:

$$\sigma(k) = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \sqrt{\langle k \rangle} \quad (3-8)$$

Własność ta stoi w sprzeczności z obserwowanym powszechnie w sieciach rzeczywistych bardzo wysokim usieciowieniem niektórych węzłów (ang. hub lub network hub). Rozkład stopni wierzchołków w sieciach rzeczywistych opisują rozkłady o ciężkich ogonach (ang. heavy tailed distribution).

Kolejną własnością odróżniającą grafy Erdos'a Reniy'go od sieci rzeczywistych jest średni współczynnik gronowania $\langle C_{rand} \rangle$. Dla grafu przypadkowego zachodzi (Albert and Barabasi 2002):

$$\langle C_{rand} \rangle = \frac{\langle k \rangle}{N} \quad (3-9)$$

Jak wynika ze wzoru (3-) - dla grafu przypadkowego średni współczynnik gronowania wraz ze wzrostem rozmiaru sieci będzie dążył do zera jak N^{-1} . W

przypadku sieci rzeczywistych współczynnik klasteryzacji przyjmuje wartości znacząco większe od zera (Albert and Barabasi 2002).

3.6. Podsumowanie

Podsumowując, rozdział ten został poświęcony reprezentacji układów złożonych za pomocą sieci złożonych. Przybliżono wielkości, za pomocą, których charakteryzuje się sieci złożone.

Wskazano na podstawie literatury, że rozkład stopni wierzchołków w sieciach złożonych ma charakter ciężkoogonowy. Podkreślono też, że sieci złożone mają istotnie odmienne charakterystyki od sieci losowych.

Zwrócono też uwagę, że współczynnik gronowania ma wartości zauważalnie wyższe w naturalnych sieciach złożonych niż ma to miejsce w sieciach losowych (RE) i znacząco wyższe dla sieci społecznych.

Uwagi te są bardzo istotne dla dalszych rezultatów pracy.

4. Modelowanie sieci społecznych

Rozdział dotyczy modelowania sieci społecznych. Przedstawia matematyczny, dynamiczny model wieloagentowy, umożliwiający tworzenie sztucznego społeczeństwa (ang. artificial society), o właściwościach podobnych do tych, obserwowanych w świecie rzeczywistym. W szczególności, w modelu uwzględniono fakt, że każdy agent posiada indywidualny zestaw cech, co pociąga za sobą odmienne zachowania różnych agentów wobec tych samych warunków otoczenia.

Zaproponowany model jest modelem układu złożonego. Analiza jego właściwości prowadzona jest przy użyciu reprezentacji jednowarstwowej sieci złożonej. Przyjęty sposób reprezentowania układu złożonego jest na chwilę obecną najpowszechniej stosowany wśród badaczy, co daje największe możliwości porównania z wynikami innych badań i analiz.

W rozdziale 5 zostanie pokazane w oparciu o przedstawiony formalny model, że reprezentująca sztuczną społeczność dynamiczna sieć relacji, charakteryzuje się: ciężkoogonowym (ang. heavy tailed) rozkładem liczby nawiązywanych i utrzymywanych przez agentów kontaktów (relacji) reprezentowanym przez rozkład stopni wierzchołków grafu oraz wysoką wartością współczynnika gronowania (klasteryzacji) sieci.

W niniejszym rozdziale zawarto:

- Przegląd literaturowy wraz z przedstawieniem przemyśleń autora dotyczących trudności występujących zarówno przy ankietowym sposobie zbierania informacji o rzeczywistych sieciach społecznych jak i ograniczeniach w przeprowadzaniu eksperymentów społecznych.
- Omówiono użyteczność zastosowania iteracyjnego podejścia do tworzenia wieloagentowych modeli symulacyjnych procesów socjo-ekonomicznych.
- Zaprezentowano opis wieloagentowego, dynamicznego, parametrycznego modelu tworzenia sztucznego społeczeństwa nazwanego modelem Piskor-Ignatowicza–Zachary (Zachara and Piskor-Ignatowicz 2016).

Zaproponowany model stanowi podstawę programu symulującego sztuczne społeczeństwo i zostanie wykorzystany w modelowaniu procesu kształtowania cen przedstawionym w następnym rozdziale.

Dla uniknięcia nieporozumień natury terminologicznej należy zauważyć, że słowo „relacja” ma wiele znaczeń. W kontekście tej pracy używane są w dwóch (a nawet trzech) znaczeniach: społeczno-psychologicznym, matematycznym i potocznym.

Spółeczno-psychologiczny termin, określa relacje zachodzące pomiędzy ludźmi. Relacje te mogą być różnego typu, ale w modelu rozpatrywany jest tylko **jeden typ relacji**, mianowicie relacja znajomości (uwzględniany jest natomiast dynamicznie zmieniający się stopień siły tej relacji, począwszy od „dalekich znajomych” aż do przyjaźni).

Termin ten używany jest też w znaczeniu potocznym, to jest, jako „posiadanie relacji z inną osobą”, czyli utrzymywanie kontaktów. W tym sensie każdy z agentów może „mieć wiele relacji”.

W znaczeniu matematycznym, relacja, jako podzbiór iloczynu kartezjańskiego, będzie modelowała występujące relacje pomiędzy agentami.

4.1. Przegląd literatury

W rozdziale trzecim przedstawiono charakterystyki złożonych sieci, które mogą reprezentować szerokie spektrum układów złożonych. W tym rozdziale ograniczymy się tylko do tych układów złożonych, którymi są sieci społecznościowe.

Stosunki społeczne oraz wynikające z nich sieci społecznościowe stanowią kluczowy element naszego społeczeństwa. Od dwudziestu lat badacze zajmują się naturalnymi sieciami społecznymi. Jednak, tak jak w przypadku wielu innych zjawisk występujących w naturze, trudno jest stwierdzić jak taka sieć mogłaby wyglądać w nieco zmienionych okolicznościach. W odpowiedzi na to pytanie pomóc mogą modele i symulacje. Dysponując właściwym modelem można stworzyć szeroką gamę sieci i badać ich właściwości. Sieci tego rodzaju można wykorzystywać do testowania i weryfikacji rozmaitych hipotez. W niniejszym rozdziale przedstawiam model wieloagentowy umożliwiający tworzenie sieci o właściwościach podobnych do tych obserwowanych w świecie rzeczywistym, takich jak wysoka wartość

współczynnika gronowania (klasteryzacji) sieci czy ciężkoogonowy (ang. heavy tailed) rozkład relacji w ich obrębie.

Model przetestowano dla istotnego zakresu parametrów początkowych. W dalszej części pracy pokazano, w jaki sposób wspomniane parametry wpływają na właściwości generowanej sieci.

Analiza sieci społecznościowych niesie ze sobą liczne istotne zastosowania. Można z niej korzystać przy modelowaniu rozmaitych procesów, takich jak chociażby rozprzestrzenianie się chorób (Pastor-Satorras et al. 2015), wpływ działań marketingowych (Weng, Menczer, and Ahn 2013), identyfikowanie grup przestępczych (Dajda et al. 2014) czy rozpowszechnianie informacji i niepokoju (Malik and Mucha 2013). Niewątpliwie pożądana jest tu możliwość swobodnego zmieniania parametrów sieci i przewidywania ich wpływu na badany proces.

Właśnie w tym zakresie modelowanie i symulacje przynoszą najwięcej korzyści.

Ponieważ opis analityczny sieci społecznościowych jest bardzo trudny, zamiast niego, do charakterystyki sieci, używa się parametrów statystycznych. Podejście to wywodzi się z dziewiętnastowiecznej termodynamiki, a współcześnie jest powszechnie stosowane w wielu gałęziach nauki jak choćby fizyka statystyczna czy fizyka układów złożonych.

Jedną z podstawowych cech sieci społecznościowej jest ciężkoogonowy rozkład liczby relacji między uczestnikami sieci (Vázquez et al. 2006). O tego rodzaju rozkładzie pisano w kontekście relacji osobistych (Ball and Newman 2013), (Mislove et al. 2007), liczby połączeń telefonicznych (Blondel, Decuyper, and Krings 2015) i wielu innych.

Nie są jednak jeszcze znane szczególne właściwości takiego ciężkoogonowego rozkładu liczby relacji. Za najpowszechniejsze uważa się dwa rodzaje rozkładu: rozkład potęgowy (Muchnik et al. 2013) i rozkład logarytmiczno-normalny (N. Z. Gong et al. 2012). Trudności w ich rozróżnieniu (ze względu na ich podobieństwo w określonym przedziale wartości) nakreślono czytelnie w pracy (Clauset, Shalizi, and Newman 2009).

Inny aspekt bliskiego związku między oboma rozkładami zawarto w pracy Mitzenmachner'a (Mitzenmacher 2004), gdzie zwrócono uwagę na bardzo podobny

mechanizm tworzenia zarówno rozkładu logarytmiczno-normalnego jak i potęgowego.

Dla celów oceny wyników symulacji w niniejszej pracy, uzyskany rozkład liczby relacji agentów porównywany jest z rozkładem potęgowym. Rozkład ów występuje często w sieciach naturalnych, a dobrze znana publikacja Clauset'a (Clauset, Shalizi, and Newman 2009) podaje wartości referencyjne współczynnika skalowania α stanowiącego jeden z głównych parametrów go opisujących.

Kolejnym istotnym parametrem sieci społecznościowych jest średni współczynnik gronowania (klasteryzacji) $\langle C \rangle$. Wywodzi się on, z teorii grafów i wykorzystywany jest w socjologii do oceny procesów powstawania grup i więzi grupowych. Stopniowo stał się również jednym z głównych parametrów opisujących strukturę sieci społecznościowej (Albert and Barabasi 2002), (M.E.J. Newman 2003).

Albert i Barabasi podają listę współczynników klasteryzacji rozmaitych sieci (Albert and Barabasi 2002). Ich praca pokazuje, że w sieciach naturalnych i społecznościowych przekraczających 1000 wierzchołków faktyczna wartość współczynnika klasteryzacji wynosi, co najmniej 0.08, podczas, gdy wartość teoretyczna współczynnika klasteryzacji dla sieci losowej jest znacznie poniżej 0.01.

Tradycyjne badania socjologiczne bazowały na ankietach. Generowało to wysokie koszty i było podatne na najróżniejsze przekłamania, przykładowo, gdy osoby ankietowane odczuwały wstyd lub spodziewały się, że dana czynność spotka się ze złą oceną społeczną, odpowiadały nieszczerze. Mimo, że opracowano metody pozwalające ograniczyć wpływ tego rodzaju zjawisk poprzez formułowanie pytań tak, by zminimalizować chęć nieszczerych odpowiedzi oraz móc oszacować poziom, na jakim występują przekłamania to wciąż tam gdzie można zestawić „twarde dane” z wynikami ankiet otrzymuje się rozbieżności (Shenton 2004).

Koleją trudnością występującą przy badaniach ankietowych jest wrażliwość na problemy statystyczne związane z doбором reprezentatywnej próby. Dla sieci o potęgowym lub logarytmiczno-normalnym rozkładzie stopni wierzchołków dobór prób metodą Monte Carlo prowadzi do bardzo słabego reprezentowania w próbie wierzchołków o najwyższych stopniach, podczas gdy wkład od nich pochodzący może być najbardziej istotny z punktu widzenia badanej właściwości sieci. Jako przykład można wziąć zdolność przesyłania informacji w sieci Internet. Uszkodzenie nawet 99% losowo wybranych węzłów nie pozbawi zdolność tej sieci do przekazu

informacji, ale wystarczyłoby uszkodzić zaledwie parę procent najbardziej usieciowionych wierzchołków (ang. network hub) tej sieci by nastąpił drastyczny spadek prędkości przesyłu danych aż do pełnego rozpadu sieci na izolowane obszary (Cohen et al. 2011).

Współcześnie opracowano wprowadzić metodę wyszukiwania wierzchołków o najwyższych stopniach (ang. tail-scope method) (Eom and Jo 2015), ale nie ma jeszcze poprawnie opracowanych metod jak sprawić by próba była reprezentatywna z punktu widzenia badanej właściwości/wielkości charakteryzującej wierzchołki sieci społecznej. Metoda ta bazuje na paradoksie przyjaźni (ang. friendship paradox) polegającym na tym, że losowo wybrana osoba posiada średnio mniej przyjaciół niż średnio jej przyjaciele. Paradoks ten jest konsekwencją ciężkoogonowego rozkładu stopni wierzchołków i stosuje się do wielu wielkości takich jak: zarobki, ilość partnerów seksualnych, ilość kontaktów telefonicznych i tak dalej.

Analiza sieci społecznościowych wkroczyła w nową erę w momencie szerokiego rozpowszechnienia Internetu i pojawienia się danych związanych z interakcjami między ludźmi (Ugander et al. 2011), (Zygmunt and Gliwa 2015). Dane te umożliwiają prowadzenie znacznie bardziej szczegółowych analiz, jednak tylko, jeśli chodzi o pewną klasę sieci.

Ponadto z analizą istniejących sieci społecznościowych wiąże się jeszcze jedna istotna wada. Wgląd w dynamikę tychże sieci, czyli to, w jaki sposób sieć może zareagować na określone zmiany lub bodźce, jest mocno ograniczony ze względów zarówno etycznych, jak i ekonomicznych. Obecnie tego rodzaju analizy można prowadzić prawie wyłącznie w sztucznym środowisku, wykorzystując symulacje lub gromadząc dane historyczne i zakładając, iż charakter oddziaływań socjo-ekonomicznych nie uległ znaczącej zmianie w badanym okresie czasu.

Wykorzystanie symulacji pozwala obserwować wpływ zmian parametrów na strukturę powstającej sieci. Technika symulacyjna jest stosunkowo tania, szybka i

umożliwia porównanie wyników uzyskanych przy różnych parametrach początkowych do danych zbieranych dla rzeczywistych sieci.

Podstawowym problemem symulacji pozostaje jednak wybór i sposób modelowania wszystkich uczestników i procesów koniecznych do uzyskania wystarczająco dużego podobieństwa do świata rzeczywistego.

Podobnie jak w przypadku większości modeli, sprawdza się tu podejście krok po kroku, gdzie kolejne elementy są oceniane, dodawane lub zmieniane tym samym ulepszając model i zmniejszając różnicę między wynikami rzeczywistymi a tymi uzyskanymi przez symulację.

Niemniej jednak zazwyczaj można zastosować kilka różnych modeli, a możliwość ich porównania zależy od dostępnych (lub wybranych) wartości obserwowanych. Ciekawy przykład struktury samoorganizującej się w oparciu o lokalne systemy agentowe przedstawiono w pracy (De Caux et al. 2014), tam, jednak, uwagę poświęcono, lokalnemu współczynnikowi klasteryzacji i występującemu w tworzonej sieci, zjawisku „małego świata”.

Wcześniejszą próbę budowy, sztucznych sieci społecznościowych opisano w artykule (Jin, Girvan, and Newman 2001), gdzie zaproponowano prostszy model, tworzący sieć o właściwościach zbliżonych do rzeczywistych sieci społecznościowych. Nawiązywanie więzi w tym modelu, odbywa się przez spotkania agentów. Niestety, aby otrzymać pożądane podobieństwo do rzeczywistych rozkładów, założono, że siła relacji podtrzymywana jest głównie przez wspólnych przyjaciół. Oznacza to, iż każdy agent musi posiadać informację: ile wspólnych znajomych ma z danym agentem i to właśnie ta wiedza decyduje o sile więzi agent-agent, w tym modelu. Stanowi to mało realistyczne założenie, gdyż w rzeczywistości ludzie dowiadują się, jakich znajomych posiada dana osoba dopiero, gdy, znają ją bardzo dobrze.

Model Piskor-Ignatowicza-Zachary, prezentowany w niniejszej pracy, realizuje tworzenie więzi między agentami wyłącznie w oparciu o cechy samych agentów, takie jak ich przyjacielskość i podobieństwo zainteresowań. Tym samym, stanowi dowód konstrukcyjny tworzenia dynamicznej sieci złożonej, w oparciu wyłącznie o oddziaływanie i własności samych agentów bez konieczności korzystania z lokalnych lub globalnych informacji, do których agenci nie mają (lub mogą nie mieć) dostępu.

Innymi słowy, w naszym modelu proces samoorganizacji jest w pełni spontaniczny i równoległy, jak ma to miejsce w układach złożonych obserwowanych w naturze.

4.2. Proponowany formalny model sieci społecznej.

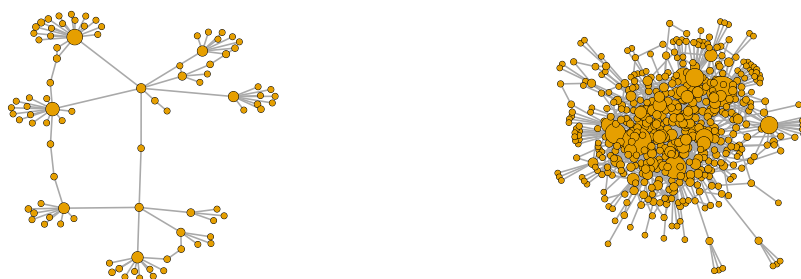
Model przedstawiony w niniejszej pracy oparto na analizie koncepcyjnej i poddano walidacji przez porównanie uzyskanych parametrów sieci wynikowej z wartościami zaobserwowanymi w prawdziwych sieciach społecznościowych i opublikowanymi w innych artykułach. Cechuje go elastyczność i wielość parametrów, które można wykorzystać do tworzenia szerokiej gamy sieci. Tworzenie i symulowanie sieci zależy od działań i interakcji poszczególnych agentów, przekładając się na samoorganizację sieci (podobnie jak w scenariuszach rzeczywistych). W ramach zaproponowanego tu podejścia, relacje lokalne pomiędzy agentami decydują o postaci tworzonej sieci społecznościowej. Co bardzo istotne, do jej powstania w wyniku niskopoziomowych oddziaływań agent-agent, nie są potrzebne żadne globalne informacje.

Podejście to daje również nowe możliwości badania dynamiki zmian zachodzących w sieci. Aby umożliwić tworzenie i rozwój wirtualnej sieci społecznościowej opracowano dedykowane środowisko symulacyjne. Środowisko to pozwala na wykorzystanie pół-autonomicznych agentów mogących ‘żyć’ i wchodzić w interakcje z innymi agentami. Podejście to wykorzystano już z powodzeniem w licznych symulacjach społecznych (Jin, Girvan, and Newman 2001), (Wąs and Lubaś 2014). Podstawowe zasady rządzące agentami, ich interakcjami i tworzeniem systemów wieloagentowych można znaleźć w kilku źródłach; np. (Lin 2007), dlatego też niniejsza praca przywołuje jedynie kluczowe aspekty środowiska symulacyjnego.

Każdy agent dysponuje właściwościami wpływającymi na jego zachowanie zgodnie z modelem agenta przedstawionym poniżej. Przykładowo „osobowość” agenta ma wpływ na liczbę i siłę możliwych do uzyskania przez niego i utrzymania relacji z innymi agentami.

Początkowo środowisko zasiedla zadana liczba agentów, a właściwości każdego z nich są przydzielane losowo zgodnie z zadaniem przedziałem i rozkładem w jego obrębie. Następnie, w każdym cyklu symulacji każdy z agentów jest wywoływany i wykonywana jest jego procedura. Każdy cykl składa się ze „spotkań” agentów z innymi agentami w środowisku wirtualnym i rozwoju ich obopólnych relacji, co

opisano bardziej szczegółowo w dalszej części niniejszego rozdziału. W toku symulacji agenci tworzą sieć relacji, co widać na rysunku Rys. 4-1 Przykładowa (pod)sieć relacji położonych nie dalej niż trzy przeskoki, dla losowo wybranego agenta. Po 20 początkowych cyklach symulacji (po lewej) i po 100 cyklach (po prawej).. Rycina ta przedstawia losowo wybraną część stworzonej sieci; tzn. losowo wybranego agenta i wszystkich pozostałych agentów, których odległość (rozumiana, jako ilość krawędzi w najkrótszej ścieżce patrz rozdział 3.2) od niego nie przekracza 3.



Rys. 4-1 Przykładowa (pod)sieć relacji położonych nie dalej niż trzy przeskoki, dla losowo wybranego agenta. Po 20 początkowych cyklach symulacji (po lewej) i po 100 cyklach (po prawej).

Widoczny na powyższym rysunku bardzo dynamiczny wzrost liczby więzi agenta pomiędzy cyklem $t = 20$ a cyklem $t = 100$ zachodzi jedynie na samym początku symulacji, gdy numer cyklu $t < t_{stab}$ i w miarę upływu czasu (rozumianego, jako numery cyklu symulacji), ulega spowolnieniu by w pełni ustabilizować się dla $t \approx t_{stab}$, podlegając jedynie fluktuacjom związanym z dynamiką tworzenia i rozpadu więzi pomiędzy agentami (bardziej szczegółowo zagadnienia te zostaną omówione w dalszej części tego rozdziału).

4.2.1. Model agenta

Modelowanie sieci społecznej polega na modelowaniu powiązań pomiędzy poszczególnymi osobnikami reprezentowanymi przez agentów. Formalnie, polega na

modelowaniu relacji, w jakie poszczególni agenci wchodzi z innymi agentami oraz dynamicznych zmian tych relacji.

W modelu zakłada się, że zbiór (populacja) agentów jest stały. Podobnie zakłada się, że agenci rozmieszczeni są (zamieszkują) w wyróżnionym, stałym dla danego eksperymentu symulacyjnego, obszarze o kształcie prostokątnym, o wymiarach odpowiednio x_{max} , y_{max} . Długości boków tego prostokąta stanowią parametry, które można zmieniać w kolejnych badaniach symulacyjnych.

W modelu uwzględnia się szereg wielkości charakteryzujących agentów, w szczególności uwzględnia się fakt, że mają oni różne „osobowości”. W przypadku analizy tworzenia i dynamiki zmian sieci społecznych „osobowość agenta” reprezentuje te jego cechy, które mają wpływ na model sieci, a więc pozyskiwanie znajomych i podtrzymywanie więzi z nimi. W przedstawionym poniżej modelu sieci, jako istotne właściwości wpływające na zachowanie agenta przyjęto obok jego osobowości E , lokalizację L (miejsce zamieszkania) oraz jego relacje z innymi agentami.

Formalnie, zbiór agentów będzie oznaczany, jako A (tak samo jak zbiór wierzchołków grafu). Każdego agenta $a \in A$ charakteryzują następujące wielkości:

$L(a)$ - lokalizacja agenta a określona współrzędnymi $(x(a), y(a))$, niezmiennymi w całym okresie symulacji T_{sym} .

$E(a)$ - zestaw wielkości charakteryzujących osobowość agenta, stałych w całym okresie symulacji T_{sym} .

Natomiast wielkości charakteryzujące stany agentów będą zmieniać się dynamicznie. Stan agenta w cyklu o numerze t , określa relacja znajomości/przyjaźni z innymi agentami oznaczana, jako $R_{known}^t(a)$ oraz charakterystyka tej relacji, określona w dalszej części tego rozdziału.

4.2.1.1. Lokalizacja i położenie agentów

W modelu **lokalizacje agentów** generowane są losowo, jako pozycje w dwuwymiarowej przestrzeni euklidesowej. Pozycje te są rozłożone losowo z rozkładem jednorodnym w rozważanej przestrzeni. W poszczególnych eksperymentach symulacyjnych (patrz rozdział 5) zarówno rozważany obszar jak i lokalizacje agentów mogą się zmieniać.

W modelu korzysta się z pojęcia globalnej gęstości agentów ρ_{glob} (liczba agentów na jednostkę kwadratową rozpatrywanej powierzchni) wyliczanej, jako:

$$\rho_{glob} = \frac{|A|}{S}$$

Gdzie $|A| = N$ -- moc zbioru agentów A , stanowiąca parametr w kolejnych eksperymentach symulacyjnych, zaś S - pole rozpatrywanego obszaru.

Uwzględnienie lokalizacji agenta $L(a)$ pozwala dla każdego agenta a , wyodrębnić podzbiór agentów będących jego sąsiadami, oznaczany, jako $N_{neig}(a)$. Wyróżnienie tego podzbioru ma znaczenie dla określania częstości jego spotkań z innymi agentami. Jako sąsiadów agenta a , rozumie się tych agentów, którzy mieszkają w odległości mniejszej niż tak zwany „zasięg sąsiedztwa” - r_{neigh} podawany, jako parametr symulacji.

Niech $d(a_j, a_k)$ oznacza odległość Euklidesową pomiędzy agentami a_j i a_k o lokalizacjach odpowiednio: $L(a_j) = (x(a_j), y(a_j))$ i $L(a_k) = (x(a_k), y(a_k))$.

Zbiór sąsiadów $N(a_j)$ agenta a_j jest zdefiniowany:

$$N_{neig}(a_j) = \{a_k: d(a_j, a_k) < r_{neigh}\}$$

Natomiast zbiór agentów z dalszego otoczenia („nie sąsiadów”) $O(a_j)$:

$$O(a_j) = \{a_k: d(a_j, a_k) \geq r_{neigh}\}$$

Oczywiście zachodzi $O(a_j) = A \setminus N_{neig}(a_j)$.

Dla każdego agenta a_j definiuje się ponadto tzw. lokalną gęstość sąsiadów - ρ_{loc} jako:

$$\rho_{loc}(a_j) = \frac{|N_{neig}(a_j)|}{\pi(r_{neigh})^2}$$

4.2.1.2. Osobowość agenta

Osobowość agenta może uwzględniać różne cechy (jak pokazano dalej w rozdziale 6 dotyczącym agentów kupujących), ale dwie z nich są istotne ze względu na tworzenie i utrzymywanie relacji z innymi agentami. Są to: towarzyskość (przyjacielskość) i zainteresowania. Im bardziej dany agent jest „towarzyski” tym łatwiej nawiązuje relacje z innymi agentami. W modelu uwzględnia się różne natężenia tej cechy za pomocą współczynnika przyjacielskości oznaczanego, jako $f(a)$. Współczynnik ten przyjmuje wartości procentowe z zakresu od 0% do 100%. Wartość $f(a)$ nadawana jest w sposób losowy każdemu agentowi, zgodnie z zadany rozkładem i pozostaje stała w trakcie danego eksperymentu symulacyjnego.

W stworzonym modelu przyjęto, że dla nawiązywania i podtrzymywania relacji, istotna jest, nie tylko zdolność agentów do zawierania znajomości, (której miarą jest współczynnik przyjaźni $f(a)$), ale także na ile „podobni” do siebie są agenci, a konkretnie, na ile podobne są ich zainteresowania. W tym celu, do modelu wprowadzono obszary zainteresowań, których ilość ustala się przy pomocy parametru - in . Wspólne zainteresowania agentów mają wpływ na możliwość powstania relacji między nimi i siłę tej relacji.

Zbiór zainteresowań i obszary zainteresowań stanowią rozszerzenie modelu pierwotnie przedstawionego w (Zachara, Majchrzyk-Zachara, and Piskor-Ignatowicz 2013). Celem wprowadzenia tych obszarów było przybliżenie środowiska świata rzeczywistego, gdzie ludzie różnią się profilami zainteresowań i łatwiej nawiązują relacje z osobami o zbliżonych zainteresowaniach. Jak zostanie pokazane dalej, liczba obszarów zainteresowań wpływa zauważalnie na wyniki symulacji i charakter powstających rozkładów relacji.

Niech rozpatrywany w modelu zbiór obszarów zainteresowań składa się z in różnych obszarów, gdzie in jest parametrem modelu.

Osobowość agenta definiujemy, jako:

$$E(a) = (f(a), w_1(a), w_2(a), \dots, w_{in}(a))$$

Gdzie: $f(a)$ – współczynnik „przyjaźni” agenta a , $w_i(a)$, $i = 1, 2, \dots, in$ – współczynnik (stopień) zainteresowania agenta a , i -tym obszarem zainteresowań, przy czym zachodzi warunek normalizacyjny:

$$\forall a \in A \sum_{i=1}^{in} w_i(a) = 1$$

Każdy eksperyment symulacyjny, rozpoczyna się od generacji zbioru agentów A , przy czym dla każdego agenta $a \in A$, generowane są losowo (zgodnie z przyjętym rozkładem) wartość przyjaźni $f(a)$ oraz wszystkie współczynniki zainteresowań $w_i(a)$, $i = 1, 2, \dots, in$ przy zachowaniu warunku normalizacyjnego.

4.2.1.3. Charakterystyka relacji między agentami

Więzi pomiędzy agentami modelowane są za pomocą dynamicznie zmieniającej się relacji.

Niech $R_{known}^t \subset A \times A$ oznacza relację znajomości (przyjaźni) pomiędzy wszystkimi agentami w cyklu t . W modelu przyjęto, iż relacja ta jest symetryczna, (na

chwile obecną jedynie takie dane empirycznych są dostępne, a więc założenie o symetrii relacji jest konieczne do zachowania weryfikowalności modelu). Formalnie:

$$\forall t \in T \forall a \in A \quad (a_i, a_j) \in R_{known}^t \Leftrightarrow (a_j, a_i) \in R_{known}^t$$

Każdy element (a_i, a_j) relacji R_{known}^t jest charakteryzowany dwoma wielkościami: siłą relacji w cyklu t oznaczaną s^t oraz czasem jej trwania g^t . Obie wielkości mogą się zmieniać w kolejnych cyklach na skutek spotkań agentów.

Zatem dla relacji R_{known}^t definiujemy dwie poniższe funkcje:

Funkcja $s^t: R_{known}^t \rightarrow [s_{threshold}, 1]$ określa procentową siłę więzi w cyklu t między każdą parą agentów $(a_i, a_j) \in R_{known}^t$. Zakłada się, że jest ona jednakowa dla obu agentów, w każdym cyklu t , czyli:

$$s^t(a_i, a_j) = s^t(a_j, a_i)$$

Siła więzi s^t , jest wzmacniana wskutek spotkań agentów i osłabiana w przypadku ich braku. Dla urealnienia modelu, wprowadzono parametr $s_{threshold} \in (0,1)$ określający progową wartość, poniżej której zakłada się, że więź pomiędzy agentami zostaje zerwana. W takiej sytuacji, czyli jeśli $s^t(a_i, a_j) < s_{threshold}$ to elementy (a_i, a_j) oraz (a_j, a_i) są usuwane ze zbioru R_{known}^t (i w konsekwencji, ze zdefiniowanych poniżej zbiorów $R_{known}^t(a_j)$ i $A_{known}^t(a_j)$).

Funkcja $g^t: R_{known}^t \rightarrow \mathbb{N}$, gdzie $\mathbb{N}$ oznacza zbiór liczb naturalnych. Funkcja g^t , określa czas trwania relacji (wiek relacji), rozumiany, jako ilość cykli od nawiązania relacji. Początkowo między agentami nie występują relacje i dopóki nie zostaną nawiązane to zarówno siła relacji jak i jej wiek równe są 0. Po powstaniu, relacje rozwijają się i zmieniają w sposób opisany w dalszej części rozdziału.

Na podstawie powyższych relacji można łatwo wyznaczyć, dla każdego agenta a_j zbiór jego znajomych w cyklu t oznaczony, jako $A_{known}^t(a_j)$:

$$A_{known}^t(a_j) = \{a_k \in A: (a_j, a_k) \in R_{known}^t\}$$

Oznaczmy $R_{known}^t(a_j)$ - podzbiór relacji R_{known}^t zawierający te pary, których pierwszym elementem jest a_j , zaś $s^t/R_{known}^t(a_j)$ oraz $g^t/R_{known}^t(a_j)$ restrykcję odpowiednio funkcji s^t oraz g^t do podzbioru $R_{known}^t(a_j)$.

Stan agenta a_j w cyklu t zdefiniowany jest przez:

zbiór $A_{known}^t(a_j)$, funkcję $s^t/R_{known}^t(a_j)$ oraz funkcję $g^t/R_{known}^t(a_j)$.

4.2.2. Dynamika rozwoju sieci

Modelowanie rozwoju sztucznej sieci społecznościowej ma przybliżać powstawanie i utrzymywanie się relacji znajomości w środowisku świata rzeczywistego, zatem odbywa się w oparciu o koncepcję spotkań/kontaktów i wynikających z nich dynamicznych zmian relacji między uczestnikami. W wyniku pierwszego spotkania relacja pomiędzy agentami może być nawiązana, a w wyniku kolejnych spotkań będzie wzmacniana.

Jednak w świecie rzeczywistym występują procesy odpowiedzialne zarówno za tworzenie jak i za rozpad relacji znajomości. Stabilny system można uzyskać tylko, jeśli między siłami tymi zachodzi równowaga. Gdyby nie istniały procesy niszczące relacje w końcu każdy agent miałby znajomość z każdym innym agentem, co oczywiście jest nierealistyczne i niepożądane. Zatem w modelu założono, że w przypadku braku spotkań, relacja znajomości jest osłabiana, a po pewnej ilości cykli bez spotkań może zniknąć.

Celem obliczeń modelu w cyklu t jest wyznaczenie dla wszystkich agentów: $a_j \in A$, nowych wartości ich stanów, które będą obowiązywały w kolejnym cyklu $(t + 1)$, to jest wyznaczenie nowych zbiorów $A_{known}^{t+1}(a_j)$, oraz nowych funkcji: $s^{t+1}/R_{known}^{t+1}(a_j)$, i $g^{t+1}/R_{known}^{t+1}(a_j)$.

4.2.2.1. Spotkania agentów

Modelowane spotkania agenta a z innymi agentami zależą od przynależności do zdefiniowanych wcześniej zbiorów sąsiadów $N_{neig}(a)$ (lub zbioru pozostałych $O(a)$) oraz zbioru znajomych $A_{known}^t(a)$, w poszczególnych cyklach.

Idea modelowania spotkań dla każdego cyklu $t: 1 < t \leq T_{sym}$, jest następująca

Dla potrzeb symulacji wprowadza się parametr V_{mm} określający nieprzekraczalną liczbę spotkań w pojedynczym cyklu, której żaden agent nie może przekroczyć.

Natomiast maksymalna ilość spotkań, jakie dany agent a może odbyć w jednym cyklu jest różna dla różnych agentów i zależy od wartości ich współczynnika przyjaźliwości $f(a)$ i wynosi $[f(a)V_{mm}]$.

W każdym cyklu agenci angażują się w pewną ilość spotkań przypadkowych oraz intencjonalnych (zaplanowanych). **Spotkania przypadkowe** oznaczają losowe

spotkania z losowo wybranymi agentami. **Spotkania intencjonalne**, natomiast dotyczą znajomych (przyjaciół) agenta.

W każdym cyklu t , każdy agent a_j , dokonuje $m_{los}(t, a_j) \in \mathbb{N}$, spotkań przypadkowych oraz $m_f(t, a_j) \in \mathbb{N}$ spotkań intencjonalnych ze znajomymi. Łączna ilość spotkań, jaką może odbyć agent a_j w cyklu t symulacji, wyraża się wzorem:

$$\forall a_j \in A \forall t \in T \quad m_{los}(t, a_j) + m_f(t, a_j) = [f(a_j)V_{mm}]$$

Spotkania przypadkowe dotyczą sąsiadów (agentów ze zbioru $N_{neig}(a_j)$) oraz agentów z dalszego otoczenia (ze zbioru $O(a_j)$):

W każdym cyklu t danej symulacji, dla każdego agenta a_j losowana jest liczba spotkań z sąsiadami: $n_{los_{neigh}}(t, a_j)$ oraz liczba spotkań z agentami dalszego otoczenia: $n_{los_{other}}(t, a_j)$

przy czym zachodzi:

$$\forall t \in T \quad n_{los_{neigh}}(t, a_j) + n_{los_{other}}(t, a_j) = m_{los}(t, a_j) < [f(a_j)V_{mm}]$$

W symulacji wprowadzono parametry $c_{los_{neigh}}$ oraz $c_{los_{other}}$ pozwalające sterować średnią liczbą spotkań agenta a_j , zarówno z jego sąsiadami jak i agentami znajdującymi się w odległości większej niż promień sąsiedztwa. W kolejnych cyklach symulacji $n_{los_{neigh}}(t, a_j)$ oraz $n_{los_{other}}(t, a_j)$ przyjmują różne wartości, przy czym ich średnie $\langle n_{los_{neigh}}(a_j) \rangle$, $\langle n_{los_{other}}(a_j) \rangle$ pozostają proporcjonalne do parametrów symulacji $c_{los_{neigh}}$ oraz $c_{los_{other}}$, formalnie:

$$\langle n_{los_{neigh}}(a_j) \rangle \sim c_{los_{neigh}}$$

$$\langle n_{los_{other}}(a_j) \rangle \sim c_{los_{other}}$$

Mechanizmem mającym znaczący wpływ na strukturę wynikowej sieci społecznej są spotkania, w których bierze udział trzech agentów. W modelu podobnie jak w realnym świecie na spotkanie ze znajomym można przyjść ze swoim przyjacielem. Aby zrealizować taki scenariusz w modelu wprowadzono parametr V_{pf} – sterujący częstotliwością spotkań z „przyjaciółmi przyjaciół”.

Gdy dochodzi do spotkania obcych agentów a_m z a_n , w cyklu symulacji t , czyli $(a_m, a_n) \notin R_{known}^{t-1}$, relacja znajomości zostanie utworzona z prawdopodobieństwem:

$$P_{relation}(a_m, a_n) = \frac{f(a_m) + f(a_n)}{2}$$

Nowo utworzonej relacji nadawana jest początkowa wartość siły relacji:

$$s^t(a_m, a_n) = s_{start}$$

gdzie s_{start} – parametr symulacji

Równocześnie, dla każdej nowopowstałej relacji przypisywany jest jednostkowy wiek, czyli:

$$g^t(a_m, a_n) = 1$$

W przypadku agentów znanych, czyli gdy $(a_i, a_j) \in R_{known}^{t-1}$, każde kolejne spotkanie wzmacnia siłę relacji o wartość:

$$\Delta s^t(a_i, a_j) \sim \frac{w_{mutual}(a_i, a_j)}{G_{time} g^t(a_i, a_j)}$$

Gdzie: G_{time} – parametr symulacji odpowiadający za czułość zmian relacji z wiekiem, $g^t(a_i, a_j)$ – wiek relacji znajomości agenta a_i , z agentem a_j , mierzony jako ilość cykli trwania relacji, $w_{mutual}(a_i, a_j)$ – miara podobieństwa zainteresowań agentów a_i oraz a_j , postaci:

$$w_{mutual}(a_i, a_j)$$

$$= \frac{1}{\min(w_1(a_i), w_1(a_j), \dots, w_{in}(a_i), w_{in}(a_j))} \sum_{k=1}^{k=in} \max(w_k(a_i), w_k(a_j))$$

Po zrealizowaniu powyższych obliczeń, dla każdego agenta a_j w danym cyklu t modyfikowany jest przejściowo zbiór jego znanych $A_{known}^t(a_j)$, do postaci:

$$A_{known}^{*t}(a_j) = \{a_k : s^t(a_j, a_k) > s_{threshold}\}$$

Spotkania intencjonalne:

Dla każdego agenta a_j elementy zbioru $A_{known}^{*t}(a_j)$, porządkowane są w ciąg nierosnący ze względu na siłę relacji $s^t(a_j, a_k)$ i agent a_j dokonuje spotkań z $m_f(t, a_j)$ pierwszymi agentami z ciągu swoich znanych (czyli tymi agentami z którymi siła relacji jest największa).

4.2.2.2. Osłabianie siły relacji – brak spotkań

Jak dotąd omówiono rozmaite czynniki związane z powstawaniem i wzmacnianiem relacji między agentami. Jak wspomniano, dla stabilności modelu wprowadzono też mechanizm osłabiania relacji znajomości, aż do jej rozpadu (zaniku).

Po dokonaniu wszystkich obliczeń modelujących spotkania agenta a_j w danym cyklu t (przypadkowe i intencjonalne) wyznaczany jest zbiór agentów $A_{meet}^t(a_j)$ oraz $A_{non-meet}^t(a_j)$ których zna agent a_j , a z którymi odpowiednio nastąpiło lub nie nastąpiło spotkanie w tym cyklu.

$$A_{non-meet}^t(a_j) = A_{known}^{*t}(a_j) \setminus A_{meet}^t(a_j)$$

Siła relacji znajomości agenta a_j , ze wszystkimi agentami ze zbioru $A_{non-meet}^t(a_j)$ zostaje zmniejszona o wartość:

$$\delta s^t(a_j, a_i) = s^t(a_j, a_i)(1 - s_{decay})$$

gdzie: s_{decay} – parametr symulacji odpowiadający za słabnięcie relacji.

Jeśli dla któregoś z agentów znanych z agentem a_j siła relacji, spadnie poniżej wartości progowej s_{tresh} , to taki agent zostaje usunięty ze zbioru znanych agenta a_j .

4.3. Podsumowanie

Przedstawiony w rozdziale 4 wieloagentowy model tworzenia sztucznej sieci społecznościowej wykorzystano następnie, jako środowisko do działania agentów sprzedawców i kupujących. W rozdziale 5 pokazano wpływ zmian parametrów modelu na otrzymywane sieci wynikowe.

5. Koncepcja badań symulacyjnych sztucznych sieci

Rozdział ten przedstawia koncepcję przeprowadzonych badań symulacyjnych dla zaproponowanego w rozdziale 4. formalnego modelu sztucznych sieci społecznościowych. Celem badania jest analiza wpływu zmian parametrów modelu na wynikową charakterystykę sieci (patrz Tabela 5-1). Badane będą estymatory rozkładów stopni wierzchołków sieci (ilości znajomych poszczególnych agentów). Jako rozkład referencyjny przyjmowany jest rozkład potęgowy, ze względu na fakt, że dostępne dane o rzeczywistych sieciach społecznościowych były charakteryzowane (opisywane) tym rozkładem. Badano również średni współczynnik gronowania $\langle C \rangle$ generowanych przez model sieci.

Badania symulacyjne przeprowadzono w następujący sposób:

1. Ustalono, po jakiej liczbie cykli T_{stab} stabilizują się wielkości charakteryzujące tworzoną sztuczną sieć społecznościową takie jak: rozkład stopni wierzchołków czy współczynnik gronowania.
2. Zbadano metodą Monte Carlo czy liczba cykli z punktu 1 nie zależy od parametrów modelu.
3. Poprzez wielokrotne generowanie sieci społecznościowych przy ustalonych parametrach wejściowych, zbadano stabilność charakterystyk opisujących wynikowe sztuczne sieci społecznościowe.
4. Przy ustalonych pozostałych parametrach modelu zmieniano z ustalonym krokiem jeden z parametrów modelu badając jego wpływ na wielkości charakteryzujące generowaną sieć dla ilości cykli większej niż T_{stab} .

Do charakterystyki sieci złożonych reprezentujących sztuczne społeczeństwo zastosowano wielkości zestawione w tabeli Tabela 5-1.

Tabela 5-1 Zestawienie wielkości używanych w niniejszej pracy przy charakteryzowaniu sieci złożonych.

Wybrane parametry zastosowane do charakterystyki generowanych sieci złożonych, (dokładniejszy opis użytych wielkości umieszczono poniżej)	
α	Estymator parametru skalowania dla rozkładu potęgowego stopni wierzchołków sieci
x_{min}	Parametr odcięcia, określa minimalną wartość $x \geq x_{min}$ powyżej której stosowalny jest rozkład potęgowy
$< C >$	Średni współczynnik gronowania (klasteryzacji) sieci
$p - \text{wartość}$	Wynik testu zgodności

x_{min} to wartość począwszy, od której dana wielkość podlega określonemu rozkładowi prawdopodobieństwa (w tym wypadku rozkład stopni wierzchołków podlega rozkładowi potęgowemu). Jak wspomniano w pracy (Clauset, Shalizi, and Newman 2009), potęgowe prawo skalowania stosuje się często tylko do wartości przekraczających pewne x_{min} dla danych empirycznych. Ten sam artykuł zaleca stosowanie testu Kołmogorowa-Smirnowa (Press et al. 1992) do oszacowania takiej wartości x_{min} przy której dopasowanie rozkładu potęgowego do danych doświadczalnych jest najlepsze.

Estymator parametru skalowania α , otrzymano metodą największej wiarygodności, zgodnie z zaleceniami zawartymi w artykule (Clauset, Shalizi, and Newman 2009) dla danych dyskretnych. Jest on postaci:

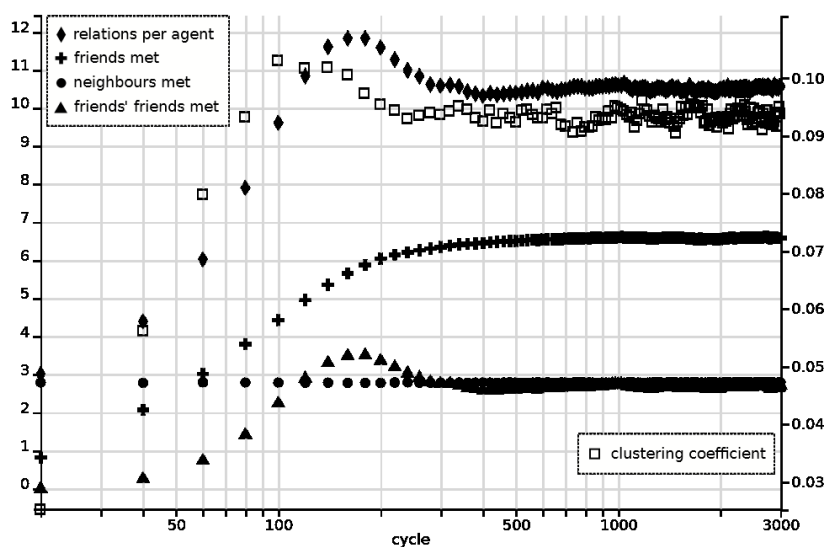
$$\alpha \cong 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min}-0.5} \right]^{-1} \quad (5-1)$$

gdzie n – ilość danych empirycznych podlegających rozkładowi potęgowemu, $x_i: i = 1, 2, \dots, n$ – dane doświadczalne, x_{min} – progowa wartość powyżej której dane podlegają rozkładowi potęgowemu.

Test zgodności jest standardową metodą stosowaną do oceny hipotezy, zgodnie, z którą dane obserwowane przybierają pewien rozkład. Przy obliczaniu odległości danych ocenianych od rozkładów wygenerowanych sztucznie, można obliczyć p – wartość (ang. p – value) (Wasserstein and Lazar 2016), określającą wiarygodność hipotezy. Dla celów niniejszej pracy p – wartość wyliczono za pomocą metody bootstrap z pakietu *R* *powerLaw* (Gillespie 2015), i przyjęto, że p – wartość $> 0,1$ wskazuje, iż potęgowe prawo skalowania jest hipotezą wiarygodną (Clauset, Shalizi, and Newman 2009).

Średni sieciowy współczynnik klasteryzacji (gronowania) jest wartością średnią lokalnego współczynnika klasteryzacji. Klasteryzacja lokalna to stosunek liczby istniejących trójkątów relacji między określonym agentem a innymi, które mają z nim relacje, do liczby wszystkich trójkątów możliwych do utworzenia dla tego agenta i jego sąsiadów.

W celu zbadania jak zmieniają się w funkcji numeru cyklu symulacji t , wybrane wielkości charakteryzujące tworzoną sieć społecznościową, wyliczano je, co dwadzieścia cykli symulacji. Otrzymane wielkości przedstawiono w formie graficznej na ilustracji Rys. 5-1.



Rys. 5-1 Na wykresie przedstawiono jak zmieniają się wybrane wielkości w zależności od numeru cyklu symulacji t , szczegółowy opis w tekście.

Łatwo zauważyć, że zaprezentowane wielkości osiągają swoje maksima pomiędzy setnym a sto pięćdziesiątym cyklem i stabilizują się powyżej trzechsetnego cyklu, podlegając dalej jedynie niewielkim fluktuacjom. Można tym samym przyjąć, że stworzony model sztucznego społeczeństwa osiąga dynamiczną stabilność parametrów statystycznych dla ilości cykli $T_{stab} \sim 300$ cykli. Wielkości przedstawione powyżej na Rys. 5-1 to: $\blacklozenge$ - średnia liczba relacji tworzonych przez agentów, $\square$ - średni współczynnik gronowania (klasteryzacji, uwaga: wartości dla współczynnika klasteryzacji należy odczytywać korzystając z prawej skali na rysunku), $+$ - średnia ilość spotkań intencjonalnych w danym cyklu, $\bullet$ - średnia liczba spotkań sąsiadów, $\blacktriangle$ - średnia liczba spotkań, na które znajomy agent przyprowadza swojego znajomego.

W tabeli Tabela 5-2 poniżej zamieszczono przykładowe wartości parametrów wejściowych modelu sztucznych społeczności oraz zaprezentowano wybrane charakterystyki uzyskanych układów złożonych reprezentowanych, jako jednowarstwowe sieci złożone. Na potrzeby porównania wygenerowanej sztucznej sieci społecznościowej do dostępnych danych literaturowych, dopasowano potęgowy rozkład stopni wierzchołków wygenerowanej sieci.

Tabela 5-2 Przykładowe wartości startowe wybranych parametrów symulacji oraz otrzymane wartości wynikowe dla sześciu eksperymentów symulacyjnych.

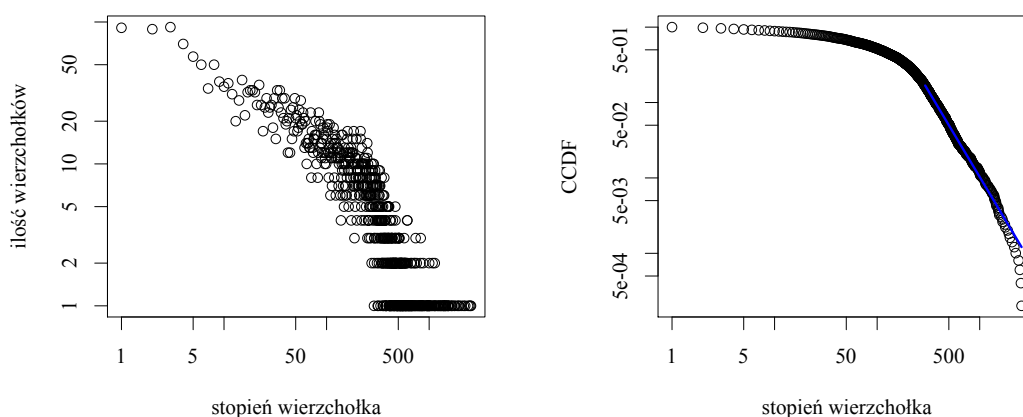
Wybrane parametry symulacji		Wyniki symulacji	
Rozmiar populacji $ A $	5000	α	3.4 ± 0.2
Ilość cykli symulacji T	3000	x_{min}	76 ± 29
Ilość obszarów zainteresowań ni	15	$p - \text{wartość}$	$0.56(> 0.1)$
Maksymalna ilość spotkań V_{mm}	20	Średni współczynnik gronowania $< C >$	0.095 ± 0.05
Parametr słabnięcia relacji S_{decay}	0.02	Średnia ilość spotkań obcych agentów w cyklu t	1.00 ± 0.11
c_{los_neigh}	0.01	Średnia ilość spotkań sąsiadów w cyklu t	2.79 ± 0.24
c_{los_other}	0.0002	Średnia ilość spotkań przyjaciół w cyklu t	6.69 ± 0.73
Zasięg sąsiedztwa r_{neighb}	0.1	Średnia ilość spotkań przyjaciół swoich przyjaciół	2.71 ± 0.29
Szansa spotkania przyjaciela swojego przyjaciela	0.05		

W niniejszym rozdziale omówiony zostanie wpływ parametrów symulacji na uzyskiwane rozkłady relacji celem określenia zakresu parametrów dających sieci

o właściwościach rozkładu relacji podobnych jak w sieciach naturalnych (Clauset, Shalizi, and Newman 2009).

Dla celów referencyjnych przedstawiono i omówiono bardziej szczegółowo pojedynczy scenariusz symulacji. Scenariusz ten oparto na wartościach parametrów symulacji dość bliskich środkom przedziałów koniecznych do uzyskania wyników zawierających się w granicach rozkładów obserwowanych w świecie rzeczywistym. Kluczowe parametry symulacji wymieniono w Tabeli 5-2.

Wynikowy rozkład ilości relacji tworzonych przez agentów (stopni wierzchołków) przedstawiono w lewej części ilustracji Rys. 5-2, a po prawej stronie zaprezentowano odpowiadający tym danym CCDF (ang. Complementary Cumulative Distribution Function) wraz z dopasowanym rozkładem potęgowym.



Rys. 5-2 Po lewej stronie ilustracji przedstawiono rozkład stopni wierzchołków w sztucznej sieci społecznościowej tworzonej przez 5000 agentów. Po prawej stronie uwidoczniono jeden minus dystrybuenta eksperymentalna (CCDF) tej samej sieci wraz z dopasowaniem rozkładu potęgowego. Szczegółowy opis umieszczono w tekście poniżej.

Zwróćmy uwagę, że:

$$CCDF = 1 - F(x) = 1 - P(X \leq x) = P(X > x) \quad (5-2)$$

czyli w naszym konkretnym przypadku $CCDF = P(K > k)$ gdzie k stopień wierzchołka (konkretna wartość), K – zmienna losowa odpowiadająca stopniowi wierzchołka.

Jak widać na wspomnianej rycinie, rozkład ma ciężki ogon i jest bardzo podobny do tych obserwowanych w rzeczywistych sieciach bezskalowych (Clauset, Shalizi, and Newman 2009). Parametry rozkładu (przedstawione w Tabeli 5-2 jako *wyniki*

symulacji) także potwierdzają jego podobieństwo do wspomnianych już sieci bezskalowych, gdyż współczynnik α mieści się w przedziale obserwowanym dla tego rodzaju sieci i jednocześnie wynik testu zgodności p – wartość $> 0,1$ pociąga za sobą nikłe prawdopodobieństwo losowego charakteru otrzymania dobrego dopasowania rozkładu ciężkoogonowego.

Przykładowo obserwowane parametry można porównać do zbiorów e-mailowych książek adresowych opisanych w (M. E. J. Newman, Forrest, and Balthrop 2002) i Tabeli 6.1 w (Clauset, Shalizi, and Newman 2009). Dla sieci o bardzo podobnych rozmiarach (4581 autentycznych e-mailowych książek adresowych) i charakterze (kontakty) parametry rozkładu są również bardzo zbliżone:

	Populacja	α	x_{min}	x_{max}	$< C >$
Rozmiar książki	4581	3.5 ± 0.6	57 ± 21	333	0.13
Wyniki symulacji	5000	3.4 ± 0.2	76 ± 29	417	0.09

Najłatwiej zauważalną różnicą jest niższy współczynnik klasteryzacji (0.09 wobec 0.13), który pełniej omówiony zostanie we wnioskach niniejszego rozdziału.

Prezentowany w niniejszej pracy model wieloagentowy sztucznego społeczeństwa został stworzony na podstawie analizy koncepcyjnej zachowań społecznych ludzi i licznych symulacji. W dalszej części rozdziału przedstawiam wyniki badań symulacyjnych wpływu zmiany parametrów modelu na właściwości wynikowych sieci społecznych. Zaprezentowane poniżej wyniki analizy właściwości modelu zostały już opublikowane wraz z prezentowanymi wykresami w czasopiśmie Computing in Science & Engineering (Zachara and Piskor-Ignatowicz 2016).

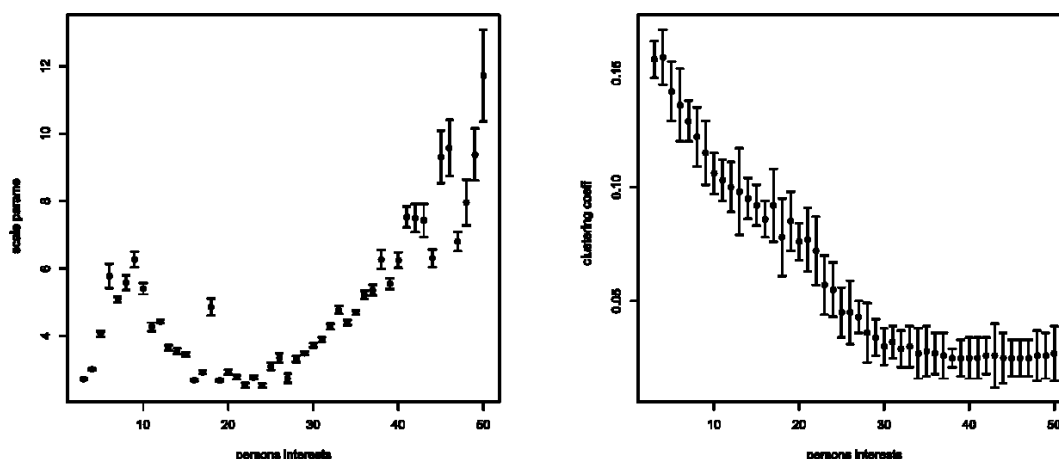
5.1. Parametry modelu odpowiadające za powstawanie więzi między agentami

Pośród parametrów używanych w prezentowanym w niniejszej pracy modelu sztucznego społeczeństwa można wydzielić te z nich, które mają związek z nawiązywaniem więzi między agentami, jak i te, które wiążą się z zanikiem znajomości. Do pierwszej z grup można zaliczyć: ilość obszarów zainteresowań in , nieprzekraczalną liczbę spotkań V_{mm} (jakie agent może odbyć w danym cyklu), parametr V_{pf} – sterujący prawdopodobieństwem spotkania „przyjaciela naszego przyjaciela”, promień sąsiedztwa r_{neighb} oraz parametr c_{los_other} sterującego w symulacji prawdopodobieństwem spotkania obcego agenta.

W celu zwiększenia czytelności grupa parametrów odpowiadających zanikowi znajomości zostanie omówiona w kolejnym podrozdziale.

Pierwszym z omawianych tu parametrów modelu jest liczba obszarów zainteresowań agentów (in). Zgodnie z właściwościami agentów omówionymi w poprzednim rozdziale, proces nawiązania znajomości pomiędzy dwoma agentami jest tym bardziej prawdopodobny im bardziej podobne są ich zainteresowania. W przeprowadzonych eksperymentach symulacyjnych, osobowości agentów były generowane losowo z jednostajnym (zwanym też jednorodnym – ang. uniform distribution) rozkładem cech, w szczególności tak też generowano rozkład zainteresowań każdego z agentów. W efekcie zwiększanie wartości parametru ni skutkuje większą różnorodnością profili zainteresowań agentów, a tym samym zmniejsza szanse na spotkanie agentów z podobnymi profilami zainteresowań, co przekłada się na charakterystyki tworzonych sieci społecznościowych.

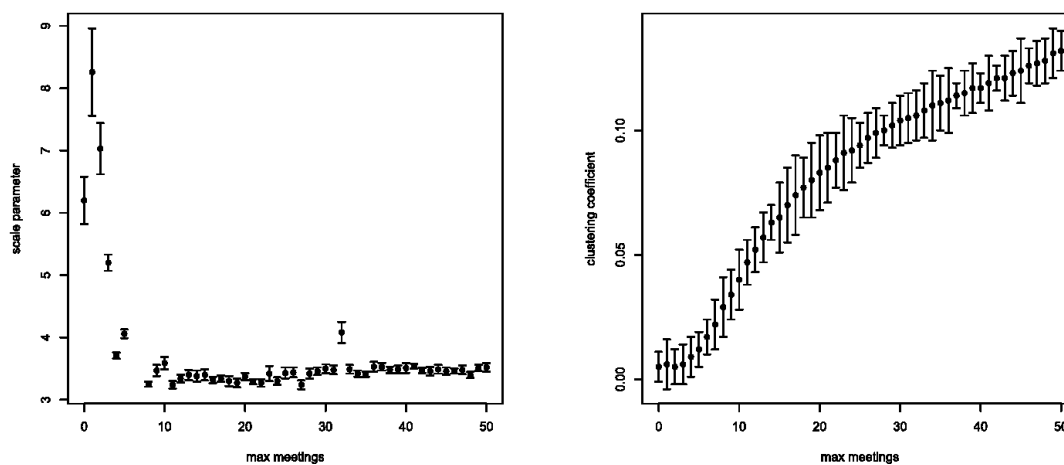
Poniżej przedstawiono zależność wartości estymatora wykładnika rozkładu potęgowego stopni wierzchołków α i średniego współczynnika gronowania $\langle C \rangle$ tworzonej sieci wynikowej w funkcji in Rys. 5-3.



Rys. 5-3 Na lewym wykresie przedstawiono zmianę wartości estymatora wykładnika α dopasowanego rozkładu potęgowego stopni wierzchołków wynikowej sieci w funkcji ilości obszarów zainteresowań in (persons interests). Po prawej stronie przedstawiono wpływ zmiany tejże ilości obszarów zainteresowań na średni współczynnik gromadzenia $\langle C \rangle$ tworzonych sieci.

Im większa jest wartość parametru in tym mniej prawdopodobne staje się spotkanie agentów z pasującymi do siebie zainteresowaniami, czyli tym mniejsza jest szansa na utworzenie relacji między agentami. Warto zauważyć, że zmiana współczynnika skalowania α nie jest monotoniczna ze zmianą ilości obszarów zainteresowań agentów in , tworząc maksimum lokalne około wartości $in = 10$ i minimum lokalne około $in = 25$. Obserwowany charakter zmian wartości wykładnika α w funkcji in jest związany ze zmianami w strukturze tworzonych sieci społecznościowych.

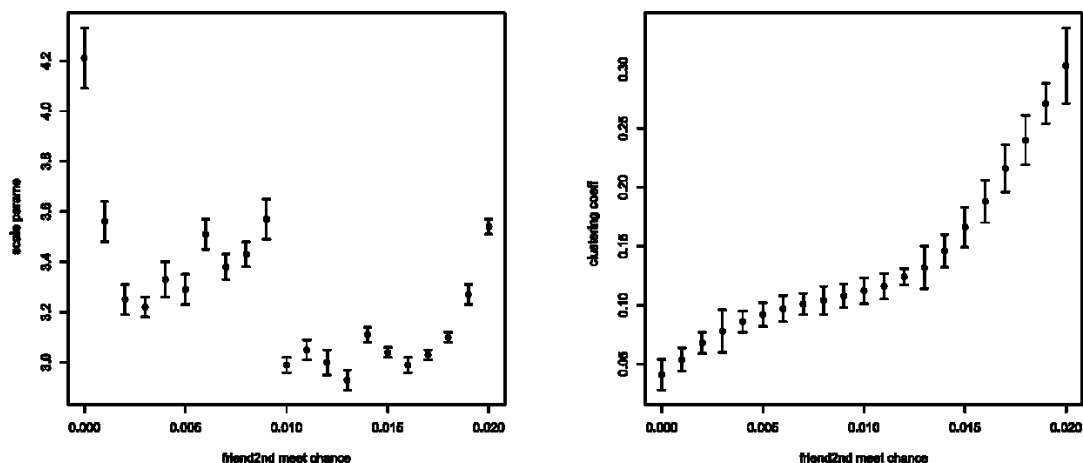
Wpływ parametru symulacji V_{mm} , - (maksymalnej liczby spotkań), jakie każdy agent może odbyć w cyklu t na charakterystyki wynikowej sieci przedstawiono na poniższej ilustracji Rys. 5-4.



Rys. 5-4 Wpływ maksymalnej liczby spotkań V_{mm} jaką może odbyć każdy agent w pojedynczym cyklu symulacji: na wykładnik α dopasowywanego rozkładu potęgowego (po lewej stronie) oraz na średni współczynnik klasteryzacji C (po prawej stronie ilustracji) dla symulowanych sieci wynikowych.

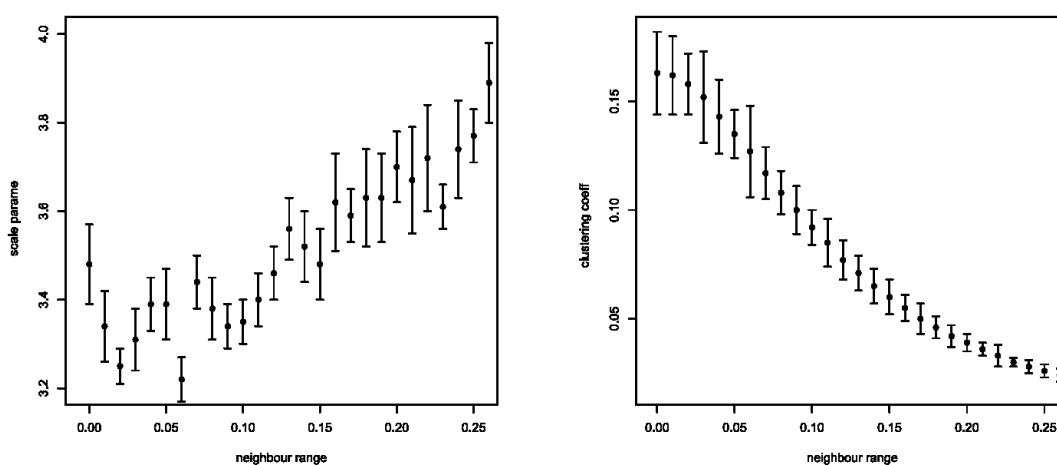
Liczba spotkań, jakie agent może odbyć w każdym cyklu symulacji ma wpływ głównie na współczynnik klasteryzacji. Dla bardzo niskich wartości parametru V_{mm} , współczynnik skalowania α - przyjmuje bardzo wysokie wartości, wskazując na nienaturalną strukturę sieci ze względu na brak możliwości nawiązywania wystarczająco wielu relacji między agentami. Takie zachowanie modelu jest zgodne z obserwacjami doświadczalnymi. U ludzi, którzy mają mniej czasu (np. bardzo dużo pracują, oglądają nadmiernie telewizję czy spędzają zbyt wiele czasu w Internecie (Nie, Hillygus, and Erbring 2000)) następuje osłabianie ich więzi społecznych. Gdyż ilość czasu dostępna w ciągu doby dla każdego człowieka jest stała, a wymienione powyżej aktywności zabierają czas dostępny na rzeczywiste kontakty. W przypadku Internetu o ile czas jest poświęcany na udzielanie się w mediach społecznościowych to nastąpi skurczenie się sieci kontaktów rzeczywistych na korzyść wzrostu liczby kontaktów wirtualnych.

Mechanizmem mającym znaczący wpływ na strukturę wynikowej sieci społecznej są spotkania, w których bierze udział trzech agentów. W modelu podobnie jak w realnym świecie na spotkanie ze znajomym można przyjść ze swoim przyjacielem. Aby zrealizować taki scenariusz w modelu wprowadzono parametr V_{pf} – sterujący częstotliwością spotkań z „przyjaciółmi przyjaciół”. Na ilustracji Rys. 5-5 pokazano wpływ zmian wartości parametru V_{pf} na charakterystyki tworzonej sieci.



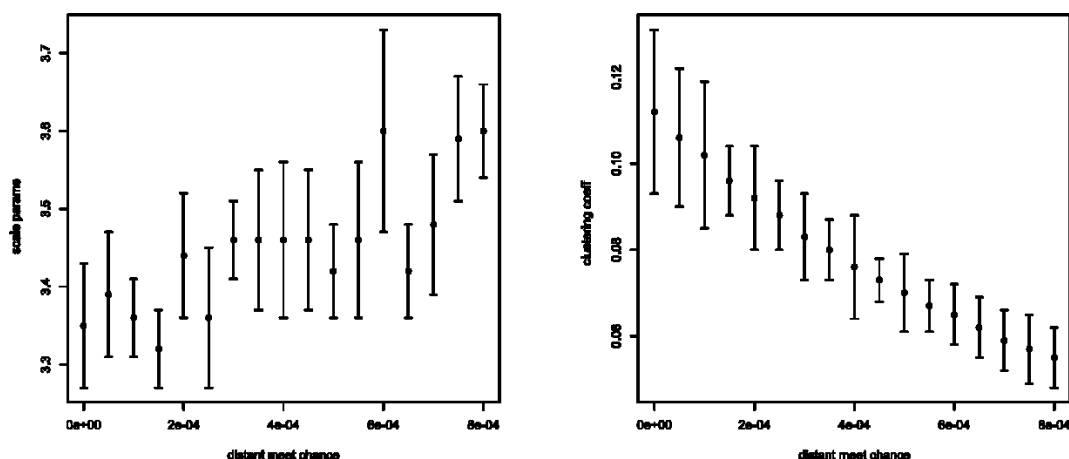
Rys. 5-5 Na lewym wykresie przedstawiono zależność dopasowanego wykładnika α rozkładu potęgowego stopni wierzchołków w generowanej sieci, w funkcji wartości parametru modelu V_{pf} (friend2nd meet chance), a na prawym wykresie pokazano, odpowiadający tworzonej sieciom średni współczynnik gronowania $\langle C \rangle$.

W stworzonym modelu sztucznego społeczeństwa mechanizm spotkań rozważono bardziej szczegółowo niż tylko, jako zależność procesu od nieprzekraczalnej ilości spotkań V_{mm} , które może odbyć agent w danym cyklu symulacji t . A konkretnie, wydzielono dwa współzależne mechanizmy odpowiedzialne za spotkania agenta: składową „przypadkową” oraz wolicjonalną. Do pierwszej grupy należą przypadkowe spotkania sąsiadów i obcych, gdyż wiążą się one z codzienną aktywnością agenta i zachodzą jakby „przy okazji” (nie są wynikiem planowania). Częstość spotkań sąsiadów jest funkcją parametru symulacji r_{neig} - promienia sąsiedztwa. Wpływ parametru r_{neig} na wynikowe charakterystyki tworzonej sieci przedstawiono na ilustracji poniżej Rys. 5-6.



Rys. 5-6 Na lewym wykresie przedstawiono zależność dopasowanego do rozkładu stopni wierzchołków w wynikowej sieci wykładnika potęgowego α w funkcji parametru symulacji r_{neig} (neighbour range), a na prawym wykresie pokazano średni współczynnik gronowania $\langle C \rangle$ wynikowych sieci społecznościowych w funkcji promienia sąsiedztwa r_{neig} .

Poniżej na rysunku przedstawiono zmianę charakterystyk wynikowych sieci społecznych w funkcji parametru c_{los_other} sterującego w symulacji częstością spotkań agenta z obcymi mu agentami.



Rys. 5-7 Na lewym wykresie pokazano zależność dopasowanego wykładnika α od wartości parametru symulacji c_{los_other} sterującego w symulacji prawdopodobieństwem spotkania obcego (distant meet chance), a na prawym wykresie przedstawiono zależność średniego współczynnika gronowania wynikowych sieci $\langle C \rangle$ od tegoż parametru.

Ponieważ nieprzekraczalną liczbę spotkań, jaką może odbyć agent, określa parametr V_{mm} , to zwiększenie liczby spotkań przypadkowych zmniejsza automatycznie liczbę możliwych do odbycia przez danego agenta spotkań intencjonalnych (wolincjonalnych), co znacząco wpływa na właściwości tworzonych w modelu sieci społecznościowych powodując spadek ich średniego współczynnika klasteryzacji $\langle C \rangle$.

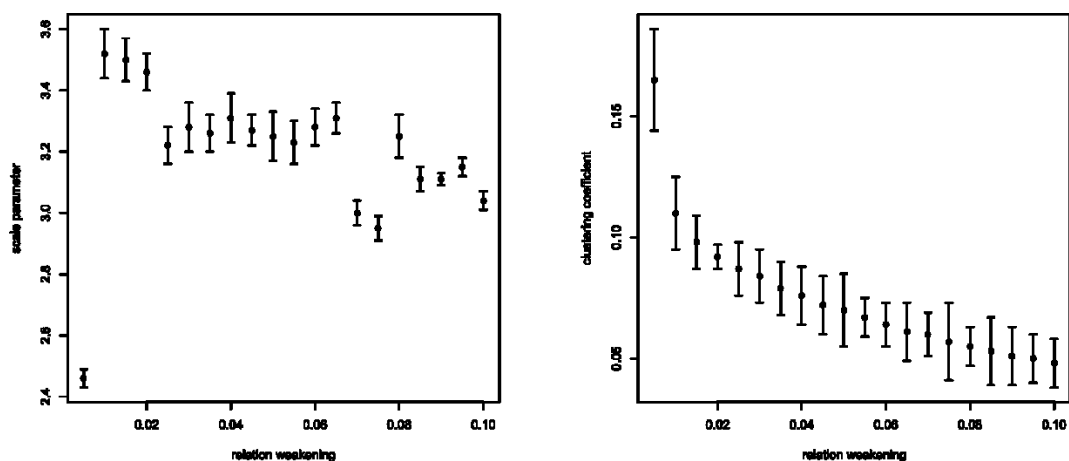
Podsumowując: dla zadanej nieprzekraczalnej liczby spotkań - V_{mm} , jaką może w danym cyklu t odbyć agent, tym większy będzie średni współczynnik klasteryzacji sieci wynikowej $\langle C \rangle$, im więcej spotkań intencjonalnych („zaplanowanych” patrz rozdział 4) może odbyć agent.

5.2. Parametry modelu odpowiedzialne za rozpad więzi między agentami

Podobnie jak w realnym świecie nie każde spotkanie się ludzi skutkuje nawiązaniem nowej znajomości tak i w przedstawionym w tej pracy modelu nie każde spotkanie agentów zakończy się utworzeniem więzi pomiędzy nimi. Jednakże, nawet wobec niewielkiej szansy pojawienia się nowej znajomości bez wprowadzenia mechanizmów osłabiających siłę znajomości pomiędzy agentami aż do pełnego rozpadu relacji z czasem, wcześniej czy później doszłoby do sytuacji gdzie „każdy zna każdego” – co oczywiście jest nierealne dla nieizolowanych społeczności. Dlatego też w modelu wprowadzono nie tylko mechanizmy odpowiadające za powstawanie

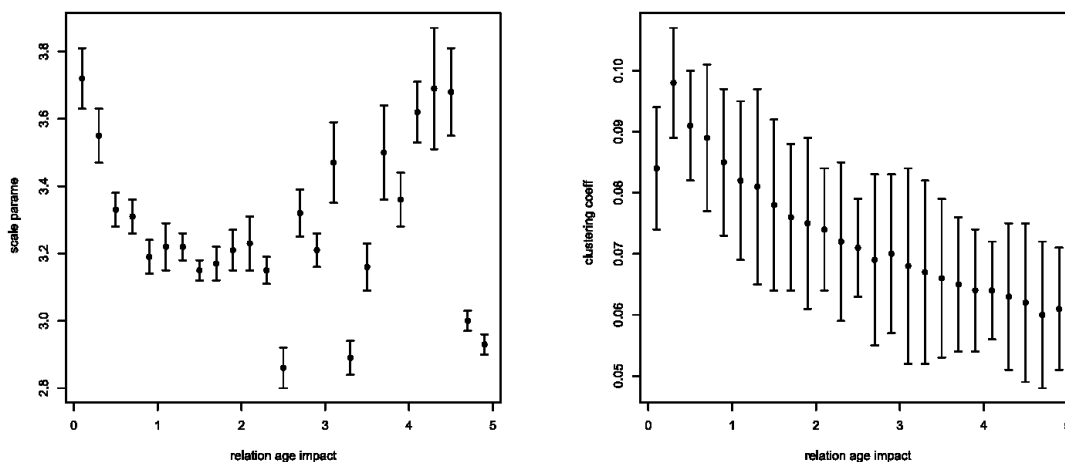
nowych czy podtrzymywanie już istniejących więzi, ale również takie, które odpowiadają procesom słabnięcia tychże relacji.

Poniżej pokazano dwa parametry wykorzystywane w modelu do sterowania procesami rozpadu. Pierwszym parametrem jest s_{decay} sterujący słabnięciem siły więzi pomiędzy agentami z każdym cyklem symulacji, w którym nie doszło do spotkania agentów. Drugim jest parametr G_{time} odpowiedzialny za spowalnianie wraz z wiekiem wpływu spotkań lub ich braku na siłę więzi między agentami, (czyli im starsza znajomość tym mniej zmienia się z czasem).



Rys. 5-8 Wpływ parametru odpowiadającego za słabnięcie siły znajomości s_{decay} (relation weakening) na wartość dopasowanego wykładnika potęgowego α i średnią wartość współczynnika gronowania $\langle C \rangle$ wyników sieci.

O ile parametr s_{decay} bezpośrednio wpływa na szybkość rozpadu relacji i zgodnie z oczekiwaniem jego zwiększanie prowadzi do krótszego „czasu życia” znajomości skutkując spadkiem średniego współczynnika gronowania tworzonych sieci, o tyle kolejny parametr G_{time} głównie działa „stabilizacyjnie” hamując wzrost liczby nowych relacji, ale nie bezpośrednio tylko, jako efekt, większej trwałości starszych relacji i „brak czasu” dla agenta na nawiązanie nowej trwałej więzi.



Rys. 5-9 Wpływ parametru odpowiadającego za słabnięcie wpływu dynamiki spotkań wraz z wiekiem relacji G_{time} (relation age impact) na wartość dopasowanego wykładnika potęgowego α i średnią wartość współczynnika gronowania $\langle C \rangle$ wynikowych sieci.

5.3. Podsumowanie

W rozdziale tym przebadano wpływ zmiany wartości parametrów modelu na postać wynikowych sieci społecznościowych. Sieci te scharakteryzowano przez dopasowany do rozkładu liczby relacji (w tworzonych sieciach) współczynnik skalowania α oraz odpowiadającej każdej z powstałych sieci średni współczynnik gronowania $\langle C \rangle$.

Pokazano, iż zaproponowany model pozwala na generację szerokiej gamy sztucznych sieci społecznościowych.

Jednym z kluczowych wniosków jest to, że aby uzyskać w prezentowanym modelu sieć o właściwościach podobnych jak obserwowane w rzeczywistych sieciach społecznościowych agenci muszą dysponować nie tylko okazją do spotkania i nawiązania relacji, ale również możliwością dobrowolnego wyboru innych agentów, z którymi się spotykają, biorąc pod uwagę ich wspólne zainteresowania. Liczba takich

dobrowolnych spotkań wydaje się być jednym z głównych czynników wpływających na parametry powstałej sieci.

Zbadano, przy jakich wartościach parametrów tworzona w modelu sieć społecznościowa posiada własności najbardziej zbliżone do obserwowanych w realnym świecie. Wyniki te są następnie używane do tworzenia społecznej sieci agentów kupujących w prezentowanym w dalszej części pracy modelu kształtowania cen.

6. Model współdziałania procesów decyzyjnych kupujących i sprzedawców

W rozdziale tym przedstawiony zostanie opis wieloagentowego, dynamicznego modelu kupna i sprzedaży, nazwanego modelem Piskor-Ignatowicza – Zachary. Model ten reprezentuje współdziałanie procesów decyzyjnych populacji kupujących i populacji sprzedawców.

Model ten stanowi podstawę programu symulującego wykorzystywanego w części eksperymentalnej.

Do modelowania dynamiki procesów decyzyjnych kupujących i sprzedawców wykorzystywany jest model sztucznej sieci społecznej przedstawiony w rozdziale 4.

Motywacje do stworzenia modelu wieloagentowego symulującego proces kształtowania się cen produktu są zarówno natury teoretycznej jak i praktycznej. Klasyczne koncepcje czy teorie ekonomiczne zakładają racjonalność rynku, równowagowy charakter procesów, jednolitość zarówno rynku jak i kupujących i sprzedających. W konsekwencji nie są w stanie ani opisać ani wyjaśnić powstania rozpiętości cenowych

Tymczasem w przypadku kupujących, obserwuje się całe spektra zachowań, od w pełni racjonalnych po całkowicie irracjonalne, takie jak gry hazardowe, zakupy usług astrologicznych czy robienie zakupów pod wpływem impulsu, skutkujące nabyciem rzeczy zupełnie kupującemu niepotrzebnych.

6.1. Ogólna koncepcja

Zanim szczegółowo przedstawimy model, podamy tutaj w sposób bardzo ogólny zależności, w oparciu o które podejmują swoje decyzje (w dowolnym cyklu) sprzedawcy i kupujący oraz wzajemne powiązania w obu tych procesach decyzyjnych. Każdy sprzedawca stara się zmaksymalizować całkowity swój zysk uzyskany w wyniku bilansu dochodów i kosztów w pewnym, ustalonym okresie bilansowania, zwanym okresem rozliczeniowym. Każdy kupujący stara się dokonać takiego zakupu, który w jego odczuciu, jest najbardziej atrakcyjnym. Zatem podstawą modelu jest

definicja dwóch zależności: zysku sprzedawcy oraz tego, co kupujący uważa za atrakcyjność zakupu.

Ogólne założenia, przy których opracowano model są następujące

- a) W modelu zaimplementowano dwie populacje agentów: sprzedawców i kupujących, oddziaływających wzajemnie na siebie.
- b) Wielkość przestrzeni jak i ilość znajdujących się w niej agentów danego typu zadawane są jako parametry symulacji. Agenci rozmieszczani są losowo zgodnie z zadaną funkcją gęstości prawdopodobieństwa. W dalszych rozważaniach, o ile nie zaznaczono inaczej, jest to rozkład jednorodny.
- c) Model uwzględnia zarówno sprzedaż w sklepach zwykłych jak i sprzedaż internetową. Ten ostatni fakt ma wpływ na koszty sprzedaży, które istotnie mogą zależeć od ilości transakcji.
- d) Dla uproszczenia w modelu zakłada się, że sprzedający handlują jednym towarem, jednakże sprzedawcy mogą handlować towarem o różnej jakości oraz różnić się też jakością sprzedaży (związanych z nią usług). Uproszczenie takie wprowadzono ze względu na eksperymenty w których badano rozkład cen towarów. Łatwo zauważyć, że jest to relatywnie słabe założenie, którego opuszczenie nie zmieni koncepcji modelowania.
- e) Każdy agent sprzedawca podejmuje decyzje dotyczące zmiany cen oraz jakości swoich usług, starając zmaksymalizować swój zysk zdefiniowany dalej za pomocą wzoru (6.1).
- f) Każdy agent kupujący może jednorazowo dokonać zakupu tylko jednej sztuki towaru. Decyzję o kupnie podejmuje w oparciu o postrzeganą przez niego (indywidualną) atrakcyjność A_{tr} potencjalnego zakupu, zdefiniowaną dalej za pomocą wzoru (6.2).
- g) Model uwzględnia, oprócz cen kupowanych towarów, także dodatkowe, tzw. niejawne koszty zakupu, z których kupujący nie zawsze zdają sobie sprawę.
- h) W modelu przewidziana jest możliwość bankructwa sprzedawców, którzy nie osiągnęli wystarczających dochodów oraz możliwość pojawienia się nowych sprzedawców w przypadku koniunktury.

- i) Model uwzględnia też proces pozyskiwania (i zapominania) informacji o sprzedawcach od znajomych
- j) Czas w modelu liczony jest za pomocą tzw. cykli oznaczających równe przedziały czasu. Numer cyklu, oznaczany symbolem t , $t = 1, 2, \dots, T_{sym}$ jest równoznaczny z momentem zakończenia cyklu. T_{sym} oznacza całkowitą ilość cykli.
- k) Wprowadzono również okresy rozliczeniowe sprzedawców składające się ze stałej ilości cykli. Numer okresu rozliczeniowego oznaczany jest jako k , $k = 1, 2, \dots, k_{max}$.

Dalsze, już bardziej szczegółowe, założenia dotyczące modelowanej rzeczywistości, będą przedstawione przy opisie modelu zachowania kupującego oraz zachowania sprzedawcy.

Zysk sprzedawcy P uzyskany w okresie rozliczeniowym, zdefiniowany jest w następujący sposób:

$$P = n(S - C_p(n)) - C_s(n) - C_c \quad (6.1)$$

gdzie:

S - cena transakcyjna towaru,

C_c - stałe koszty (np. wynajmu lub utrzymania pomieszczeń itd.),

n - ilość transakcji w okresie rozliczeniowym,

$C_s(n)$ - koszt sprzedaży n sztuk (jednostek) produktu nieliniowo zależny od n (tj. zatrudnienia pracowników, magazynowania, pakowania itd.),

$C_p(n)$ - koszt uzyskania sprzedawanego przedmiotu przy sprzedaży n sztuk.

Jeżeli zysk jest ujemny, to mamy do czynienia ze stratą.

Każdy agent kupujący stara się dokonać zakupu tak, by zmaksymalizować ocenianą przez niego atrakcyjność zakupu A_{tr} zdefiniowaną następująco:

$$A_{tr} = \alpha Q_{serv} + \beta Q_{obj} - \gamma S - \delta C_{hid} \quad (6.2)$$

gdzie:

Q_{serv} - jakość usług sprzedawcy, (uwzględniająca czas dostarczenia, sposoby uwzględniania reklamacji itp.)

Q_{obj} - jakość towaru; jest ona własnością każdego przedmiotu i uwzględnia fizyczną jakość (wartość marki itd.)

C_{hid} - suma kosztów niejawnych (ukrytych/hidden) uzyskania towaru przez kupującego (np. czas potrzebny na zgromadzenie informacji przed podjęciem decyzji o konkretnym zakupie),

$\alpha, \beta, \gamma, \delta$ - współczynniki określający indywidualne preferencje kupującego odpowiednio: α – względem jakości usług sprzedawcy,

β – względem jakości towaru,

γ – względem ceny towaru,

δ – względem kosztów niejawnych.

Jak widać, w obu zależnościach (6.1) i (6.2) występuje cena towaru, jednakże obok niej występuje jeszcze kilka dodatkowych wielkości. Tak więc z punktu widzenia kupującego, wiedza o konkretnym sprzedawcy reprezentowana jest przez czwórkę wielkości: $S, Q_{obj}, Q_{serv}, C_{hid}$.

W celu maksymalizacji zysku, każdy ze sprzedawców może modyfikować kilka z wielkości występujących we wzorze (6.1). Po pierwsze, może zmienić cenę, co prowadzi do zmiany atrakcyjności zakupu A_{tr} dla wszystkich kupujących, ale równocześnie wpływa na zysk P tego sprzedającego. To, czy zysk wzrośnie czy zmaleje, zależy od ilości transakcji jaką spowoduje zmiana ceny.

Po drugie, może zmienić jakość usług, co w modelu wyraża się zmianą składowej Q_{serv} w równaniu (6.2) i może wpływać na składową $C_s(n)$. Jednym ze sposobów zmiany jakości usług jest zmiana lokalizacji sklepu, która ma wpływ na zmianę kosztu stałego C_c w równaniu na zysk, jak i również wpływać może na koszty niejawne C_{hid} . Jeżeli zmiana lokalizacji realizowana jest jako zamknięcie nierentownej inwestycji ze względu na duże koszty, to znajduje to swoje odzwierciedlenie w modelu jako bankructwo.

6.2. Model agenta sprzedawcy

Zbiór agentów sprzedawców jest oznaczany jako V (vendors). W modelu wprowadzono tzw. okres rozliczeniowy (bilansowy) sprzedawcy. Ilość cykli

składających się na okres rozliczeniowy n_{roz} jest stały dla wszystkich sprzedawców i stanowi parametr modelu.

Suma wszystkich cykli we wszystkich okresach rozliczeniowych daje całkowitą ilość cykli T_{sym} . Zbiór (populacja) agentów sprzedawców jest stały w całym okresie rozliczeniowym, może natomiast różnić się w kolejnych okresach rozliczeniowych w wyniku bankructw oraz pojawiania się nowych sprzedawców.

Wszystkie wielkości niezmiennie w okresach rozliczeniowych i obliczane (modyfikowane) na koniec $k-1$ -szego okresu stają się obowiązujące od pierwszego cyklu następnego, k -tego okresu i są przypisane do całego tego okresu.

Po zakończeniu każdego okresu rozliczeniowego, sprzedawca podejmuje też decyzje dotyczące :

- ustalenia ceny na następny okres rozliczeniowy,
- ustalenia swojej jakości,
- ewentualnego bankructwa.

6.2.1. Obliczanie zysku sprzedawcy

Celem każdego sprzedawcy jest maksymalizacja zysku P . Zysk każdego sprzedawcy liczony jest oddzielnie na koniec każdego okresu rozliczeniowego. Cena transakcyjna towaru danego sprzedawcy jest stała w całym okresie rozliczeniowym, może natomiast zmieniać się w kolejnych okresach rozliczeniowych (na początku), w wyniku decyzji sprzedawcy.

Wprowadźmy bardziej szczegółowe oznaczenia:

k – numer okresu rozliczeniowego, $k = 1, 2, \dots, k_{max}$,

n_{roz} – ilość cykli składających się na okres rozliczeniowy,

$V^k = \{v_i^k, i = 1, 2, \dots, |V^k|\}$ - zbiór agentów sprzedawców w k -tym okresie rozliczeniowym,

$S^k(v_i^k)$ - cena jednostkowa towaru i -tego sprzedawcy w k -tym okresie rozliczeniowym,

$n^k(v_i^k)$ - ilość transakcji i -tego sprzedawcy v_i^k w k -tym okresie rozliczeniowym.

Jeśli numer sprzedawcy, bądź numer okresu rozliczeniowego nie będzie istotny, (innymi słowy jeśli zależności będą obowiązywać dla każdego sprzedawcy lub

każdego okresu rozliczeniowego), indeks k oraz numer agenta i będzie w powyższych oznaczeniach opuszczany.

Obliczając swój zysk (stratę) w dowolnym okresie rozliczeniowym k , sprzedawca uwzględnia:

- a) cenę jednostkową S^k towaru obowiązującą w okresie rozliczeniowym, tzw. cenę transakcyjną
- b) jednostkowy koszt uzyskania sprzedawanego towaru $C_p^k(n^k)$, koszt ten zależy od ilości transakcji n^k w danym okresie rozliczeniowym i zależność ta może być nieliniowa i nieciągła,
- c) stałe koszty C_c^k , niezależne od ilości transakcji (np. wynajem lub utrzymanie sklepu)
- d) koszty sprzedaży $C_s^k(n^k)$, nieliniowo zależne od ilości transakcji n^k obejmujące zatrudnienie pracowników, magazynowania, pakowania i innej obsługi sprzedaży internetowych itd.

Dla agenta sprzedawcy w k -tym okresie rozliczeniowym definiujemy funkcję zysku (odpowiednik wzoru (6.1)):

$$\forall v_i^k \in V^k \quad P^k(v_i^k) = n^k(S^k - C_p^k(n^k)) - C_s^k(n^k) - C_c^k \quad (6.3)$$

6.2.2. Algorytm decyzyjny zmieniający cenę

W każdym okresie rozliczeniowym u danego agenta sprzedawcy zostaje dokonanych n^k zakupów przez kupujących.

Sprzedawca, aby zmaksymalizować zysk, podejmuje decyzje o ewentualnej zmianie ceny obowiązującej w nowym okresie rozliczeniowym. W tym celu zapamiętuje i wykorzystuje historię zysków i zmian cen z ostatniego okresu rozliczeniowego. Dla modelu agenta sprzedawcy można badać różne algorytmy zmiany cen. Poniżej przedstawiamy podstawowy wariant.

Kierunek zmiany ceny jest określony przez efekt wcześniejszej decyzji dotyczącej ceny. Jeżeli zmiana ceny obowiązującej w poprzednim $k - 1$ okresie rozliczeniowym (wzrost lub zmniejszenie) ΔS^{k-1} wpłynęła na wzrost zysku w tym okresie ΔP^{k-1} , to kierunek zmiany ceny (wzrost lub zmniejszenie) będzie kontynuowany dla nowego k -tego okresu. Jeśli decyzja dotycząca poprzedniego okresu rozliczeniowego wpłynęła na zmniejszenie zysku, to kierunek zmiany ceny zostanie zmieniony na przeciwny. Wartość, o jaką maksymalnie może się cena

zmienić jest ograniczona. Dopuszczamy zmianę do 20 % ceny w poprzednim okresie rozliczeniowym. Bezwzględna wartość zmiany ceny, w ramach tego ograniczenia, odbywa się w pewnym zakresie losowo za pomocą funkcji stochastycznej $|\Delta s^k| = rand_s(s^{k-1})$.

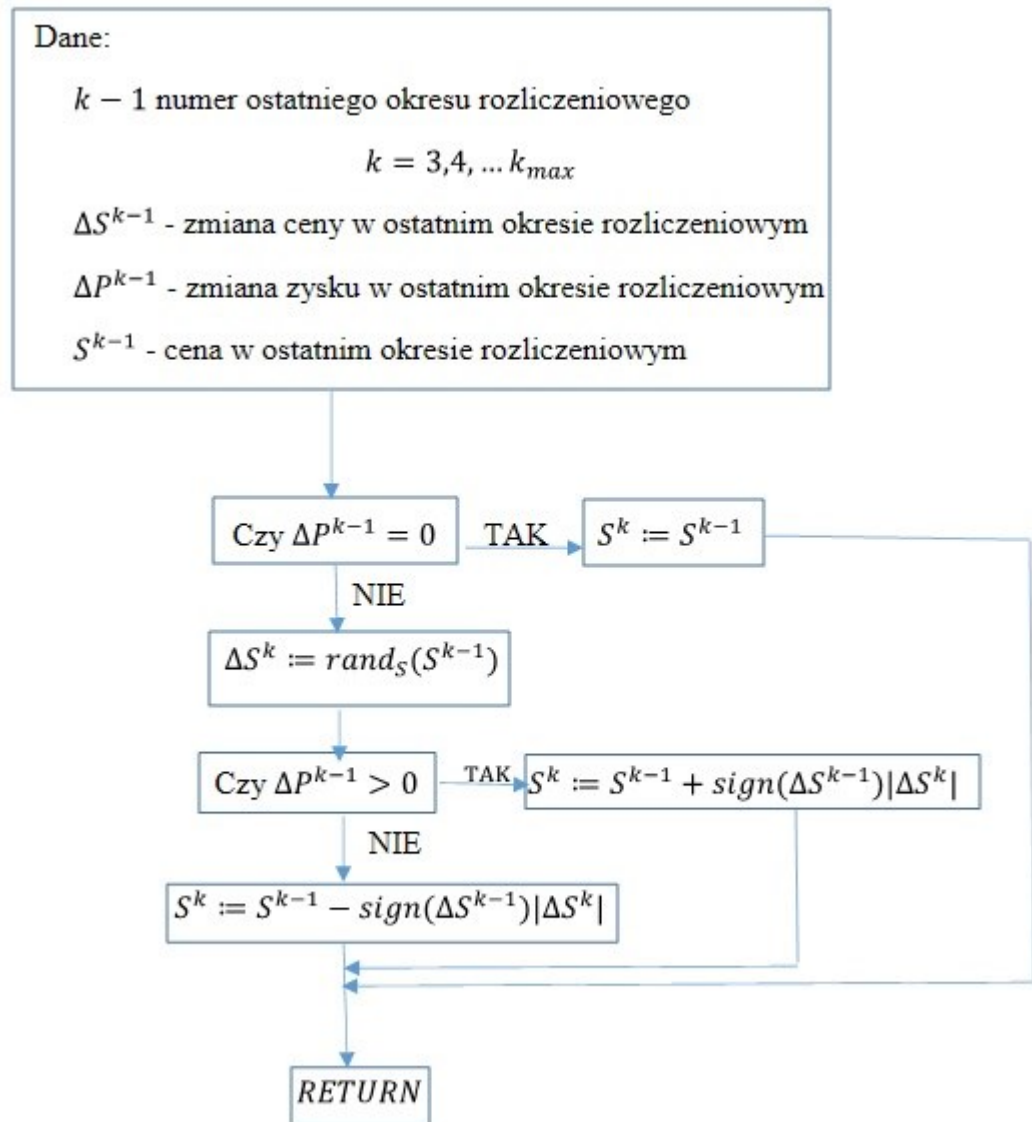
Wynikiem powyższego algorytmu jest cena S^k obowiązująca w bieżącym k -tym okresie rozliczeniowym. Niech:

$$\Delta S^{k-1} = S^{k-1} - S^{k-2}, \quad (6.4)$$

$$\Delta P^{k-1} = P^{k-1} - P^{k-2}. \quad (6.5)$$

Dla poniższego algorytmu wprowadzamy warunki początkowe, pozwalające ustalić zysk w pierwszych dwóch okresach dla wszystkich agentów sprzedawców:

S^1 – cena obowiązująca w okresie $k = 1$ oraz ΔS^2 - zmiana ceny w okresie drugim (dodatnia lub ujemna, ustalana losowo). Zgodnie ze wzorem (6.3) obliczane są zyski agentów w pierwszym i drugim okresie: P^1 oraz P^2 . Poniższy algorytm obowiązuje dla okresów rozliczeniowych $k = 3, 4, \dots, k_{max}$.



Rys 6.1. Schemat algorytmu wyznaczania nowej ceny na k -ty okres rozliczeniowy

6.2.3. Decyzja o bankructwie i pojawieniu się nowych agentów sprzedawców

W chwili początkowej (pierwszym cyklu pierwszego okresu rozliczeniowego) ilość agentów sprzedawców wynosi $|V^1|$ i jest parametrem modelu.

Zysk w okresie rozliczeniowym obliczany jest wzorem (6.3), gdzie n^k stanowi sumę transakcji z poszczególnych cykli składających się na ten okres.

Każdy sprzedawca dysponuje kapitałem, który zmienia się w poszczególnych okresach rozliczeniowych i jest uaktualniany na koniec tego okresu. Kapitał sprzedawcy v_i^k w k -tym okresie rozliczeniowym oznaczamy jako $Cap^k(v_i^k)$.

Jeśli suma zysków sprzedawcy w tym okresie jest mniejsza niż tak zwana dopuszczalna wartość długu sprzedawców $Ddop$, następuje jego bankructwo. W modelu założono, że każdy sprzedawca posiada taki sam kapitał startowy, zwany również zakładowym, oznaczany jako $Cap_0 = Cap^1(v_i^1) \quad \forall i \in 1, \dots, |V^1|$. Za dopuszczalną wartość długu $Ddop$ przyjmujemy 20 procent wartości kapitału zakładowego:

$$Ddop = -20\% Cap_0. \quad (6.6)$$

W zależności od zysków i strat poszczególnych sprzedawców w rozpatrywanym okresie rozliczeniowym mogą nastąpić zmiany ilości agentów sprzedawców. Niektórzy sprzedawcy mogą zbankrutować i zniknąć z runku jak i mogą pojawić się nowi agenci.

Nowi agenci sprzedawcy mogą pojawić się w przypadku koniunktury. W celu stwierdzenia koniunktury wprowadza się progową wartość sumarycznego zysku sprzedawców P_{tresh} .

Algorytm znikania i pojawiania się nowych agentów sprzedawców

Zmiany ilości agentów (znikanie lub pojawianie się nowych) następuje w następujących sytuacjach:

- a) agent v_i znika jeżeli bankrutuje, czyli jego zysk w pewnym okresie rozliczeniowym jest ujemny i przekracza dopuszczalną wartość długu czyli:

$$P^k(v_i^k) < Ddop \quad (6.7)$$

- b) nowi agenci sprzedawcy mogą pojawiać się w przypadku koniunktury, czyli jeśli sumaryczny zysk wszystkich agentów sprzedawców w k -tym okresie rozliczeniowym

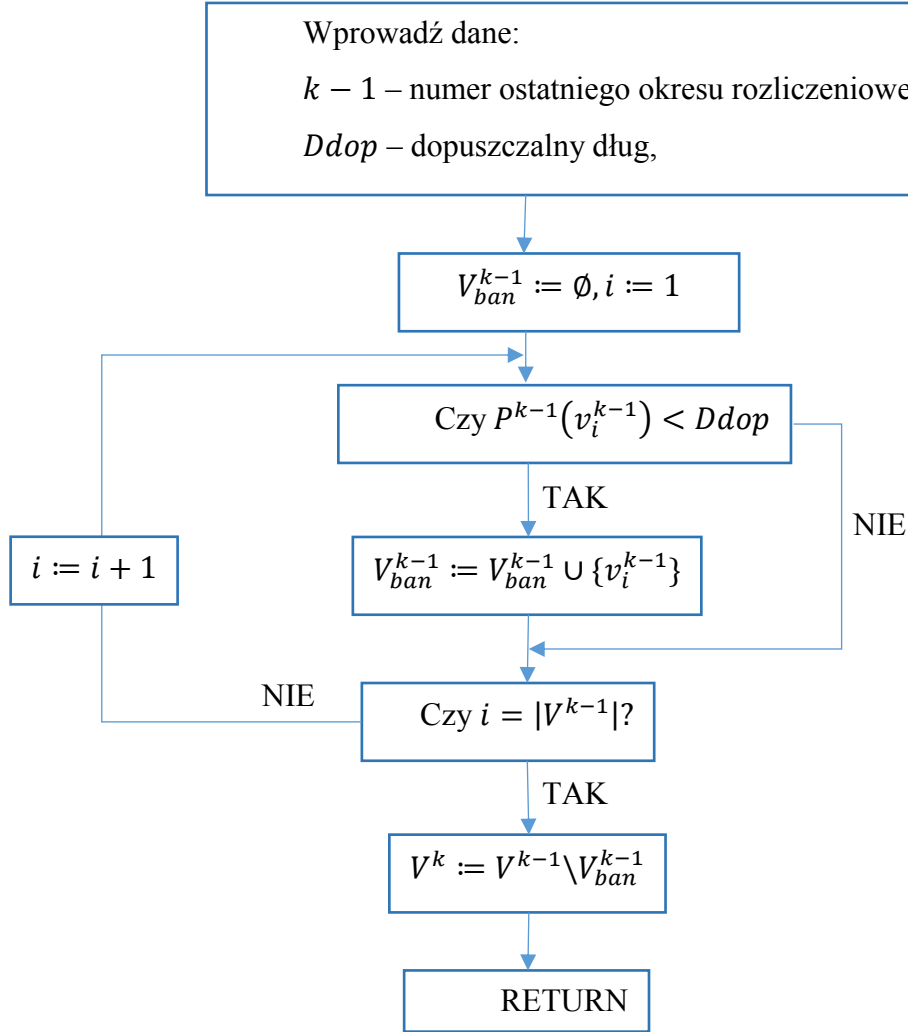
$$P_{sum}^k = \sum_{i=1}^{|V^k|} P_i^k(v_i^k) \quad (6.8)$$

- c) pomniejszony o sumę ich kapitałów zakładowych jest większy niż zadana wartość P_{tresh} , formalnie

$$P_{sum}^k - |V^k| Cap_0 \geq P_{tresh} \quad (6.9)$$

- d) Ilość nowych agentów sprzedawców zależy od wysokości tej nadwyżki ponad P_{tresh} .

Algorytmy dla obu powyższych przypadków zostały przedstawione na kolejnych dwóch rysunkach. Algorytm wyznaczający nowy zbiór V^k agentów sprzedaży zmodyfikowany w wyniku bankructw. Oznaczmy V_{ban}^k – podzbiór sprzedawców, którzy bankrutują w k -tym okresie rozliczeniowym.



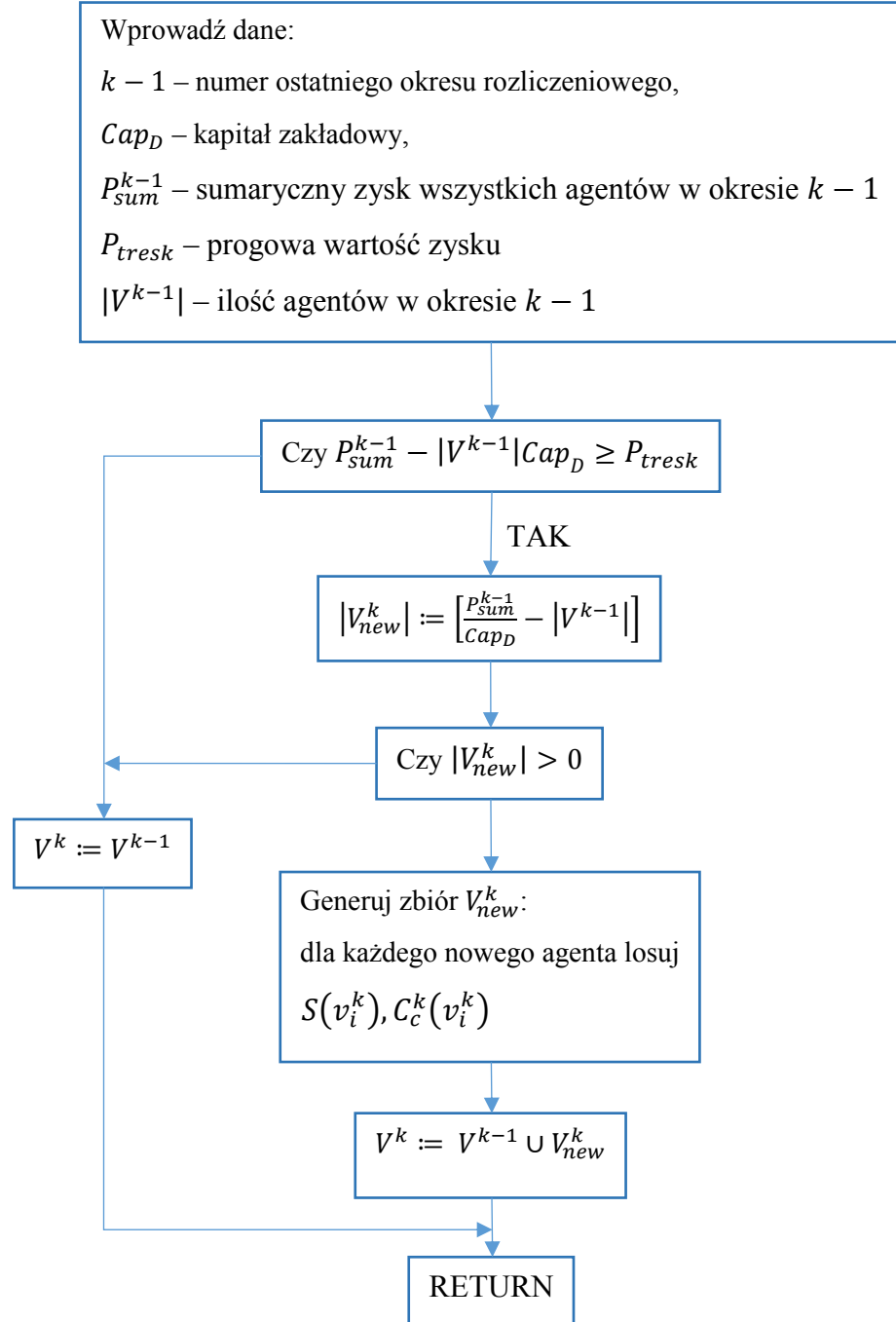
Rys. 6.2. Algorytm usuwania agentów sprzedawców (bankructwo)

Wynikiem powyższego algorytmu jest zbiór pozostających agentów sprzedawców, czyli agentów którzy nie zbankrutowali i zostają na następny okres rozliczeniowy.

Algorytm pojawiania się nowych agentów sprzedawców

Wynikiem tego algorytmu jest sprawdzenie czy pojawią się w k -tym okresie rozliczeniowym nowi agenci, a jeżeli tak, to wygenerowanie podzbioru nowych agentów i ostatecznie zbioru agentów sprzedawców w k -tym okresie rozliczeniowym.

Oznaczmy V_{new}^k - zbiór nowych agentów sprzedawców, którzy pojawiają się na początku (w pierwszym cyklu) k -tego okresu rozliczeniowego. Algorytm pojawiania się nowych agentów sprzedawców ma postać:



Rys. 6.3. Algorytm pojawiania się nowych agentów sprzedawców.

Ostatecznie nowy zbiór agentów w k -tym okresie rozliczeniowym jest sumą mnogościową zbiorów wyznaczonych przez oba algorytmy.

6.3. Modelowanie procesów decyzyjnych Agentów Kupujących

W modelu zakłada się, że zbiór (populacja) agentów kupujących jest stały, jakkolwiek stan posiadania kupujących, ich wiedza konieczna do podjęcia decyzji zakupowych oraz relacje z innymi agentami kupującymi zmieniają się dynamicznie.

Zbiór agentów kupujących (buyers) będzie oznaczany jako B . Agenci kupujący, poprzez kontakty z innymi agentami kupującymi tworzą sieć społeczną, (której model przedstawiono w rozdziale 4) i dzięki temu uzyskują informacje i gromadzą wiedzę o agentach sprzedających. Dlatego też model działania agentów kupujących zawiera w sobie model sztucznej sieci społecznej.

W modelu agenta kupującego, oprócz wielkości charakteryzujących agenta $a \in A$ podanych w rozdziale 4, uwzględnia się szereg dodatkowych, charakteryzujących go wielkości, mających wpływ na indywidualne preferencje przy dokonywaniu zakupów.

6.3.1. Model agenta kupującego

Każdego agenta kupującego $b \in B$ charakteryzują następujące wielkości, niezmiennie w całym przebiegu symulacji: $L(b)$ (jak poprzednio) położenie kupującego b określone współrzędnymi $(x(b), y(b))$, $E(b)$ - zestaw wielkości charakteryzujących osobowość agenta kupującego, będący rozszerzeniem osobowości zdefiniowanych w rozdz. 4, bowiem w modelu osobowości agenta kupującego $E(b)$, oprócz przyjaźniowości i zainteresowań pozwalających nawiązywać i utrzymywać kontakty ze znajomymi uwzględnia się dodatkowo takie cechy jak preferencje wysokiej jakości towaru i usług związanych z zakupem, preferencje niskiej ceny, niechęć do ponoszenia dodatkowych, niejawnych kosztów zakupu i skłonność do szybkiego nudzenia się posiadanymi przedmiotami.

Stopnie natężenia poszczególnych cech charakteryzowane są za pomocą współczynników: $\alpha, \beta, \gamma, \delta, w_{bor}$ określających indywidualne preferencje kupującego odpowiednio:

α – względem jakości usług sprzedawcy, $\alpha \in [0,1]$

β – względem jakości towaru, $\beta \in [0,1]$

γ – względem ceny towaru, $\gamma \in [0,1]$

δ – względem kosztów niejawnych, $\delta \in [0,1]$

w_{bor} - znudzenia się posiadanym przedmiotem.

Zatem osobowość agenta kupującego definiujemy, jako:

$$E(b) = (f(b), (w_1(b), w_2(b), \dots, w_{in}(b)), \alpha(b), \beta(b), \gamma(b), \delta(b), w_{bor}(b)) \quad (6.10)$$

gdzie: $f(b)$ – współczynnik „przyjacielskości” agenta b zaś $w_i(b)$, $i = 1, 2, \dots, in$ – współczynniki zainteresowania agenta b , i -tym obszarem zainteresowań, zdefiniowane w rozdziale 4.

Stan agenta kupującego w cyklu t zdefiniowany jest w modelu za pomocą trzech wielkości:

- relacji znajomości z innymi agentami $R_{known}^t(b)$ oraz jej charakterystyki, (zdefiniowanych w rozdziale 4).
- wiedzy $GK^t(b)$ jaką posiada agent kupujący b w cyklu t ; wiedza ta dotyczy znajomości sprzedawców (podzbiór znanych sprzedawców $V_{known}^t(b)$ oraz wielkości (parametrów) ich charakteryzujących,
- zbioru $O^t(b)$ przedmiotów posiadanych przez agenta kupującego b w cyklu t ; wraz z informacją dotyczącą stopnia w jakim są jeszcze są dobre (nie zużyte).

Zbiór B obejmuje całą populację potencjalnych agentów kupujących. W każdym z cykli kupujący może dokonać zakupu jednej jednostki towaru od dowolnego agenta sprzedawcy lub też zaniechać zakupu. Stąd wprowadzamy oznaczenie podzbioru agentów dokonujących zakupu w cyklu t jako B_{zak}^t .

Kupujący dokonuje decyzji zakupu biorąc pod uwagę:

- stan posiadania, czyli zbiór posiadanych w danym cyklu przedmiotów; zbiór ten może się zmieniać w poszczególnych cyklach na skutek zużycia istniejących przedmiotów i zakupu nowych,
- tzw. stopień znudzenia posiadanymi przedmiotami, zależny od osobowości kupującego,
- znajomość sprzedawców i posiadane, charakteryzujące ich informacje: S , Q_{obj} , Q_{serv} , C_{hid} .

Opierając się na tych informacjach i uwzględniając swoje preferencje osobowościowe, kupujący wyznacza i realizuje w każdym cyklu swoją decyzję dotyczącą zakupów.

W modelu zachowania agenta kupującego, w każdym cyklu uwzględniane są następujące etapy, związane z realizacją zakupów:

- **etap 1:** podjęcie decyzji o potrzebie zakupu
- **etap 2:** wybór sprzedawcy, u którego zrealizowany zostanie zakup
- **etap 3:** dokonanie zakupu, uaktualnienie zbioru posiadanych przedmiotów i ich stanu

6.3.2. Etap 1 - podjęcie decyzji o potrzebie zakupu

W wyniku realizacji omawianego etapu, zostaje określony podzbiór agentów B_{zak}^t dokonujących zakupy w cyklu t .

Każdy agent kupujący posiada zbiór przedmiotów, których stan pogarsza się z czasem ze względu na zużywanie się (psucie). Zbiór przedmiotów składa się z kilku przedmiotów tego samego rodzaju (analizowanego w modelu), różniących się stopniem zużycia, przyjmuje się bowiem w modelu, że posiadane przedmioty podlegają psuciu losowo. Stopień „nie zużycia” modelowany jest jako zmieniający się zgodnie z rozkładem eksponentialnym $e^{\lambda\tau}$ gdzie $\lambda < 0$, zaś τ jest czasem jaki upłynął od momentu zakupu (ilość cykli)..

Uświadomienie sobie potrzeby zakupu i podjęcie decyzji o tej potrzebie zależy od dwóch czynników: stopnia zużycia posiadanych przedmiotów oraz indywidualnego stopnia znudzenia agenta posiadanymi przedmiotami

Każdy posiadany przedmiot $o \in O(b)$ jest scharakteryzowany przez:

- czas zakupu, oznaczany jako numer cyklu w którym został zakupiony - $t_{zak}(o)$
- stopień „nie zużycia” (not worn out), zmieniający się losowo w czasie (malejący) i oznaczany jako $w^\tau(o)$, gdzie $\tau = t - t_{zak}(o)$,

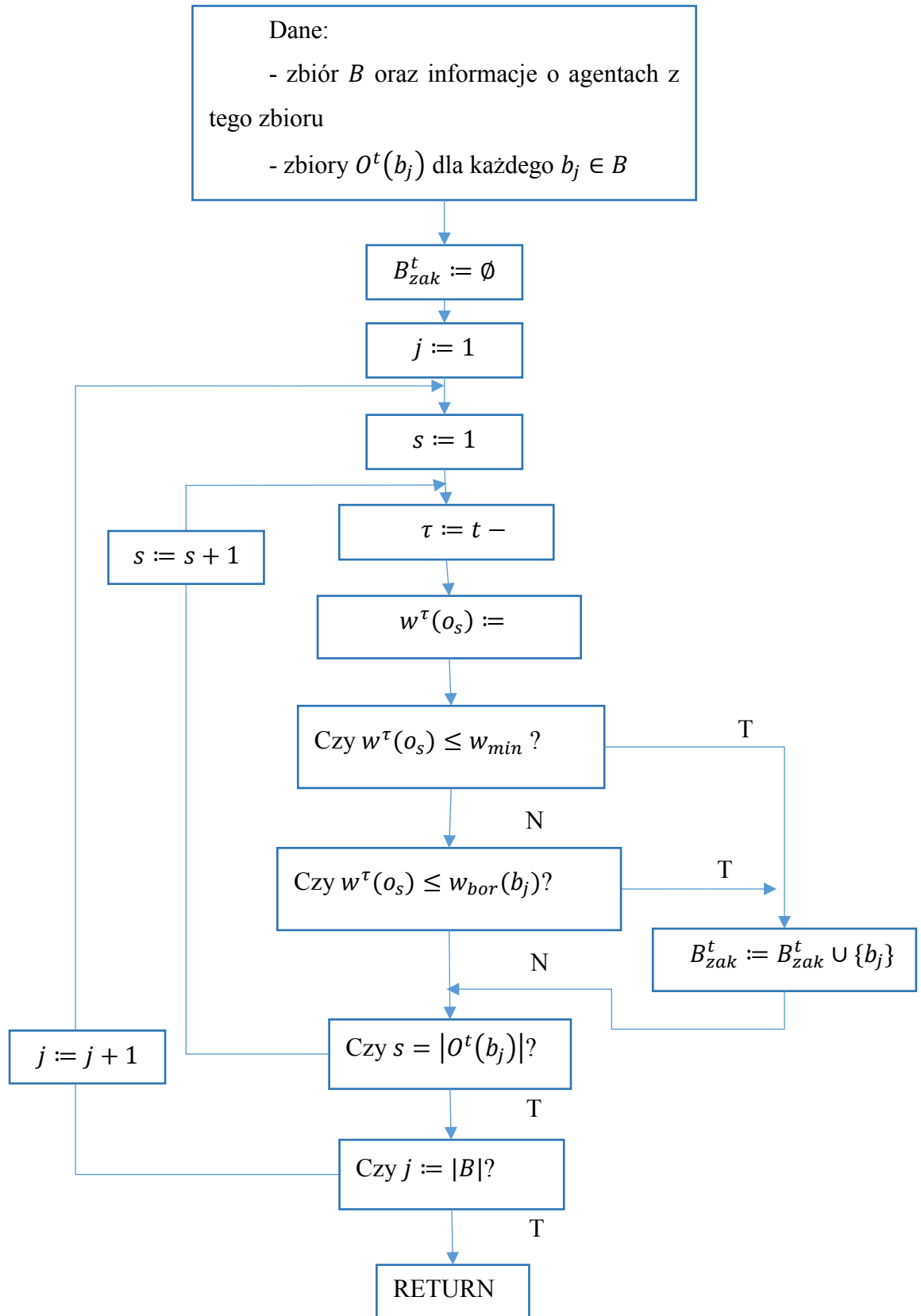
W chwili zakupu (w cyklu o numerze t_{zak}) stopień ten ma wartość równą 1, zatem $w^0(o) = 1$, zaś w kolejnych cyklach, dla $\tau > 0$ maleje przyjmując wartość:

$$w^\tau(o) = rand_w(\tau), \quad (6.11)$$

gdzie $rand_w(\tau)$, jest funkcją losową o rozkładzie $e^{\lambda\tau}$, λ jest stałą, $\lambda < 0$.

Minimalna, akceptowalna przez każdego agenta kupującego wartość stopnia „nie zużycia” w_{min} jest stałą i stanowi parametr modelu. Jeśli zatem stopień „nie zużycia” osiągnie wartość mniejszą lub równą od w_{min} , to każdy agent kupujący podejmuje decyzję o potrzebie zakupu. Jak powiedziano, decyzja o potrzebie zakupu zależy też od indywidualnej cechy osobowości agenta, a dokładniej od wartości tzw. współczynnika znudzenia $w_{bor}(b)$, który jest jednym z parametrów charakteryzujących osobowość agenta kupującego. Jeśli $w_{bor}(b) > w_{min}$, to agent podejmuje decyzję o dokonaniu zakupu, mimo, że przedmiot nie jest całkowicie zużyty.

Algorytm poniższy wyznacza zbiór B_{zak}^t (dla początku cyklu t) agentów, którzy dokonują zakupów w cyklu t na podstawie informacji z końca poprzedniego cyklu $t - 1$, s oznacza indeks przedmiotu w zbiorze $O^t(b_j)$.



Rys. 6.4 Wyznaczanie zbioru agentów B_{zak}^t , agentów realizujących zakupy w cyklu t .

6.3.3. Etap 2 - wybór sprzedawcy

Agent kupujący dokonuje wyboru agenta sprzedawcy biorąc pod uwagę :

- znajomość sprzedawców i posiadane o nich informacje (wiedzę)
- swoje preferencje reprezentowane za pomocą wielkości składających się na tzw. osobowość agenta kupującego.

6.3.3.1. Wiedza agenta kupującego

W modelu zakłada się, że kupujący może dokonywać zakupu tylko u sprzedawców, o których istnieniu wie i pamięta. Co więcej, posiada o nich wiedzę wyspecyfikowaną poniżej. Agent kupujący b_j pozyskuje wiedzę o sprzedawcach od innych kupujących, którzy są jego znajomymi. Może też pozyskać ją bezpośrednio od sprzedawcy (reklama, internet). W modelu przewidziane są i zaimplementowane w oprogramowaniu symulacyjnym różne mechanizmy pozyskiwania i uaktualniania wiedzy o sprzedawcach. Nie będą tu jednak formalnie przedstawiane, ze względu na szczegółowość i długość opisu różnych alternatyw. Oznaczmy jako $V_{known}^t(b)$ zbiór sprzedawców, znanych i pamiętanych przez agenta b w cyklu t . Wiedza, jaką posiada agent b_j o sprzedawcy v_i w cyklu t (należącym do k -tego okresu rozliczeniowego sprzedawców) oznaczana jako $K^t(b_j, v_i)$, stanowi n -tkę (krotkę) specyfikującą wiedzę o istnieniu danego sprzedawcy (identyfikator sprzedawcy) i jego atrybutach takich jak - $S^k(v_i^k)$ - cena jednostkowa towaru i -tego sprzedawcy w k -tym okresie rozliczeniowym i odpowiednio, jakość towaru $Q_{obj}(v_i^k)$, jakość usług sprzedawcy $Q_{serv}(v_i^k)$, oraz koszty ukryte $C_{hid}(v_i^k)$. Wiedza agenta b_j w cyklu t o wszystkich sprzedawcach ze zbioru $V_{known}^t(b)$, przechowywana jest w uporządkowanym zbiorze krotek i oznaczanym jako $GK^t(b_j)$.

Wiedza ta, uzyskana podczas wcześniejszych spotkań agenta kupującego ze swoimi znajomymi, może być aktualna (jeśli jacyś sprzedawcy nie zbankrutowali i nie zmienili swoich cen oraz innych parametrów) bądź nieaktualna, a zmiany mogą być zarówno pozytywne dla kupujących jak i negatywne.

Zatem model uwzględnia opóźnienia w informacji o sprzedawcach. Pozyskiwanie wiedzy następuje podczas spotkań w modelu sztucznej sieci społecznej. Podczas spotkania dwóch agentów, w tym przypadku kupujących, z pewnym prawdopodobieństwem następuje przekazanie informacji o sprzedawcy. Równocześnie w modelu uwzględniane jest też zapominanie informacji związane z

dwoma mechanizmami zapominania.. Pierwszy polega na wprowadzeniu zróżnicowanej pojemności pamięci agentów kupujących (za pomocą wewnętrznego parametru modelu o losowo przydzielanych wartościach). Informacje o sprzedawcach szeregowane są zgodnie z kolejnością ich otrzymania. Po wypełnieniu pamięci usuwana jest najstarsza informacja.

Drugi mechanizm polega na losowym usunięciu informacji z pamięci. W tym celu w modelu wprowadzony został parametr cmf (customer memory factor). W symulacjach przeprowadzonych w ramach pracy parametr cmf został przyjęty jako 0,01. Czyli z prawdopodobieństwem 0, 01 każda informacja posiadana przez agenta kupującego może zostać zapomniana w danym cyklu.

Po otrzymaniu nowej informacji o sprzedawcy, może ona też zostać zapomniana z prawdopodobieństwem 0,01 w danym cyklu i (losowo przydzielany parametr agenta – customer broadcast through wynoszący 0,01)

6.3.3.2. Atrakcyjność zakupu

Każdy agent dokonujący zakupu w cyklu t należącym do k -tego okresu rozliczeniowego wybiera sprzedawcę, u którego chce dokonać zakupu, na podstawie indywidualnie przez niego szacowanej atrakcyjności zakupu. Formalnie, dla każdego $b_j \in B_{zak}^t$, dla każdego sprzedawcy $v_i \in GK^t(b_j)$, obliczana jest tzw. indywidualna atrakcyjność zakupu $A_{tr}^t(b_i, v_j)$ zdefiniowana wzorem:

$$A_{tr}^t(b_i, v_j) = \alpha_j Q_{serv}(v_i^k) + \beta_j Q_{obj}(v_i^k) + \gamma_j S^k(v_i^k) + \delta_j C_{hid}(v_i^k) \quad (6.12)$$

Zbiór $V_{known}^t(b)$ jest porządkowany zgodnie z malejącą wartością atrakcyjności. Wybierany jest pierwszy dostawca.

6.3.4. Etap dokonywania zakupu i uwzględnienie nieaktualnej wiedzy

Po wyborze sprzedawcy następuje realizacja zakupu. Jednakże wiedza kupującego może nie być aktualna, zwłaszcza gdy pozyskana została w cyklach należących do wcześniejszych okresów rozliczeniowych wybranego sprzedawcy. W modelu uwzględnia się fakt, że wybrany sprzedawca może już nie istnieć lub też cena lub jakość jego towaru są już zmienione (informacje zapisane w zbiorze $GK^t(b_j)$). W pierwszej sytuacji wybrany zostaje następny sprzedawca, zaś w drugiej sytuacji następuje ponowne obliczenie $A_{tr}^t(b, v)$ dla aktualnych wartości parametrów

sprzedawcy. Jeśli atrakcyjność wybranego sprzedawcy nie jest gorsza, to następuje realizacja zakupu u niego. W przeciwnym wypadku, zakup będzie realizowany u następnego z kolei sprzedawcy ze zbioru $V_{known}^t(b)$. W obu przypadkach zbiór $V_{known}^t(b)$ oraz wiedza $GK^t(b_j)$ zostają zaktualizowane.

Etap realizacji zakupu kończy wywołanie i zrealizowanie algorytmu, który dla każdego agenta $b \in B_{zak}^t$ zmienia **stan** posiadanych przez niego przedmiotów $O'(b)$ i ich charakterystyk. Zbiór $O'(b)$ jest powiększany o nowy przedmiot na skutek zakupu, zaś przedmiot o_s , dla którego $w^\tau(o_s) \leq w_{min}$ lub $w^\tau(o_s) \leq w_{bor}(b_j)$ zostaje usunięty. Ponadto, następuje zwiększenie τ i modyfikacja stopni „nie zużycia” pozostałych przedmiotów.

Po wykonaniu w rozważanym cyklu t , wszystkich algorytmów modelujących zachowanie agentów kupujących, wywoływane są algorytmy dotyczące agentów sprzedawców w celu wzajemnego powiązania procesów decyzyjnych obu typów agentów.

Po pierwsze realizowane są algorytmy zwiększające ilość transakcji i zysk tych agentów sprzedawców, którzy w tym cyklu zostali wybrani przez kupujących. Po drugie, jeśli cykl t jest równocześnie ostatnim cyklem któregośkolwiek okresu rozliczeniowego, następuje wywołanie i realizacja algorytmów obliczających ewentualne zmiany cen i pozostałych parametrów sprzedawców, modelowanie bankructw i pojawiania się nowych sprzedawców (algorytmy przedstawione w podrozdziale 6.2).

7. Analiza rozkładu cen produktu w sklepach internetowych

W niniejszym rozdziale przedstawiono wyniki analizy rozkładów cen dziesięciu losowo wybranych produktów z pośród tych, które są dostępne w więcej niż szesnastu różnych sklepach internetowych skatalogowanych na portalu Ceneo.

Celem analizy jest zbadanie charakterystyk rozkładów cen rzeczywistych produktów. Otrzymane wyniki zostaną użyte do weryfikacji zgodności rozkładów cen uzyskiwanych w modelu Piskor-Ignatowicza-Zachary z rozkładami cen realnych produktów.

7.1. Metodyka analizy statystycznej pozyskanych danych

Dla każdego modelu kluczowym zagadnieniem jest porównanie otrzymywanych w nim wyników z rzeczywistymi danymi. Dotyczy to zarówno modeli, których celem jest parametryczny opis realnego świata lub jego aspektu, jak i modeli posiadających określoną moc predykcji. Do pierwszej grupy możemy zaliczyć modele, które imitują rozkład jakiejś własności bez wnikania w fizykę procesów. Najczęściej są to modele wieloparametrowe, opisujące przy pomocy funkcji matematycznych doświadczalnie obserwowane rozkłady. Drugą grupę tworzą modele symulujące fizyczne aspekty oddziaływania bądź dynamiki procesu - od tych modeli często oczekujemy zdolności predykcyjnych. Dla obu grup modeli ostatecznym weryfikatorem są dane z pomiarów.

W celu weryfikacji modelu Piskor-Ignatowicza-Zachary dokonano pomiarów rzeczywistych rozkładów cen w sklepach internetowych. Użyto w tym celu dane/wyniki wszystkich sklepów dostępnych w porównywarce cen Ceneo.pl. Analizę danych doświadczalnych przeprowadzono przy użyciu środowiska R (R Core Team 2017) i wykorzystano funkcje z pakietu 'poweRlaw' (Gillespie 2015).

Do sprawdzenia gaussowskości rozkładów zaprezentowanych danych, zastosowano test Shapiro-Wilka służący do sprawdzenia normalności rozkładu. Hipoteza zerowa testu Shapiro-Wilka zakłada, iż rozkład jest normalny. Jeżeli p – wartość $<$ przyjęty poziom istotności, to odrzucamy hipotezę zerową

o gaussowskości rozkładu (Shapiro and Wilk 1965). Warto w tym miejscu wspomnieć, iż test Shapiro-Wilka ma największą moc statystyczną spośród wszystkich znanych testów używanych do sprawdzania normalności rozkładu (i jest ona nawet większa niż moc testu Kołmogorowa-Smirnowa).

Ponieważ we wszystkich zbadanych rozkładach cen produktów zaobserwowano występowanie asymetrii rozkładu w kierunku wyższych cen, kolejnym krokiem analizy danych jest przetestowanie hipotez o ciężkoogonowym charakterze rozkładów cen.

Wykresy rozkładu $P(X > x)$ (ang. CCDF - Complementary Cumulative Distribution Function) od ceny X wyrażonej w [PLN], dla wszystkich badanych danych sporządzone są na skali logarytmiczno-logarytmicznej i przejawiają liniowy charakter począwszy od pewnej wartości x_{min} . Zaobserwowana liniowa zależność na skali logarytmiczno-logarytmicznej może sugerować, iż mamy do czynienia z rozkładem potęgowym, ale bynajmniej jeszcze tego nie dowodzi. Koniecznym jest, więc przeprowadzenie odpowiednich testów statystycznych. O skomplikowaniu tego typu badań może świadczyć fakt, iż poprawną ich metodologię opisano dopiero w 2009 roku. Procedura postępowania w takich przypadkach została zaprezentowana w pracy (Clauset, Shalizi, and Newman 2009) i jest następująca:

- Wyznaczenie wartości x_{min} oraz wartości wykładnika rozkładu potęgowego α .
- Wyliczenie dobroci dopasowania pomiędzy danymi a rozkładem potęgowym. Do wyznaczenia dobroci dopasowania używa się metody bootstrap, pozwalającej wyznaczyć zarówno p – wartość, jak i niepewności parametrów dopasowania. Jeżeli otrzymana p – wartość jest większa od 0.1, to uznajemy dopasowanie rozkładu potęgowego, jako rozsądne, w przeciwnym razie odrzucamy tę hipotezę.

Jak wiadomo, p – wartość nie mówi o prawdopodobieństwie prawdziwości uzyskanego dopasowania (nie ma zależności, że im większa p – wartość tym lepsze dopasowanie). p – wartość, jest testem typu 0 lub 1 to jest, jeżeli: p – wartość < 0.1 , to odrzucamy dany rozkład, a gdy p – wartość > 0.1 , to na poziomie istotności 0.1, nie ma podstaw odrzucić danego rozkładu.

- Porównanie rozkładu potęgowego do rozkładów alternatywnych: logarytmiczno-normalnego, wykładniczego, Weibulla. Do porównania

dobroci dopasowania rozkładu potęgowego w stosunku do rozkładu alternatywnego, stosujemy test Vuong'a.

Do wyznaczenia x_{min} i α wykorzystuje się statystykę Kołmogorowa-Smirnowa:

$$D = \max_{x > x_{min}} (F(x) - F^0(x)) \quad (7-1)$$

Gdzie $F^0(x)$ oraz $F(x)$ to odpowiednio dystrybuanty modelu i danych. Procedura do wyznaczenia estymatora x_{min} polega na znalezieniu takiej wartości, x_{min} dla której D przyjmuje minimum. W trakcie poszukiwania optymalnej wartości x_{min} dla każdego badanego x wyznacza się odpowiadający mu wykładnik α przy użyciu estymatora największej wiarygodności odpowiednio dla danych ciągłych postaci:

$$\alpha = 1 + n[\sum_{i=1}^n \ln \frac{x_i}{x_{min}}]^{-1} \quad (7-2)$$

Oraz dla danych dyskretnych, stosujemy aproksymację postaci:

$$\alpha \cong 1 + n[\sum_{i=1}^n \ln \frac{x_i}{x_{min}-0.5}]^{-1} \quad (7-3)$$

Pomimo, iż dla konkretnej wartości x_{min} niepewność wyznaczenia parametru skali α dla rozkładu potęgowego jest wyrażona analitycznie, to, aby wyznaczyć niepewność x_{min} stosuje się metodę bootstrap (Efron and Tibshirani 1993). Metoda ta pozwala również na wyznaczenie p – wartości jak i niepewności pozostałych parametrów rozkładu. Metoda bootstrap jest uniwersalna i może być stosowana praktycznie dla każdego typu rozkładu. Dla N elementowego zbioru obserwacji polega na N krotnym losowaniu ze zwracaniem elementów z oryginalnego zbioru obserwacji, a następnie B krotnym powtórzeniu całej procedury (gdzie $B \gg 1$). Dla każdej iteracji powtarza się ponowne wyznaczanie estymatorów interesujących nas wielkości, dzięki czemu otrzymujemy rozkład prawdopodobieństwa szukanych estymatorów. Warto w tym miejscu wspomnieć, iż wariancja odchylenia standardowego badanego estymatora jest funkcją zarówno liczebności obserwacji N jak i ilości bootstrapów B i ma postać:

$$var(\sigma_{boot}) \sim \frac{1}{N^2} + \frac{1}{NB} \quad (7-4)$$

Dlatego w praktyce używamy dużej liczby bootstrapów B , najczęściej rzędu 1000 czy nawet 10000. Oczywiście idealnie było by mieć również dużą ilość obserwacji doświadczalnych N , niestety w wielu praktycznych sytuacjach nie mamy

wpływu na ilość dostępnych danych, jak choćby na ilość sklepów internetowych, w których sprzedawany jest dany towar.

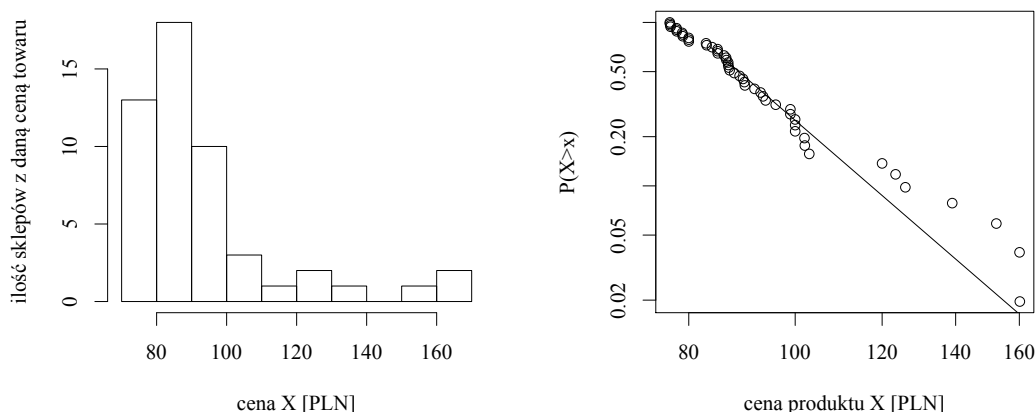
Kolejnym krokiem analizy danych jest sprawdzenie alternatyw dla rozkładu potęgowego. Dla każdego badanego rozkładu podobnie jak dla rozkładu potęgowego stosujemy metodę opartą na statystyce D Kołmogorowa-Smirnowa i wyznaczamy x_{min} , oraz parametry rozkładu w ten sposób by uzyskać minimalizację D . Metodę bootstrap stosujemy do wyznaczenia niepewności otrzymanych parametrów i wyznaczenia p – wartości. W celu porównania badanego rozkładu z rozkładem potęgowym stosujemy test Vuong’a (Vuong 1989). Test ten oparty jest na ilorazie wiarygodności (ang. likelihood ratio test) i wykorzystuje kryterium Kullback-Leiblera (ang. Kullback-Leibler criterion) do wyboru modelu. Hipoteza zerowa postuluje, że oba modele są równie bliskie danym doświadczalnym. Hipoteza alternatywna mówi, iż jeden z modeli jest bliższy. Statystyka testowa R , jest stosunkiem logarytmów wiarygodności pomiędzy dwoma porównywanymi modelami. Znak R wskazuje, który model jest lepszy. W celu uzyskania informacji czy otrzymany znak R nie jest wynikiem fluktuacji należy uwzględnić otrzymaną z testu p – wartość, jeżeli jest mniejsza od 0.1, to możemy traktować, iż znak R nie jest wynikiem fluktuacji i wynik testu jest znaczący statystycznie. Dokładny opis metody można znaleźć w dodatku C w pracy (Clauset, Shalizi, and Newman 2009).

W dalszej części przedstawiam wyniki analizy statystycznej cen dziesięciu losowo wybranych produktów. Analizowane dane zostały pobrane z portalu Ceneo piętnastego lipca 2016 roku.

7.1.1. Badania rozkładu cen czajnika „Philips HD 4646/00 Biały”

Poniżej na rysunku po lewej stronie zaprezentowano przykład rozkładu ceny czajnika marki Philips dostępnego w 51 sklepach internetowych, oraz odpowiadający

mu rozkład $P(X > x)$ od ceny X wyrażonej w [PLN], przedstawiony po prawej stronie ilustracji Rys. 7-1.

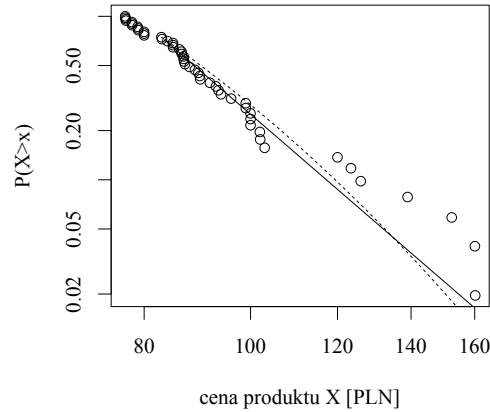


Rys. 7-1 Po lewej stronie rysunku znajduje się histogram przedstawiający rozkład ilości sklepów internetowych z daną ceną czajnika “Philips HD 4646/00 Biały”. Po prawej stronie ilustracji przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwarte kółka przedstawiają dane doświadczalne. Linia ciągła odpowiada dopasowanej zależności potęgowej.

Na rysunku Rys. 7-1 po lewej stronie przedstawiono na osi rzędnych cenę czajnika, a na osi odciętych ilość sklepów internetowych, w których można kupić ten czajnik w zadanym przedziale cenowym. Zaprezentowany rozkład ma asymetryczny kształt z długim ogonem w kierunku większych cen. Otrzymana z testu Shapiro-Wilka wartość statystyki $W = 0.74$, oraz p – wartość $= 4.7e - 08$, co pozwala na wykluczenie hipotezy o gaussowskim rozkładzie cen czajnika “Philips HD 4646/00 Biały”.

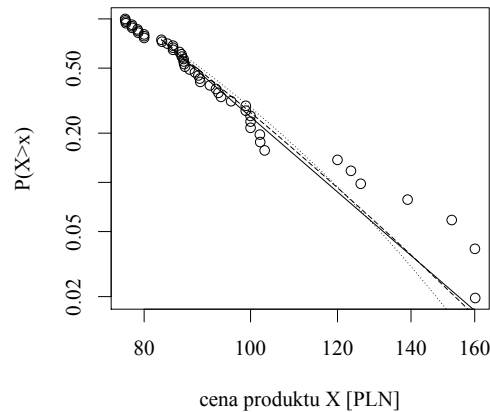
W celu weryfikacji hipotezy o ciężkoogonowym rozkładzie cen, na obrazku Rys. 7-1 po prawej stronie przedstawiono zależność $P(X > x)$ od ceny X wyrażonej w [PLN] wraz z dopasowanym rozkładem potęgowym o następujących wartościach parametrów: $x_{min} = 83.0 \pm 4.2$; $\alpha = 6.8 \pm 1.4$. Otrzymana dla tego dopasowania p – wartość $= 0.66 > 0.1$, podobnie jak wartość testu Kołmogorowa-Smirnowa $KS = 0.074$, mniejsza od wartości krytycznej na poziomie ufności $Cl = 0.9$ dla ilości pomiarów $N = 36$, wynoszącej $KS_{critical} = 0.19$, wskazują na dobrą, jakość dopasowania. Niemniej zgodnie z zaleceniem zawartym w pracy (Clauset, Shalizi, and Newman 2009), przetestowane zostaną również dopasowania alternatywnych

rozkładów ciężkoogonowych, przy użyciu testu Vuong’a (Gillespie 2014)(Vuong 1989).



Rys. 7-2 Rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono ceny czajnika „Philips HD 4646/00 Biały”. Linia ciągła odpowiada dopasowanej zależności potęgowej (jak na wcześniejszej ilustracji). Linia przerywana odpowiada zależności logarytmiczno-normalnej.

Linia przerywana na ilustracji Rys. 7-1 odpowiada zależności logarytmiczno-normalnej o następujących wartościach parametrów: $x_{min} = 76.9 \pm 2.4$; $\mu = 0.25 \pm 0.93$; $\sigma = 0.93 \pm 0.26$, oraz wartościach dopasowania: $p - value = 0.17$; $KS = 0.081$. Vuong test: $test_{statistic} = 1.04$, $p - value = 0.18$. P-wartość większa od 0.1, dla testu Vuong’a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-normalny to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu logarytmiczno-normalnego.



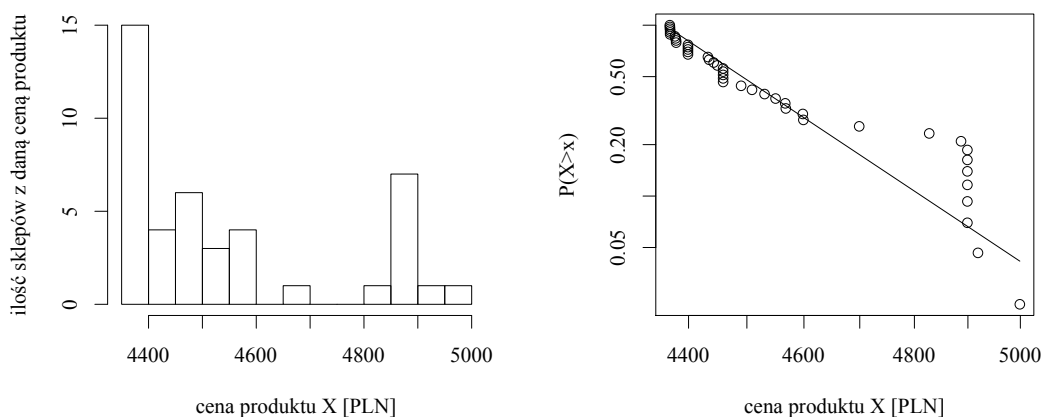
Rys. 7-3 Rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne. Linia ciągła odpowiada dopasowanej zależności potęgowej (jak na wcześniejszych ilustracjach). Linia przerywana odpowiada dopasowanemu rozkładowi Weibulla. Linia kropkowana odpowiada dopasowanemu rozkładowi wykładniczemu. Zarówno parametry dopasowanych rozkładów, jak i parametry dopasowania oraz wyniki testów Vuong'a podano poniżej w tekście.

Parametry dopasowanego rozkładu Weibulla przedstawionego linią przerywaną na rysunku Rys. 7-3 są następujące: $x_{min} = 76.9 \pm 2.7$; $k = 0.33 \pm 0.58$; $\lambda = 0 \pm 15$. Wartości parametrów dopasowania wynoszą: $p - value = 0.20$; $KS = 0.076$. Vuong test: $test_{statistic} = 0.57$, $p - value = 0.29$. p -wartość większa od 0.1, dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla. Linia kropkowana odpowiada dopasowanemu rozkładowi wykładniczemu o następujących parametrach: $x_{min} = 76.9 \pm 3.7$; $\lambda = 0.055 \pm 0.012$ i wartościach dopasowania: $p - value = 0.24$; $KS = 0.082$. Vuong test: $test_{statistic} = 1.2$; $p - value = 0.12$. P -wartość większa od 0.1, dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż wykładniczy to ilość

danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu wykładniczego.

7.1.2. Analiza rozkładu cen lodówki “Samsung RSH5ZLBG”

Rysunek poniżej przedstawia rozkład cen z wtórnym maksimum cenowym.



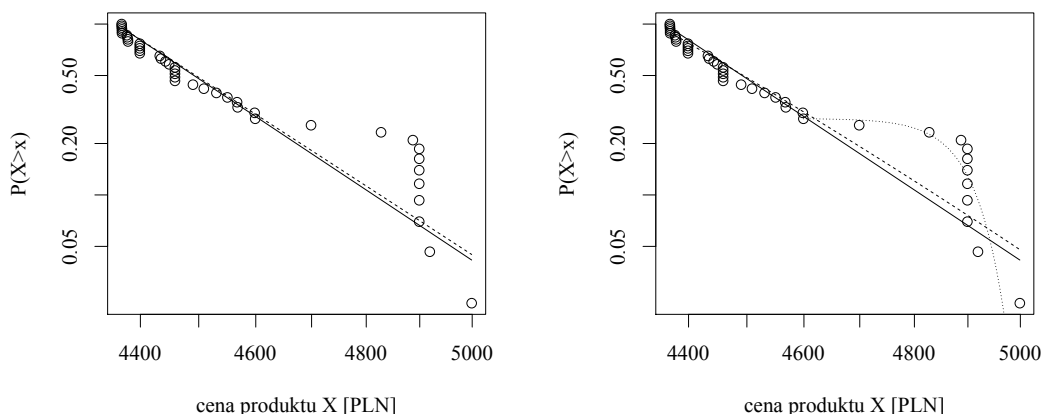
Rys. 7-3 Po lewej stronie ilustracji przedstawiono rozkład cen lodówki “Samsung RSH5ZLBG” w 44 sklepach internetowych z portalu Ceneo.pl. Po prawej stronie rysunku zaprezentowano rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne. Linia ciągła odpowiada dopasowanej zależności potęgowej. Własności rozkładów opisano poniżej w tekście.

Po lewej stronie obrazka Rys. 7-3 przedstawiono rozkład cen lodówki “Samsung RSH5ZLBG” w 44 sklepach internetowych z portalu Ceneo.pl. Dla ceny 4899.00[PLN], obserwujemy wtórne maksimum (na wykresie rozkładu $P(X > x)$ od ceny X wyrażonej w [PLN], po prawej stronie rysunku odpowiada ono „uskokowi”). Maksimum to, (o ile nie jest fluktuacją) odpowiada grupie towaru, zakupionej w innych warunkach rynkowych. Niestety na obecnym etapie badań nie znaleziono kryterium pozwalającego w łatwy sposób rozróżnić fluktuacji od przypadków odpowiadających innym warunkom rynkowym. Wydaje się jednakże, iż w celu rozróżnienia obu tych przypadków konieczna była by obserwacja zachowania w czasie, fluktuacje cenowe powinny znikać, a struktury odpowiadające różnym sytuacjom rynkowym powinny być o wiele trwalsze. Zagadnienia dynamiki cenowej nie są jednak przedmiotem niniejszej pracy.

Weryfikację hipotezy o gaussowskim rozkładzie cen lodówki przeprowadzono przy użyciu testu Shapiro-Wilka i otrzymano następujące wartości W oraz p –

wartości: $W = 0.79$; $p - value = 2.1e - 06$, co pozwala na odrzucenie hipotezy o gaussowskim rozkładzie cen lodówki “Samsung RSH5ZLBG”.

Po prawej stronie rysunku Rys. 7-4 ukazano zależność $P(X > x)$ od ceny X w [PLN], wraz z dopasowaną zależnością potęgową o następujących parametrach rozkładu: $x_{min} = 4369 \pm 18$; $\alpha = 24.2 \pm 5.2$, oraz następujących parametrach dopasowania: $p - value = 0.0022$; $KS = 0.15$. $p - wartość < 0.1$, wskazuje, że należy odrzucić hipotezę rozkładu potęgowego. Jak widać w niniejszym przykładzie, stosowanie testu $p - wartość$ daje znacznie bardziej rygorystyczne kryterium odrzucenia danego dopasowania niż test Kołmogorowa-Smirnowa, dla którego wartość krytyczna dla 41 pomiarów na poziomie ufności $Cl = 0.9$, wynosi: $KS_{critical} = 0.19$ i ponieważ $KS < KS_{critical}$, to nie mieli byśmy podstaw odrzucić rozkładu potęgowego.



Rys. 7-4 Na ilustracjach przedstawiono zależność $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne cen lodówki “Samsung RSH5ZLBG” z 44 sklepów. Linia ciągła odpowiada dopasowanej zależności potęgowej. Na wykresie po lewej stronie linią przerywaną zaznaczono dopasowany do danych rozkład logarytmiczno-normalny. Po prawej stronie rysunku linia przerywana odpowiada dopasowanej zależności wykładniczej, a linia kropkowana dopasowanemu do danych rozkładowi Weibulla. Parametry rozkładów i ich dopasowań omówiono poniżej w tekście.

Po lewej stronie rysunku Rys. 7-4 linią przerywaną przedstawiono rozkład logarytmiczno-normalny o następujących parametrach: $x_{min} = 4368.6$; $\mu = -28.49$; $\sigma = 1.28$. Parametry dopasowania wynoszą: $p - value = error$; $KS = error$, gdyż nie udało się uruchomić metody bootstrap_p w związku z niestabilnością numeryczną metody dla posiadanych danych doświadczalnych, z tego samego

powodu nie można było wyznaczyć niepewności parametrów dopasowanego rozkładu.

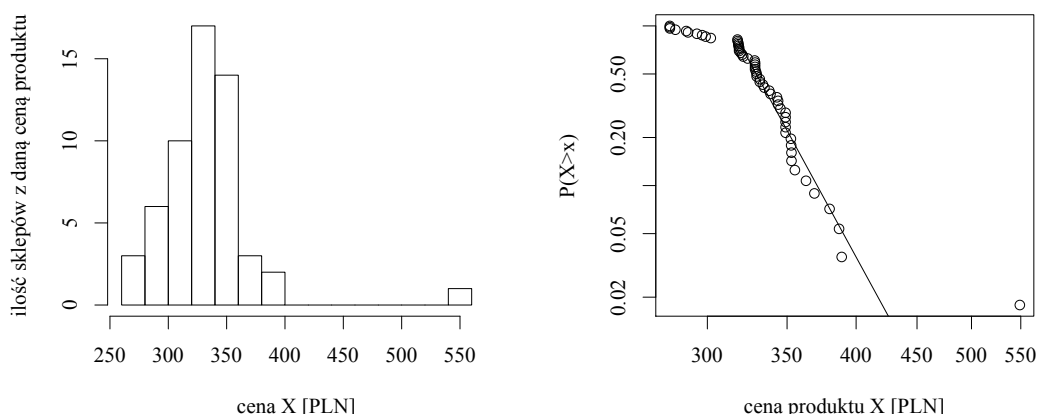
W związku z niemożnością zastosowania metody bootstrap do oceny, jakości dopasowania rozkładu logarytmiczno-normalnego, możemy posłużyć się testem Vuong'a, aby odpowiedzieć na pytanie, które z dopasowań jest lepsze potęgowe, czy logarytmiczno-normalne. Vuong test: $test_{statistic} = 0.083, p - value = 0.93$. P - wartość większa od 0.1, dla testu Vuong'a wskazuje, że, mimo iż wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-normalny to ilość danych doświadczalnych jest niewystarczająca by wynik ten był znaczący statystycznie.

Po prawej stronie ilustracji Rys. 7-4 linią przerywaną oznaczono dopasowany rozkład wykładniczy o następujących parametrach: $x_{min} = 4377 \pm 33$; $\lambda = 0.00464 \pm 0.00099$ i wartościach dopasowania: $p - value = 0.0033$; $KS = 0.15$. Otrzymana $p - wartość$ mniejsza od 0.1, pozwala na odrzucenie hipotezy, że dane są opisywane rozkładem wykładniczym.

Na ilustracji Rys. 7-4 po prawej stronie linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla uzyskane parametry rozkładu są następujące: $x_{min} = 4599 \pm 66$; $k = 87 \pm 39$; $\lambda = 4914 \pm 26$. Wartości parametrów dopasowania wynoszą: $p - value = 0.03$; $KS = 0.22$. $P - wartość < 0.1$, pozwala na odrzucenie hipotezy, iż dane są opisywane rozkładem Weibulla.

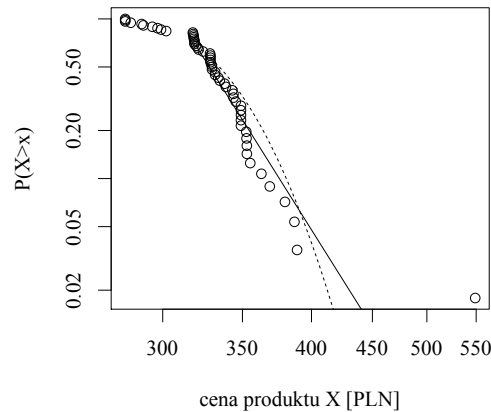
7.1.3. Badania rozkładu cen monitora „BenQ 18,5 GL955A”

Poniżej zaprezentowano rozkład cen dla monitora “BenQ 18,5 GL955A” w 56 sklepach internetowych.



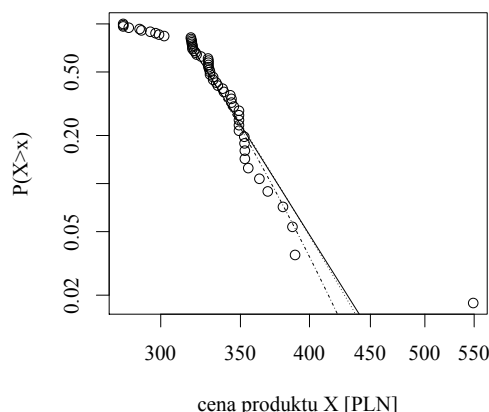
Rys. 7-5 Po lewej stronie przedstawiono rozkład cen monitora „BenQ 18,5 GL955A” w 56 sklepach internetowych z portalu Ceneo.pl. Po prawej stronie ukazano zależność $P(X > x)$ od ceny X w [PLN], wraz z dopasowaną zależnością potęgową.

Po lewej stronie ilustracji Rys. 7-5 przedstawiono rozkład cen monitora „BenQ 18,5 GL955A” w 56 sklepach internetowych z portalu Ceneo.pl, otrzymana wartość testu Shapiro-Wilka W oraz p – wartość wynoszą $W = 0.75$; $p - value = 1.8e - 08$, co pozwala na odrzucenie hipotezy o gaussowskim rozkładzie cen monitora. Po prawej stronie rysunku Rys. 7-5 ukazano zależność $P(X > x)$ od ceny X w [PLN], wraz z dopasowaną zależnością potęgową o następujących parametrach rozkładu: $x_{min} = 320.0 \pm 3.2$; $\alpha = 12.9 \pm 2.9$, oraz następujących parametrach dopasowania: $p - value = 0.057$; $KS = 0.074$. $p - wartość < 0.1$, wskazuje, że w świetle danych doświadczalnych należy odrzucić hipotezę o rozkładzie potęgowym na poziomie istotności 0.1.



Rys. 7-6 Zależność $P(X > x)$ od ceny X w [PLN], otwartymi kółkami zaznaczono ceny monitora „BenQ 18,5 GL955A”, linia ciągła odpowiada dopasowanej zależności potęgowej, linia przerywana przedstawia dopasowaną zależność logarytmiczno-normalną.

Na ilustracji Rys. 7-6 przedstawiono zależność $P(X > x)$ od ceny X w [PLN], otwartymi kółkami zaznaczono ceny monitora „BenQ 18,5 GL955A”, linia ciągła odpowiada dopasowanej zależności potęgowej, linia przerywana przedstawia dopasowaną zależność logarytmiczno-normalną o następujących parametrach rozkładu: $x_{min} = 320 \pm 22$; $\mu = 5.807 \pm 0.050$; $\sigma = 0.104 \pm 0.028$, parametry dopasowania wynoszą: $p - value = 0$; $KS = 0.16$. $P - wartość = 0$, wyznaczona metodą bootstrap pozwala na odrzucenie hipotezy rozkładu logarytmiczno-normalnego. Vuong test potwierdza lepsze dopasowanie rozkładu potęgowego: $test_{statistic} = 0.94$, $p - value = 0.17$. Niemniej $p - wartość$ większa od 0.1, dla testu Vuong’a wskazuje, na małą ilość danych doświadczalnych.



Rys. 7-7 Rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne ceny monitora „BenQ 18,5 GL955A”. Linia ciągła odpowiada dopasowanej zależności potęgowej (jak na wcześniejszych ilustracjach). Linia przerywana odpowiada dopasowanemu rozkładowi Weibulla. Linia kropkowana odpowiada dopasowanemu rozkładowi wykładniczemu. Zarówno parametry dopasowanych rozkładów, jak i parametry dopasowania oraz wyniki testów Vuong’a podano poniżej w tekście.

Parametry dopasowanego rozkładu Weibulla przedstawionego linią przerywaną na rysunku Rys. 7-7, są następujące: $x_{min} = 320.0 \pm 1.1$; $k = 0.355 \pm 0.062$; $\lambda = 0.018 \pm 0.018$. Wartości parametrów dopasowania wynoszą: $p - value = error$; $KS = 0.12$. Vuong test: $test_{statistic} = 0.36$, $p - value = 0.36$.

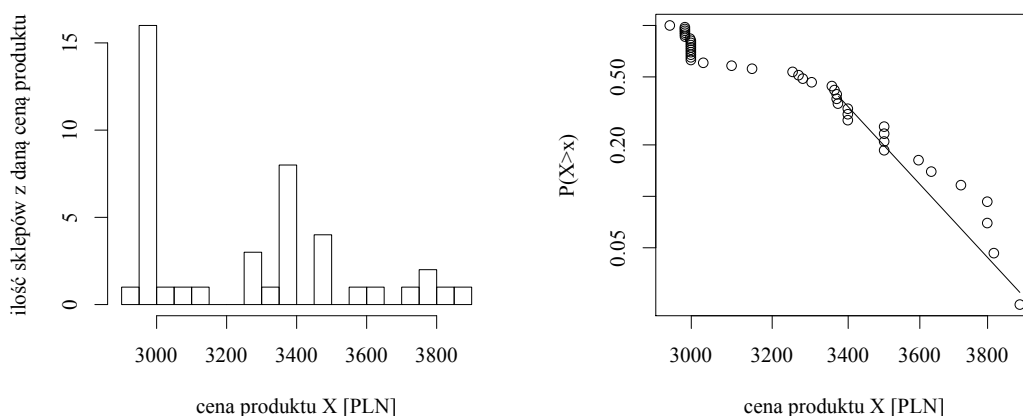
p -wartość większa od 0.1, dla testu Vuong’a wskazuje, iż, mimo, że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla.

Linia kropkowana odpowiada dopasowanemu rozkładowi wykładniczemu o następujących parametrach: $x_{min} = 320.0 \pm 5.0$; $\lambda = 0.0379 \pm 0.0082$ i wartościach dopasowania uzyskanych metodą bootstrap: $p - value = 0.12$; $KS = 0.095$. Co pozwala przyjąć hipotezę o wykładniczym charakterze rozkładu cen monitora BenQ. Vuong test: $test_{statistic} = 0.41$, $p - value = 0.34$. p -wartość większa od 0.1, dla testu Vuong’a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż wykładniczy to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu wykładniczego. Warto w tym miejscu zwrócić uwagę, że, mimo iż, parametr dopasowania rozkładu wykładniczego $p - wartość > 0.1$, a, dla dopasowania rozkładu potęgowego $p - wartość < 0.1$ (patrz opis pod rysunkiem

Rys. 7-5) to test Vuong’a wskazuje na lepsze dopasowanie rozkładu potęgowego niż wykładniczego, jednakże wskazanie to nie jest znaczące w sensie statystycznym. Podczas gdy metoda bootstrap jednoznacznie wskazuje na rozkład wykładniczy.

7.1.4. Analiza statystyczna cen lodówki „Samsung RL55VTEBG”

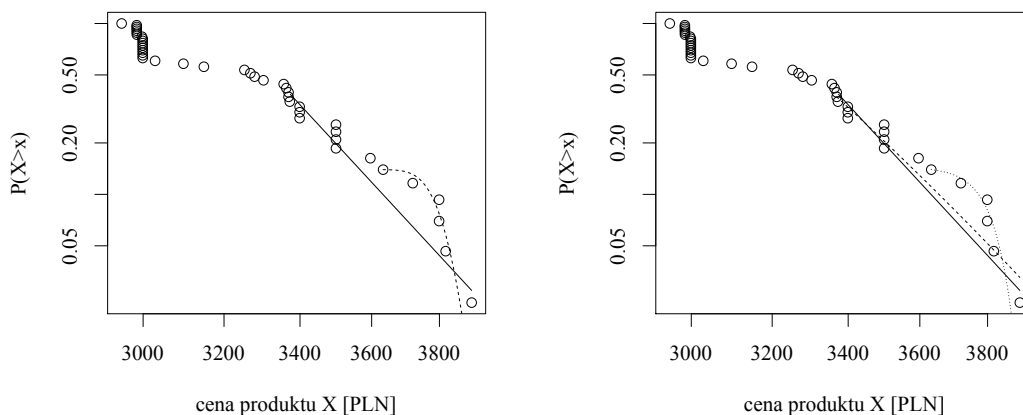
Poniżej zaprezentowano rozkład cen dla lodówki “Samsung RL55VTEBG” w 43 sklepach internetowych z portalu Ceneo.pl.



Rys. 7-8 Histogram po lewej stronie ilustracji ukazuje rozkład cen lodówki “Samsung RL55VTEBG” w 43 sklepach. Po prawej stronie rysunku pokazano zależność $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwarte kółka odpowiadają danym doświadczalnym, a linia prosta przedstawia dopasowany do nich rozkład potęgowy.

Po lewej stronie ilustracji Rys. 7-8 przedstawiono rozkład cen lodówki “Samsung RL55VTEBG” w 43 sklepach internetowych z portalu Ceneo.pl, otrzymana wartość testu Shapiro-Wilka W oraz p – wartość wynoszą $W = 0.87$; p – $value = 0.00016$, co pozwala na odrzucenie hipotezy o gaussowskim rozkładzie cen lodówki. Po prawej stronie rysunku Rys. 7-8 ukazano zależność $P(X > x)$ od ceny X w [PLN], wraz z dopasowaną zależnością potęgową o następujących parametrach rozkładu: $x_{min} = 3369 \pm 46$; $\alpha = 19.3 \pm 6.7$, oraz następujących parametrach dopasowania: p – $value = 0.14$; $KS = 0.15$. p – wartość > 0.1 , wskazuje, że dopasowanie rozkładu potęgowego jest rozsądne. Wartość krytyczna dla 17 pomiarów

na poziomie ufności $Cl = 0.9$, dla testu Kołmogorowa-Smirnowa wynosi: $KS_{critical} = 0.29$.



Rys. 7-9 Wykresy przedstawiają zależność $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwarte kółka odpowiadają danym doświadczalnym cen lodówki “Samsung RL55VTEBG”, a linia prosta przedstawia dopasowany do nich rozkład potęgowej. Na ilustracji po lewej stronie linią przerywaną przedstawiono dopasowany do danych rozkład logarytmiczno-normalny. Po prawej stronie linia przerywana odpowiada dopasowanemu rozkładowi wykładniczemu, a krzywa kropkowana dopasowanemu rozkładowi Weibulla. Parametry rozkładów i ich dopasowania omówiono poniżej w tekście.

Wykresy na obrazku Rys. 7-9 przedstawiają zależność $P(X > x)$ od ceny X w [PLN], otwartymi kółkami zaznaczono ceny lodówki “Samsung RL55VTEBG”, linia ciągła odpowiada dopasowanej zależności potęgowej. Linia przerywana na ilustracji Rys. 7-9 po lewej stronie odpowiada dopasowanemu do danych rozkładowi logarytmiczno-normalnemu o następujących parametrach: $x_{min} = 3630 \pm 170$; $\mu = 8.244 \pm 0.031$; $\sigma = 0.015 \pm 0.020$, parametry dopasowania wynoszą: $p - value = 0.69$; $KS = 0.11$. $P - wartość > 0.1$, wyznaczona metodą bootstrap wskazuje, że dopasowanie rozkładu logarytmiczno-normalnego jest sensowne. Niemniej należy zwrócić uwagę, iż rozkład logarytmiczno-normalny, został dopasowany jedynie do 6 ostatnich punktów pomiarowych, co ewidentnie grozi wystąpieniem zjawiska „nadmiernego dopasowania” (ang. overfitting)(Alstott, Bullmore, and Plenz 2014). Test Vuong’a sugeruje lepsze dopasowanie rozkładu potęgowego niż logarytmiczno-normalnego: $test_{statistic} = 0.14$, $p - value = 0.89$. Niemniej $p - wartość$ większa od 0.1, dla testu Vuong’a wskazuje, na małą ilość danych doświadczalnych, a tym samym, iż wynik testu może być odzwierciedleniem fluktuacji danych a nie rzeczywistej zależności.

Linia przerywana na ilustracji Rys. 7-9 po prawej stronie odpowiada dopasowanemu rozkładowi wykładniczemu o następujących parametrach: $x_{min} =$

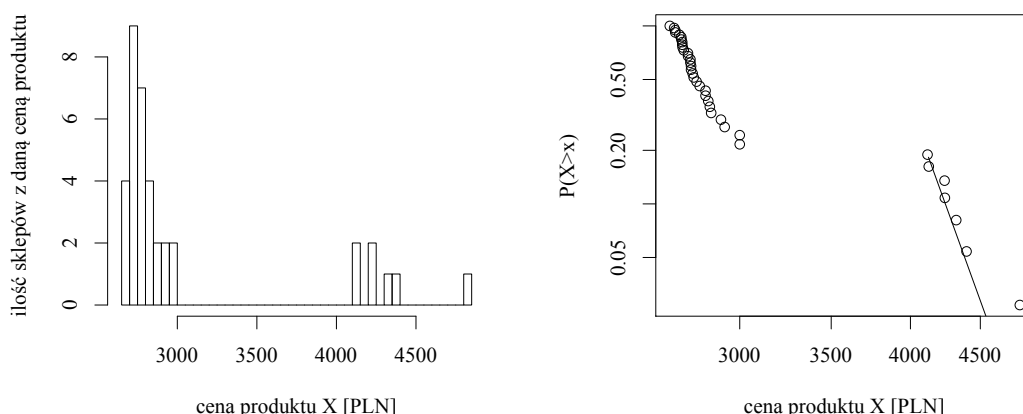
3369 ± 82 ; $\lambda = 0.0046 \pm 0.0015$ i wartościach dopasowania uzyskanych metodą bootstrap: $p - value = 0.11$; $KS = 0.14$. Co pozwala przyjąć hipotezę o wykładniczym charakterze rozkładu cen lodówki “Samsung RL55VTEBG”.

Vuong test: $test_{statistic} = -0.25$, $p - value = 0.80$. $P - wartość$, większa od 0.1, dla testu Vuong’a wskazuje, iż, mimo że wartość statystyki testowej jest mniejsza, od 0, co wskazuje, że rozkład wykładniczy jest lepiej dopasowany niż rozkład potęgowy to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu potęgowego.

Linia kropkowana na ilustracji Rys. 7-9 po prawej stronie odpowiada dopasowanemu rozkładowi Weibulla o następujących parametrach: $x_{min} = 3630 \pm 220$; $k = 64 \pm 34$; $\lambda = 3830 \pm 120$. Wartości parametrów dopasowania wynoszą: $p - value = 0.63$; $KS = 0.13$. Rozkład Weibulla podobnie jak w przypadku rozkładu logarytmiczno-normalnego, został dopasowany jedynie do 6 ostatnich punktów pomiarowych, a tym samym, sensowność tego dopasowania stoi pod bardzo dużym znakiem zapytania. Vuong test: $test_{statistic} = -0.40$, $p - value = 0.68$. $P - wartość$ większa od 0.1, dla testu Vuong’a wskazuje, iż, mimo, że wartość statystyki testowej jest mniejsza, od 0, co wskazuje, że rozkład Weibulla jest lepiej dopasowany niż rozkład potęgowy, to ilość danych doświadczalnych jest niewystarczająca by wynik testu Vuong’a móc traktować, jako znaczący statystycznie.

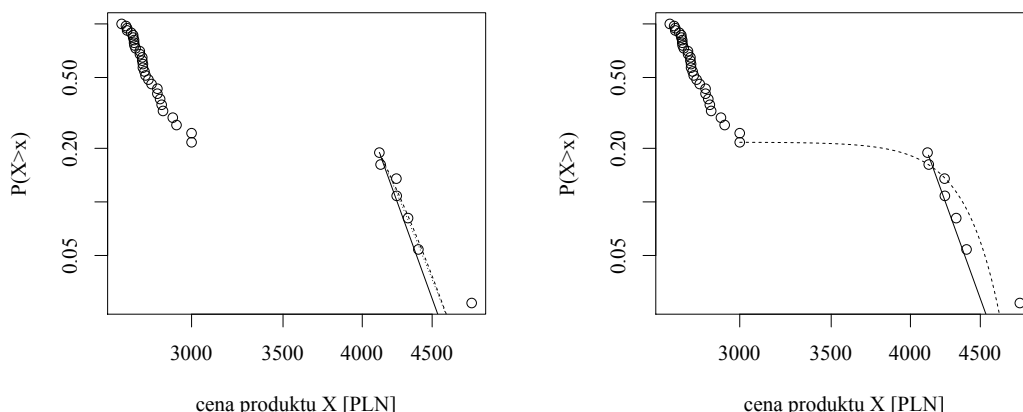
7.1.5. Badania statystyczne rozkładu cen komputera „Lenovo ThinkCenter M83 Tower (10BE001APB)”

Poniżej zaprezentowano rozkład cen dla komputera stacjonarnego “Lenovo ThinkCentre M83 Tower (10BE001APB)” w 39 sklepach internetowych z portalu Ceneo.pl.



Rys. 7-10 Po lewej stronie przedstawiono rozkład cen komputera stacjonarnego „Lenovo ThinkCenter M83 Tower (10BE001APB)”, w 39 sklepach internetowych z portalu Ceneo.pl. Po prawej stronie przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN], wraz z dopasowanym rozkładem potęgowym (linia ciągła).

Na ilustracji Rys. 7-10 po lewej stronie przedstawiono rozkład cen komputera stacjonarnego „Lenovo ThinkCenter M83 Tower (10BE001APB)” w 39 sklepach internetowych z portalu Ceneo.pl, otrzymane wartości z testu Shapiro-Wilka wynoszą: $W = 0.62$; $p - value = 1.3e - 08$, co pozwala na jednoznaczne odrzucenie hipotezy o gaussowskości rozkładu cen komputera. Po prawej stronie przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN], wraz z dopasowanym rozkładem potęgowym (linia ciągła) o następujących wartościach parametrów: $x_{min} = 4116 \pm 59$; $\alpha = 22 \pm 11$, a parametry dopasowania wynoszą: $p - value = 0.59$; $KS = 0.17$, przy wartości krytycznej testu Kołmogorowa-Smirnowa (KS) na poziomie istotności 0.9 i liczebności próby 7, wynoszącej 0.43. Tak, więc rozkład potęgowy stanowi rozsądne dopasowanie.



Rys. 7-11 Rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne ceny komputera stacjonarnego „Lenovo ThinkCenter M83 Tower (10BE001APB)”. Na ilustracji po lewej stronie linia ciągła odpowiada dopasowanej zależności potęgowej (jak na wcześniejszej ilustracji). Linia przerywana odpowiada dopasowanemu rozkładowi logarytmiczno-normalnemu, linia kropkowana odpowiada dopasowanemu rozkładowi wykładniczemu. Na rysunku po prawej stronie przedstawiono ponownie linią ciągłą rozkład potęgowy a linią przerywaną rozkład Weibulla. Zarówno parametry dopasowanych rozkładów, jak i parametry dopasowania oraz wyniki testów Vuong’a podano poniżej w tekście.

Na ilustracji Rys. 7-11 po lewej stronie linią przerywaną przedstawiono rozkład logarytmiczno-normalny o następujących parametrach: $x_{min} = 4117$; $\mu = 8.09$; $\sigma = 0.14$. Parametry dopasowania wynoszą: $p - value = error$; $KS = error$, gdyż nie udało się uruchomić metody bootstrap_p w związku z niestabilnością numeryczną metody dla posiadanych danych doświadczalnych, z tego samego powodu nie można było wyznaczyć niepewności parametrów dopasowanego rozkładu.

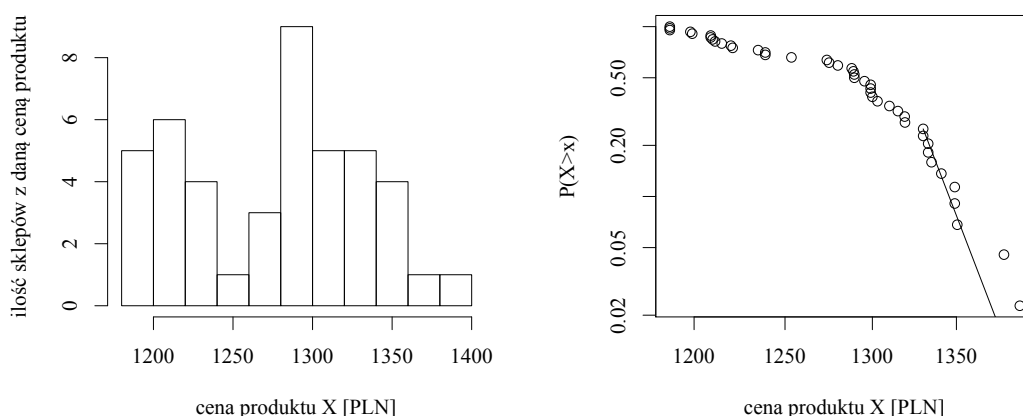
Vuong test: $test_{statistic} = 0.29$, $p - value = 0.39$. P-wartość większa od 0.1 dla testu Vuong’a wskazuje, że, mimo iż wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-normalny to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu logarytmiczno-normalnego.

Kolejnym dopasowanym rozkładem na ilustracji Rys. 7-11 po lewej stronie linią kropkowaną przedstawiono rozkład wykładniczy o następujących parametrach: $x_{min} = 4117 \pm 75$; $\lambda = 0.0043 \pm 0.0019$ i wartościach dopasowania: $p - value = 0.77$; $KS = 0.12$. Vuong test: $test_{statistic} = 0.19$, $p - value = 0.43$. P-wartość większa od 0.1, dla testu Vuong’a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż wykładniczy to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu wykładniczego.

Na ilustracji Rys. 7-11 po prawej stronie linią przerywaną przedstawiono dopasowanie rozkładu Weibulla uzyskane parametry rozkładu są następujące: $x_{min} = 3880 \pm 360$; $k = 23 \pm 15$; $\lambda = 4430 \pm 140$. Wartości parametrów dopasowania wynoszą: $p - value = 0.15$; $KS = 0.13$. Vuong test: $test_{statistic} = 1.23$, $p - value = 0.11$. P-wartość większa od 0.1 dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla.

7.1.6. Analiza statystyczna rozkładu cen pralki „Electrolux EWT1266TLW”

Poniżej zaprezentowano rozkład cen dla pralki “Electrolux EWT1266TLW” w 44 sklepach.

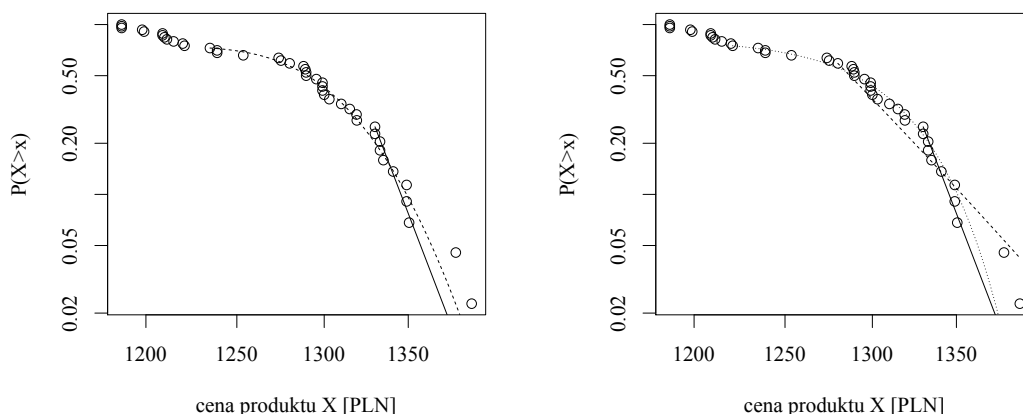


Rys. 7-12 Po lewej stronie rysunku zaprezentowano rozkład cen pralki „Electrolux EWT1266TLW” w 44 sklepach internetowych z portalu Ceneo.pl. Po prawej stronie ilustracji przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwarte kółka odpowiadają danym doświadczalnym a linia ciągła przedstawia dopasowany do danych rozkład potęgowy. Parametry rozkładu, wyniki testu Shapiro-Wilka i parametry dopasowania zamieszczono poniżej w tekście.

W celu sprawdzenia hipotezy o gaussowskości rozkładu cen pralki, przeprowadzono test Shapiro-Wilka otrzymując następujące wartości: $W = 0.94$; $p - value = 0.035$, co pozwala na odrzucenie hipotezy o gaussowskości rozkładu.

Na ilustracji Rys. 7-12 po prawej stronie przedstawiono linią ciągłą dopasowanie rozkładu potęgowego do danych doświadczalnych otrzymując następujące wartości parametrów rozkładu: $x_{min} = 1330.8 \pm 8.0$; $\alpha = 80 \pm 27$, a parametry dopasowania wynoszą: $p - value = 0.64$; $KS = 0.13$, przy wartości krytycznej testu

Kołmogorowa-Smirnowa (KS) na poziomie ufności 0.9 i liczebności próby 11, wynoszącej 0.31, czyli dopasowanie jest sensowne. Kolejnym krokiem jest sprawdzenie hipotez o alternatywnych rozkładach ciężkoogonowych.



Rys. 7-13 Na rysunku otwartymi kółkami przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN] dla pralki „Electrolux EWT1266TLW” dane uzyskane z 44 sklepów internetowych z portalu Ceneo.pl. Po lewej stronie rysunku linią przerywaną zaznaczono dopasowany do danych rozkład logarytmiczno-normalny. Po prawej stronie ilustracji widoczne jest dopasowanie rozkładów wykładniczego (linia przerywana) i rozkładu Weibulla (linia kropkowana). Parametry dopasowanych rozkładów i parametry dopasowania zamieszczono poniżej w tekście. Linia ciągła na obu ilustracjach odpowiada dopasowanemu rozkładowi potęgowemu (tak samo jak na wcześniejszym rysunku).

Na rysunku Rys. 7-13 linia ciągła odpowiada dopasowanej zależności potęgowej, linia przerywana na ilustracji po lewej stronie przedstawia dopasowaną zależność logarytmiczno-normalną o następujących parametrach rozkładu: $x_{min} = 1235 \pm 38$; $\mu = 7.176 \pm 0.015$; $\sigma = 0.0274 \pm 0.0095$, parametry dopasowania wynoszą: $p - value = 0.46$; $KS = 0.071$.

Vuong test: $test_{statistic} = 0.44$; $p - value = 0.33$. P-wartość większa od 0.1, dla testu Vuong’a wskazuje, że, mimo iż wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-normalny to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu logarytmiczno-normalnego.

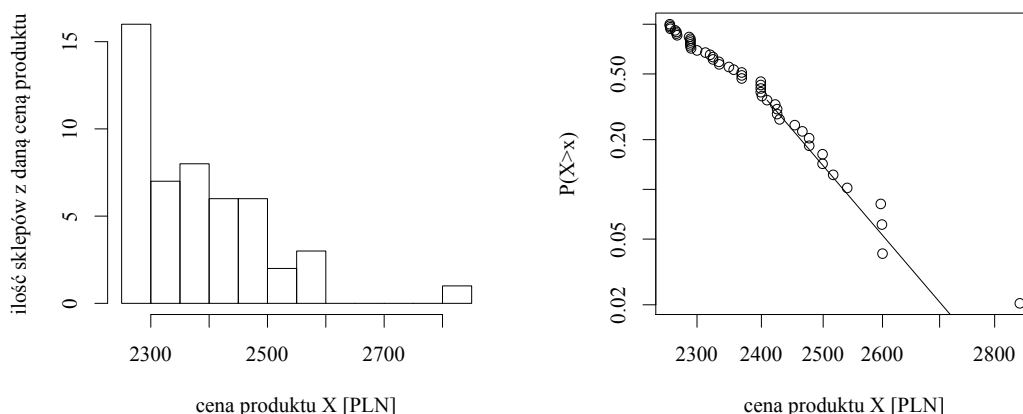
Po prawej stronie Rys. 7-13 linią przerywaną zaznaczono dopasowany rozkład wykładniczy o następujących wartościach parametrów: $x_{min} = 1280 \pm 16$; $\lambda = 0.0223 \pm 0.0037$ i wartościach dopasowania: $p - value = 0.19$; $KS = 0.14$. Vuong test: $test_{statistic} = 1.43$, $p - value = 0.076$. P-wartość mniejsza od 0.1, dla testu Vuong’a wskazuje, że przy wartości statystyki testowej większej, od 0, rozkład

potęgowej jest lepiej dopasowany niż wykładniczy i hipoteza o wykładniczym charakterze rozkładu może być odrzucona.

Po prawej stronie Rys. 7-13 linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla uzyskane parametry rozkładu są następujące: $x_{min} = 1220 \pm 13$; $k = 31 \pm 6.9$; $\lambda = 1319 \pm 12$. Wartości parametrów dopasowania wynoszą: $p - value = 0.57$; $KS = 0.08$. Vuong test: $test_{statistic} = 0.93$, $p - value = 0.18$. P-wartość większa od 0.1 dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowej jest lepiej dopasowany niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla.

7.1.7. Analiza rozkładu cen telewizora „LG 50LF652V”

Poniżej zaprezentowano rozkład cen dla telewizora „LG 50LF652V”

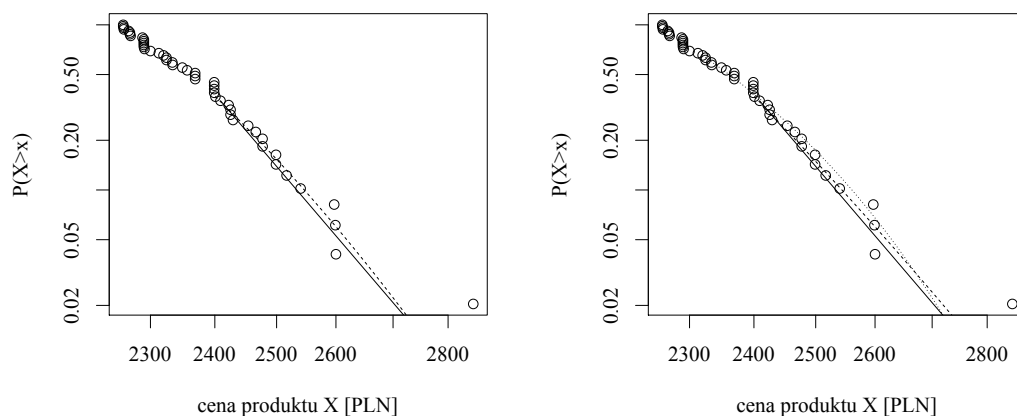


Rys. 7-14 Po lewej stronie rysunku zaprezentowano rozkład cen telewizora „LG 50LF652V” w 49 sklepach internetowych z portalu Ceneo.pl. Po prawej stronie ilustracji przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwarte kółka odpowiadają danym doświadczalnym a linia ciągła przedstawia dopasowany do danych rozkład potęgowej. Parametry rozkładu, wyniki testu Shapiro-Wilka i parametry dopasowania zamieszczono poniżej w tekście.

W celu sprawdzenia hipotezy o gaussowskości rozkładu cen telewizora „LG 50LF652V”, przeprowadzono test Shapiro-Wilka otrzymując następujące wartości: $W = 0.87$; $p - value = 5.3e - 05$, co pozwala na odrzucenie hipotezy o gaussowskości rozkładu.

Na ilustracji Rys. 7-14 po prawej stronie przedstawiono linią ciągłą dopasowanie rozkładu potęgowej do danych doświadczalnych otrzymując następujące wartości parametrów rozkładu: $x_{min} = 2409 \pm 24$; $\alpha = 25.7 \pm 5.3$, a parametry dopasowania

wynoszą: $p - value = 0.83$; $KS = 0.085$, przy wartości krytycznej testu Kołmogorowa-Smirnowa (KS) na poziomie ufności 0.9 i liczebności próby 17, wynoszącej 0.25, czyli dopasowanie jest sensowne. Kolejnym krokiem jest sprawdzenie hipotez o alternatywnych rozkładach ciężkoogonowych.



Rys. 7-15 Na ilustracji otwartymi kółkami przedstawiono dane rynkowe cen telewizora „LG 50LF652V” z 49-ciu sklepów internetowych z portalu Ceneo.pl. Po lewej stronie ilustracji przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN]. Linia ciągła przedstawia dopasowany do danych rozkład potęgowy, a linia przerywana rozkład logarytmiczno-normalny. Po prawej stronie obrazka przedstawiono dopasowanie do tych samych danych rozkładów wykładniczego (linia przerywana) i Weibulla (linia kropkowana). Parametry rozkładów, jak i dopasowania oraz wyniki testu Vuong’a omówiono poniżej w tekście.

Na rysunku Rys. 7-15 linia ciągła odpowiada dopasowanej zależności potęgowej, linia przerywana na ilustracji po lewej stronie przedstawia dopasowaną zależność logarytmiczno-normalną o następujących parametrach rozkładu: $x_{min} = 2400 \pm 16$; $\mu = 7.511 \pm 0.084$; $\sigma = 0.124 \pm 0.036$, parametry dopasowania wynoszą: $p - value = 0.86$; $KS = 0.063$.

Vuong test: $test_{statistic} = 0.18$, $p - value = 0.42$. P-wartość większa od 0.1, dla testu Vuong’a wskazuje, że, mimo iż wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-normalny to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu logarytmiczno-normalnego.

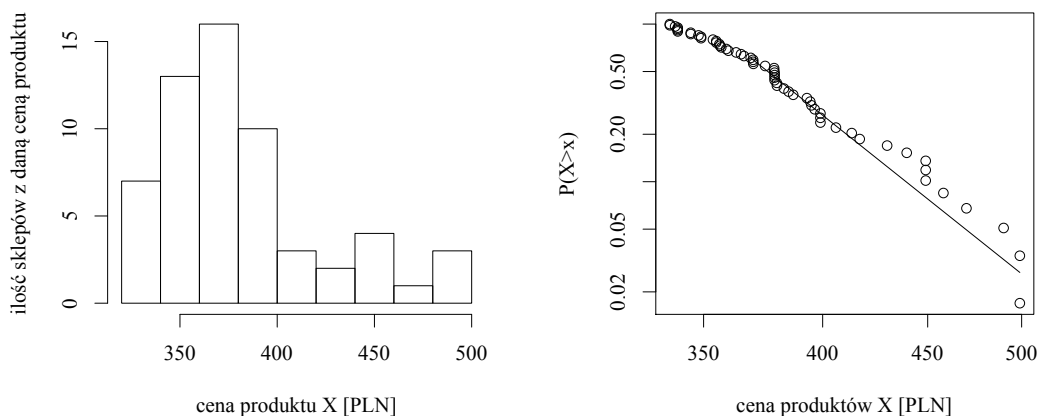
Po prawej stronie Rys. 7-15 linią przerywaną zaznaczono dopasowany rozkład wykładniczy o następujących wartościach parametrów: $x_{min} = 2401 \pm 24$; $\lambda = 0.0081 \pm 0.0016$ i wartościach dopasowania: $p - value = 0.94$; $KS = 0.064$. Vuong test: $test_{statistic} = 0.11$; $p - value = 0.46$. P-wartość większa od 0.1, dla testu Vuong’a wskazuje, że przy wartości statystyki testowej większej, od 0, rozkład

potęgowej jest lepiej dopasowany niż wykładniczy, ale ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę o wykładniczym charakterze rozkładu.

Po prawej stronie Rys. 7-15 linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla uzyskane parametry rozkładu są następujące: $x_{min} = 2324 \pm 22$; $k = 7.7 \pm 4.2$; $\lambda = 2180 \pm 240$. Wartości parametrów dopasowania wynoszą: $p - value = 0.11$; $KS = 0.10$. Vuong test: $test_{statistic} = 0.51$; $p - value = 0.30$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowej jest lepiej dopasowany, niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla.

7.1.8. Badania rozkładu cen odkurzacza „Zelmer Jupiter ZVC425HT”

Poniżej zaprezentowano rozkład cen dla odkurzacza „Zelmer Jupiter ZVC425HT”

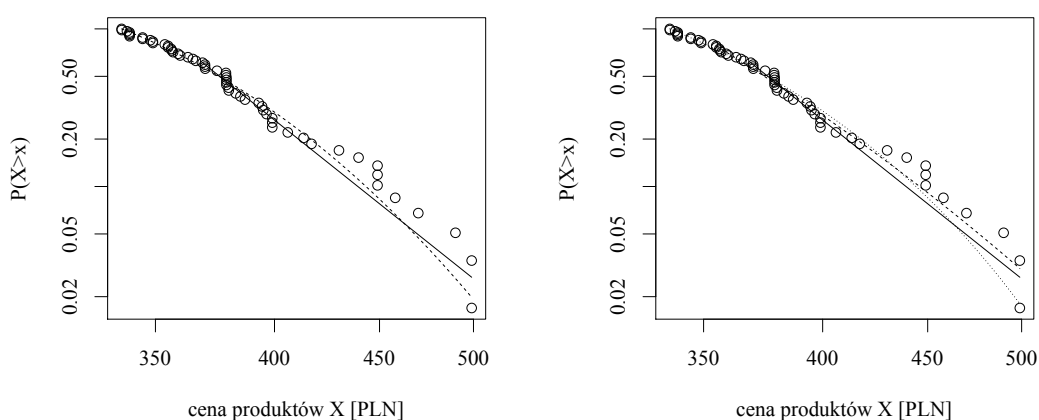


Rys. 7-16 Po lewej stronie ilustracji pokazano rozkład cen odkurzacza „Zelmer Jupiter ZVC425HT” w 59 sklepach z portalu Ceneo.pl. Po prawej stronie rysunku widać zależność $P(X > x)$ od ceny X w [PLN]. Linia ciągła przedstawia dopasowany do danych (puste kółka) rozkład potęgowej. Opis rozkładu cen umieszczono poniżej w tekście.

Hipotezę o gaussowskim rozkładzie cen odkurzacza „Zelmer Jupiter ZVC425HT” sprawdzono przy użyciu testu Shapiro-Wilka otrzymując następujące

wartości: $W = 0.88$; $p - value = 2.4e - 05$, co pozwala na odrzucenie hipotezy o gaussowskości rozkładu.

Dopasowany do danych rozkład potęgowy jest przedstawiony po prawej stronie ilustracji Rys. 7-16 linią ciągłą i ma następujące parametry: $x_{min} = 369 \pm 15$; $\alpha = 11.4 \pm 2.5$, a parametry dopasowania wynoszą: $p - value = 0.15$; $KS = 0.099$, przy wartości krytycznej testu Kołmogorowa-Smirnowa (KS) na poziomie ufności 0.9 i liczności próby 36, wynoszącej 0.20, czyli dopasowanie jest sensowne. Kolejnym krokiem jest sprawdzenie hipotez o alternatywnych rozkładach ciężkoogonowych.



Rys. 7-17 Na obu ilustracjach przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN], otwarte kółka przedstawiają dane rynkowe cen odkurzacza "Zelmer Jupiter ZVC425HT" w 59 sklepach z portalu Ceneo.pl. Linia ciągła przedstawia dopasowany do danych rozkład potęgowy (jak na wcześniejszej ilustracji). Po lewej stronie rysunku, linia przerywana odpowiada dopasowanemu do danych rozkładowi logarytmiczno-normalnemu. Po prawej stronie obrazka przedstawiono dopasowanie do tych samych danych rozkładów wykładniczego (linia przerywana) i Weibulla (linia kropkowana). Parametry rozkładów, jak i dopasowania oraz wyniki testu Vuong'a omówiono poniżej w tekście.

Na rysunku Rys. 7-17 linia ciągła odpowiada dopasowanej zależności potęgowej, linia przerywana na ilustracji po lewej stronie przedstawia dopasowaną zależność logarytmiczno-normalną o następujących parametrach rozkładu: $x_{min} = 337.0 \pm 8.7$; $\mu = 5.79 \pm 0.16$; $\sigma = 0.177 \pm 0.062$, parametry dopasowania wynoszą: $p - value = 0.25$; $KS = 0.066$.

Vuong test: $test_{statistic} = 0.32$, $p - value = 0.75$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, że, mimo iż wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-

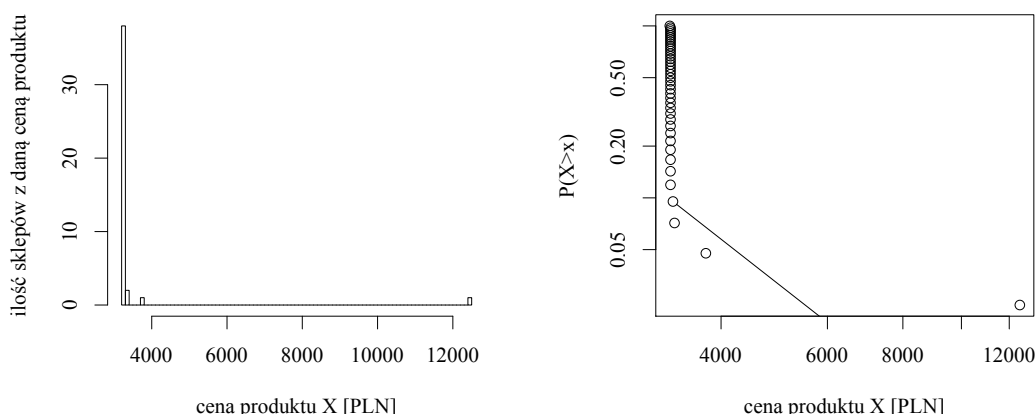
normalny to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu logarytmiczno-normalnego.

Po prawej stronie Rys. 7-17 linią przerywaną zaznaczono dopasowany rozkład wykładniczy o następujących wartościach parametrów: $x_{min} = 355 \pm 17$; $\lambda = 0.0226 \pm 0.0047$ i wartościach dopasowania: $p - value = 0.103$; $KS = 0.090$. Vuong test: $test_{statistic} = 0.15$; $p - value = 0.88$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, że przy wartości statystyki testowej większej, od 0, rozkład potęgowy jest lepiej dopasowany niż wykładniczy, ale ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę o wykładniczym charakterze rozkładu.

Po prawej stronie ilustracji Rys. 7-17 linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla uzyskane parametry rozkładu są następujące: $x_{min} = 337.0 \pm 9.2$; $k = 3.0 \pm 2.1$; $\lambda = 279 \pm 93$. Wartości parametrów dopasowania wynoszą: $p - value = 0.16$; $KS = 0.071$. Vuong test: $test_{statistic} = 0.26$; $p - value = 0.79$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany, niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla.

7.1.9. Analiza statystyczna rozkładu cen ekspresu „Nivona 767 Cafe Romatica”

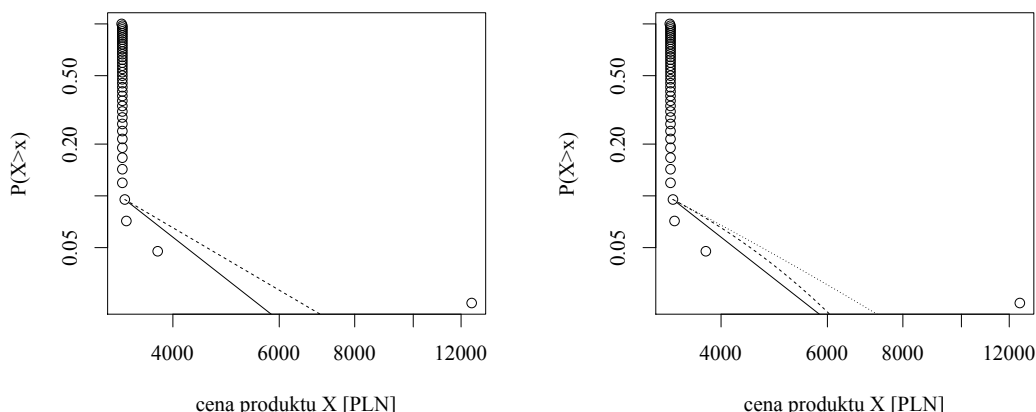
Poniżej zaprezentowano rozkład cen dla ekspresu do kawy “Nivona 767 Cafe Romatica”.



Rys. 7-18 Po lewej stronie przedstawiono histogram rozkładu cen ekspresu do kawy “Nivona 767 Cafe Romatica” w 42 sklepach internetowych z portalu Ceneo.pl. Po prawej stronie ilustracji zaprezentowano zależność $P(X > x)$ od ceny X w [PLN]. Linia ciągła przedstawia dopasowany do danych (puste kółka) rozkład potęgowy. Opis rozkładów zamieszczono poniżej w tekście.

Hipotezę o gaussowskim rozkładzie cen ekspresu do kawy “Nivona 767 Cafe Romatica” z 42 sklepów internetowych sprawdzono przy użyciu testu Shapiro-Wilka otrzymując następujące wartości: $W = 0.16$; $p - value = 3.6e - 14$, co pozwala na odrzucenie hipotezy o gaussowskim rozkładzie cen badanego ekspresu do kawy.

Dopasowany do danych rozkład potęgowy jest przedstawiony po prawej stronie ilustracji Rys. 7-18 linią ciągłą i ma następujące parametry: $x_{min} = 3330 \pm 310$; $\alpha = 3.8 \pm 5.4$, a parametry dopasowania wynoszą: $p - value = 0.88$; $KS = 0.23$, przy wartości krytycznej testu Kołmogorowa-Smirnowa (KS) na poziomie ufności 0.9 i liczebności próby 4, wynoszącej 0.56, czyli dopasowanie jest sensowne. Niemniej, biorąc pod uwagę bardzo małą ilość punktów doświadczalnych (jedynie cztery), otrzymywane wyniki powinny być traktowane z bardzo dużą rezerwą. Kolejnym krokiem jest sprawdzenie hipotez o alternatywnych rozkładach ciężkoogonowych.



Rys. 7-19 Na obu rysunkach przedstawiono rozkład $P(X > x)$ od ceny X wyrażonej w [PLN], otwarte kółka odpowiadają danym cenom ekspresu do kawy "Nivona 767 Cafe Romantica" w 42 sklepach z portalu Ceneo.pl. Linia ciągła przedstawia dopasowany do danych rozkład potęgowy (jak na wcześniejszej ilustracji). Po lewej stronie rysunku, linia przerywana odpowiada dopasowaniu do danych rozkładowi logarytmiczno-normalnemu. Po prawej stronie obrazka przedstawiono dopasowanie do tych samych danych rozkładów wykładniczego (linia przerywana) i Weibulla (linia kropkowana). Parametry rozkładów, jak i dopasowania oraz wyniki testu Vuong'a omówiono poniżej w tekście.

Linia przerywana na ilustracji Rys. 7-19 po lewej stronie przedstawia dopasowaną zależność logarytmiczno-normalną o następujących parametrach rozkładu: $x_{min} = 3330$; $\mu = -5010$; $\sigma = 49$, parametry dopasowania wynoszą: $p - value = error$; $KS = error$ gdyż nie udało się zastosować bootstrapu, ze względu na niestabilność numeryczną. Dlatego też nie udało się wyznaczyć niepewności otrzymanych parametrów rozkładu.

Vuong test: $test_{statistic} = 0.17$, $p - value = 0.86$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, że, mimo iż wartość statystyki testowej jest większa, od 0 co wskazuje, że rozkład potęgowy jest lepiej dopasowany niż logarytmiczno-normalny to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu logarytmiczno-normalnego.

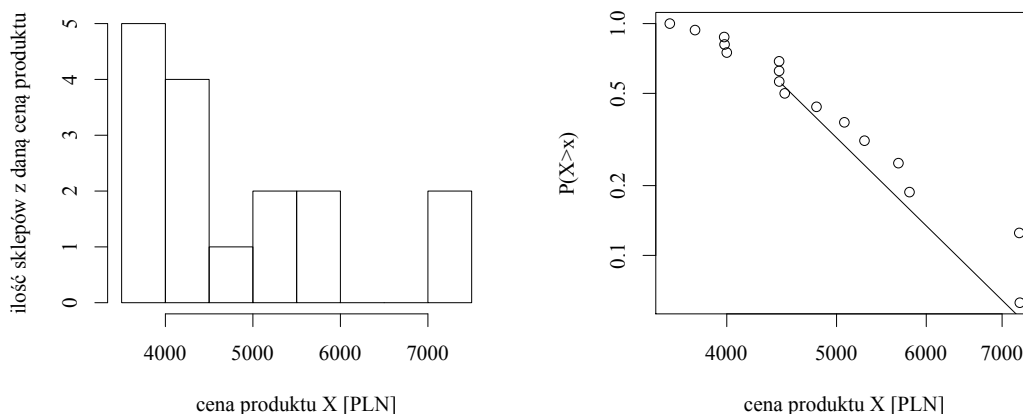
Po prawej stronie Rys. 7-19 linią przerywaną zaznaczono dopasowany rozkład wykładniczy o następujących wartościach parametrów: $x_{min} = 3330 \pm 420$; $\lambda = 0.00056 \pm 0.00047$ i wartościach dopasowania: $p - value = 0.31$; $KS = 0.28$. Vuong test: $test_{statistic} = 4.3$; $p - value = 1.7e - 05$. P-wartość mniejsza od 0.1, dla testu Vuong'a wskazuje, że przy wartości statystyki testowej większej, od 0,

rozkład potęgowy jest lepiej dopasowany niż wykładniczy i możemy odrzucić hipotezę o wykładniczym charakterze rozkładu.

Po prawej stronie ilustracji Rys. 7-19 linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla, uzyskane parametry rozkładu są następujące: $x_{min} = 3330 \pm 12$; $k = 0.2 \pm 2.5$; $\lambda = 0 \pm 1700$. Występowanie tak ogromnych rozrzutów niepewności parametrów wskazuje na niestabilność numeryczną metody bootstrap (i rzeczywiście udało się uzyskać wyniki tylko dla ilości bootstrapów $B = 6$). Wartości parametrów dopasowania wynoszą: $p - value = 0.67$; $KS = 0.29$. Vuong test: $test_{statistic} = 0.31$; $p - value = 0.76$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest większa, od 0, co wskazuje, że rozkład potęgowy jest lepiej dopasowany, niż Weibulla, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu Weibulla.

7.1.10. Badanie rozkładu cen klimatyzatora „LG G09W1 Artcool Stylist”

Poniżej zaprezentowano rozkład cen dla klimatyzatora „LG G09W1 Artcool Stylist”.

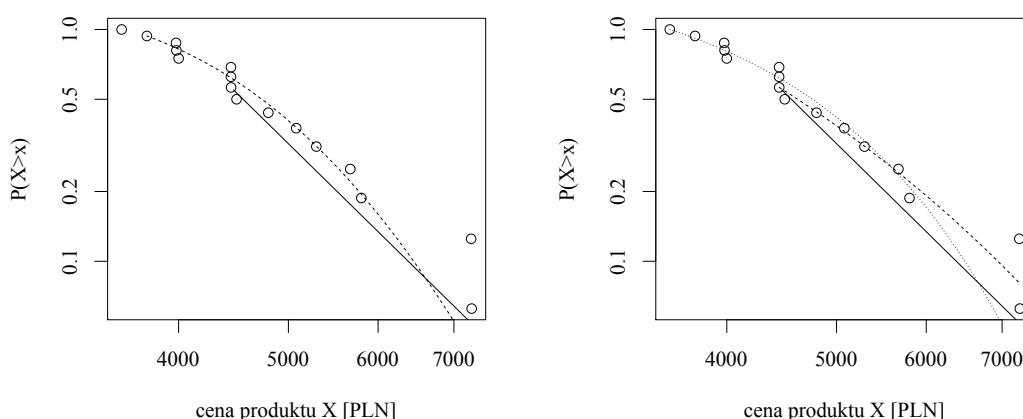


Rys. 7-20 Po lewej stronie ilustracji pokazano histogram rozkładu cen klimatyzatora „LG G09W1 Artcool Stylist” z 16 sklepów z portalu Ceneo.pl. Po prawej stronie ilustracji przedstawiono zależność $P(X > x)$ od ceny X w [PLN]. Linia ciągła odpowiada dopasowanemu do danych (puste kółka) rozkładowi potęgowemu. Wyniki analizy danych zamieszczono poniżej w tekście.

W celu weryfikacji hipotezy o gaussowskim rozkładzie cen zastosowano test Shapiro-Wilka, uzyskując następujące wartości: $W = 0.87$; $p - value = 0.030$. Otrzymana $p - value$ z testu Shapiro-Wilka jest największą otrzymaną wartością

z pośród wszystkich przebadanych rozkładów. Jako dobrą zgodność przyjmuje się p – wartość na poziomie 0.1 lub więcej. Stosunkowo duża p – wartość dla tego rozkładu, wynika z małej ilości danych (jedynie 16 sklepów) i stosunkowo niezbyt długiego ogona rozkładu w kierunku wyższych cen. Niemniej test i tak nakazuje nam odrzucenie hipotezy o gaussowskim rozkładzie cen klimatyzatora “LG G09W1 Artcool Stylist”. Gdyby towar był dostępny w większej ilości sklepów internetowych, należałoby oczekiwać dalszego gwałtownego spadku wyliczonej p – wartości.

Na rysunku Rys. 7-20 po prawej stronie umieszczono zależność $P(X > x)$ od ceny X w [PLN]. Linia ciągła przedstawia dopasowany do danych (puste kółka) rozkład potęgowy, o następujących parametrach rozkładu: $x_{min} = 4450 \pm 280$; $\alpha = 5.8 \pm 2.4$, a parametry dopasowania wynoszą: p – value = 0.63; $KS = 0.14$, wobec wartości krytycznej testu Kołmogorowa-Smirnowa (KS), na poziomie ufności 0.9 i liczebności próby 9, wynoszącej 0.39, czyli dopasowanie jest rozsądne. Kolejnym krokiem jest sprawdzenie hipotez o alternatywnych rozkładach ciężkoogonowych.



Rys. 7-21 Na ilustracji przedstawiono rozkłady $P(X > x)$ od ceny X wyrażonej w [PLN], otwarte kółka odpowiadają danym cenom klimatyzatora “LG G09W1 Artcool Stylist” w 16 sklepach z portalu Ceneo.pl. Linia ciągła przedstawia dopasowany do danych rozkład potęgowy (jak na wcześniejszym obrazku). Po lewej stronie rysunku, linia przerywana odpowiada dopasowaniu do danych rozkładowi logarytmiczno-normalnemu. Po prawej stronie obrazka przedstawiono dopasowanie do tych samych danych rozkładów wykładniczego (linia przerywana) i Weibulla (linia kropkowana). Parametry rozkładów, jak i dopasowania oraz wyniki testu Vuong’a omówiono poniżej w tekście.

Linia przerywana na Rys. 7-21 po lewej stronie przedstawia dopasowaną zależność logarytmiczno-normalną o następujących parametrach rozkładu: $x_{min} =$

3750 ± 460 ; $\mu = 8.37 \pm 0.19$; $\sigma = 0.28 \pm 0.11$, parametry dopasowania wynoszą: $p - value = 0.36$; $KS = 0.10$, a tym samym dopasowanie jest sensowne.

Vuong test: $test_{statistic} = -0.17$, $p - value = 0.86$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, że, mimo iż wartość statystyki testowej jest mniejsza, od 0 co wskazuje, że rozkład logarytmiczno-normalny jest lepiej dopasowany niż rozkład potęgowy, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu potęgowego.

Po prawej stronie ilustracji Rys. 7-21 linią przerywaną zaznaczono dopasowany rozkład wykładniczy o następujących wartościach parametrów: $x_{min} = 4450 \pm 960$; $\lambda = 0.00069 \pm 0.00021$ i wartościach dopasowania: $p - value = 0.93$; $KS = 0.079$.

Vuong test: $test_{statistic} = 0.058$; $p - value = 0.95$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, że przy wartości statystyki testowej większej, od 0, rozkład potęgowy jest lepiej dopasowany niż wykładniczy, jednakże nie możemy odrzucić hipotezy o wykładniczym charakterze rozkładu, ze względu na niewystarczającą ilość danych doświadczalnych.

Po prawej stronie ilustracji Rys. 7-21 linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla, uzyskane parametry rozkładu są następujące: $x_{min} = 3560 \pm 380$; $k = 2.6 \pm 2.9$; $\lambda = 4300 \pm 1300$. Wartości parametrów dopasowania wynoszą: $p - value = 0.42$; $KS = 0.097$.

Vuong test: $test_{statistic} = -0.19$; $p - value = 0.85$. P-wartość większa od 0.1, dla testu Vuong'a wskazuje, iż, mimo że wartość statystyki testowej jest mniejsza, od 0, co wskazuje, że rozkład Weibulla jest lepiej dopasowany, niż rozkład potęgowy, to ilość danych doświadczalnych jest niewystarczająca by odrzucić hipotezę rozkładu potęgowego.

7.1.11. Parametryczny opis funkcji gęstości prawdopodobieństwa rozkładu cen produktu

Alternatywą dla opisu cen produktów przy pomocy rozkładu potęgowego mogłaby być również funkcja zaproponowana przez autora w pracy (Zachara and Piskor-Ignatowicz 2010) mająca postać:

$$f = a \left(\frac{1}{1 + \exp\left(\frac{s-s_1}{\sigma_1}\right)} + \frac{1}{1 + \exp\left(\frac{s-s_2}{\sigma_2}\right)} - 1 \right) \quad (7-5)$$

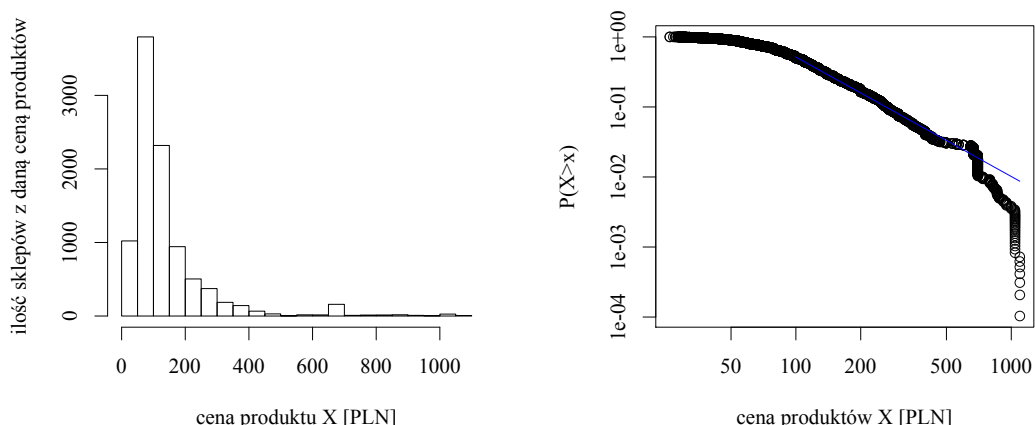
Gdzie a - parametr skalujący S - cena S_1 , S_2 - parametry decydujące o szerokości wierzchołka funkcji, σ_1 , σ_2 - parametry odpowiadające za nachylenie zboczy funkcji. Normalizując pole pod wykresem zaprezentowanej funkcji począwszy od $S = S_0$ tak, że $f(S_0) = 0$ otrzymujemy funkcję gęstości prawdopodobieństwa. Aby opisywała ona rozkłady cen, należy znormalizować również wykresy z rzeczywistymi danymi tak, aby pole powierzchni pod wykresem wynosiło 1.

Rozkład cen dla każdego produktu ma opisaną powyżej postać o ile produkt był dostarczony w tym samym czasie (w tych samych warunkach rynkowych). W przeciwnym razie na wykresie pojawią się charakterystyczne zęby/pofalowania - każdy, dające się opisać, jako superpozycja z nową funkcją f .

Ciekawą własnością zaobserwowaną dla omówionego rozkładu jest fakt, iż, ceny całej grupy produktów, np. ceny wszystkich czajników też mają postać funkcji f . Jednakże, funkcja ta mimo swej dużej uniwersalności ma istotne wady, posiada aż 5 parametrów i autorowi pomimo prób nie udało się wyprowadzić jej postaci z jakiegoś konkretnego modelu fizycznego, a więc jest to przykład matematycznego modelu parametrycznego, potencjalnie bez żadnej siły predykcji. Co gorsza, dla rozkładów posiadających zęby/pofalowania, aby opisać taki rozkład musielibyśmy robić superpozycję kilku funkcji f , co prowadziłoby do bardzo szybkiego wzrostu liczby parametrów naszego modelu. Dlatego też nie będę zajmował się dalej opisem parametrycznym przy użyciu funkcji f .

7.1.12. Analiza statystyczna rozkładu cen w ramach jednej kategorii produktu

Poniżej zaprezentowano rozkład cen dla całej kategorii czajniki to jest pokazano rozkład cen wszystkich czajników ze wszystkich sklepów internetowych skatalogowanych na portalu ceneo.

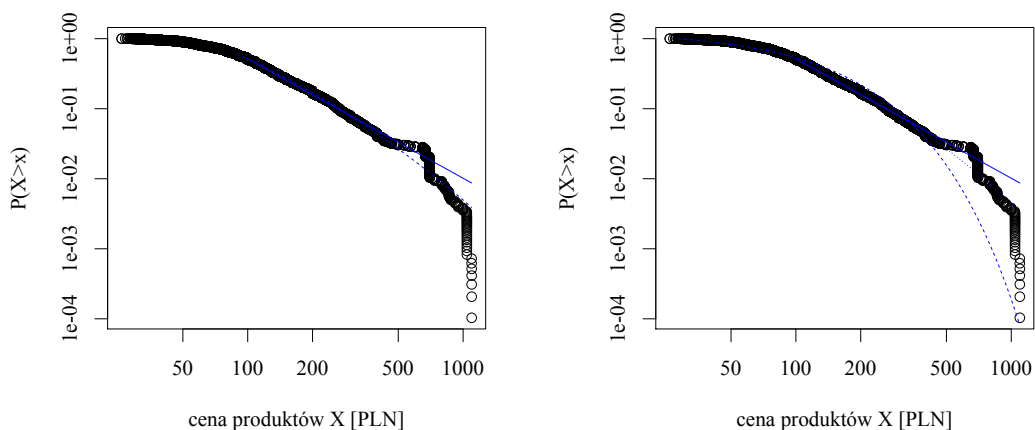


Rys. 7-22 Po lewej stronie rysunku zaprezentowano rozkład cen całej kategorii czajniki elektryczne (wszystkich czajników elektrycznych) we wszystkich sklepach internetowych dostępnych na portalu Ceneo.pl. Po prawej stronie ilustracji umieszczono zależność $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne. Linia ciągła odpowiada dopasowanej zależności potęgowej. Opis parametrów rozkładu i dopasowania umieszczono poniżej w tekście.

Po lewej stronie ilustracji Rys. 7-22 przedstawiono rozkład cen całej kategorii czajniki elektryczne (wszystkich czajników elektrycznych) we wszystkich sklepach internetowych dostępnych na portalu Ceneo.pl. Otrzymana dla reprezentatywnej próbki danych (4999 cen) z całej populacji cen czajników (9670 cen) wartość testu Shapiro-Wilka W oraz p – wartość wynoszą $W = 0.64$; p – value = $2.2e - 16$, co pozwala na odrzucenie hipotezy o gaussowskim rozkładzie cen całej kategorii czajniki. Na marginesie warto tu dodać uwagę, iż konieczność wybrania reprezentatywnej próby zamiast przeprowadzić test Shapiro-Wilka dla całej populacji cen czajników wynika z ograniczenia użytego algorytmu do 5000 rozmiaru analizowanej nim populacji (R Core Team 2017).

Po prawej stronie rysunku Rys. 7-22 ukazano zależność $P(X > x)$ od ceny X w [PLN], wraz z dopasowaną zależnością potęgową o następujących parametrach rozkładu: $x_{min} = 96.9 \pm 2.4$; $\alpha = 2.704 \pm 0.031$, oraz następujących parametrach dopasowania: p – value = 0; $KS = 0.038$. p – wartość = 0, wskazuje, że należy odrzucić hipotezę rozkładu potęgowego, co jest zgodne z otrzymaną wartością

statystyki Kołmogorowa-Smirnowa KS większą niż wartość krytyczna testu na poziomie ufności $Cl = 0.9$, dla rozmiaru populacji rzędu $N \sim 10^4$ wynosząca 0.012.



Rys. 7-23 Na ilustracjach przedstawiono zależność $P(X > x)$ od ceny X wyrażonej w [PLN]. Otwartymi kółkami zaznaczono dane doświadczalne cen całej kategorii czajniki. Linia ciągła odpowiada dopasowanej zależności potęgowej. Po lewej stronie linią przerywaną zaznaczono dopasowany do danych rozkład logarytmiczno-normalny. Po prawej stronie rysunku linia przerywana odpowiada dopasowanej zależności wykładniczej, a linia kropkowana dopasowanemu do danych rozkładowi Weibulla. Parametry rozkładów i ich dopasowań omówiono poniżej w tekście.

Na ilustracji Rys. 7-23 po lewej stronie linią przerywaną przedstawiono rozkład logarytmiczno-normalny o następujących parametrach: $x_{min} = 89.0 \pm 4.0$; $\mu = 2.67 \pm 0.38$; $\sigma = 1.327 \pm 0.085$. Parametry dopasowania wynoszą: $p - value = 0$; $KS = 0.025$, a więc odrzucamy hipotezę o rozkładzie logarytmiczno-normalnym. Warto w tym miejscu zwrócić uwagę na charakterystyczny „schodek” na wykresie $P(X > x)$, znajdujący się pomiędzy ceną 462 PLN, a 705 PLN, odpowiadający grupie czajników luksusowych. Podobne intrygujące zależności pojawiają się również dla innych grup produktów, jednakże badanie ich będzie przedmiotem dalszych prac i nie mieści się w ramach obecnego projektu.

W związku z faktem, odrzucenia zarówno rozkładu potęgowego jak i logarytmiczno-normalnego nie ma sensu stosować testu Vuong’a, gdyż odpowiedź na pytanie, który z niepasujących rozkładów pasuje lepiej? - Jest nieinteresująca.

Kolejnym dopasowanym rozkładem na ilustracji Rys. 7-23 po prawej stronie ilustracji jest rozkład wykładniczy przedstawiony linią przerywaną, ma on następujące parametry: $x_{min} = 28.3 \pm 5.7$; $\lambda = 0.00882 \pm 0.00012$ i wartościach dopasowania: $p - value = 0$; $KS = 0.082$.

Na ilustracji Rys. 7-23 po prawej stronie linią kropkowaną przedstawiono dopasowanie rozkładu Weibulla uzyskane parametry rozkładu są następujące: $x_{min} = 89.0 \pm 1.8$; $k = 0.294 \pm 0.028$; $\lambda = 0.45 \pm 0.22$. Wartości parametrów dopasowania wynoszą: $p - value = 0$; $KS = 0.026$. Tak, więc również rozkład Weibulla został odrzucony.

7.2. Wnioski z przeprowadzonej analizy rozkładów cen

W rozdziale tym opisano metodykę przeprowadzonych badań eksperymentalnych mających na celu przeanalizowanie rozkładów rzeczywistych cen w sklepach internetowych oraz przedstawiono rezultaty badań dla poszczególnych, losowo wybranych dziesięciu produktów które są dostępne w więcej niż szesnastu różnych sklepach internetowych skatalogowanych na portalu Ceneo.

Zastosowana metodyka badań stanowi powszechnie uznaną i zalecaną metodykę analizy rozkładów statystycznych. Logika badania statystycznego rozkładów cen produktów jest następująca.

1. Testem Shapiro-Wilka sprawdzono czy dane mają rozkład Gaussa. We wszystkich przypadkach nastąpiło odrzucenie hipotezy o Gaussowskim rozkładzie cen.
2. W związku z obserwowaną asymetrią rozkładów cen w stronę wyższych wartości, kolejno przetestowano hipotezy:
 - rozkład jest potęgowy,
 - rozkład jest logarytmiczno-normalny,
 - rozkład jest wykładniczy,
 - rozkład jest rozkładem Weibulla
3. Zastosowano test Vuong'a pozwalający określić czy posiadana jest wystarczająca ilość danych, by można wykazać, który z rozkładów lepiej opisuje dane.

Uzyskane rezultaty badań i ich analiza nie tylko nie potwierdzają ale wręcz stoją w sprzeczności z modelami ekonomicznymi, postulującymi Gaussowski rozkład cen, a tym bardziej z koncepcją jednej ceny. Ponadto wykazano, że nie można odrzucić

hipotezy o ciężkoogonowym rozkładzie cen. **Fakt ten potwierdza część tezy postawionej w rozdziale 1**

Należy zwrócić uwagę, że jedną z cech charakterystycznych rozkładów cen są bardzo duże rozrzuty cen i pewna dynamika zmian, która jednak nie jest przedmiotem niniejszych badań. Dynamika zmian cen produktów w sklepach internetowych jest o wiele wolniejsza niż w wypadku cen akcji. Główne różnice przejawiają się na znacząco większej czułości cen w sklepach internetowych na zmiany kursów walut niż w przypadku akcji, przy znacząco mniejszej czułości na wskaźniki makroekonomiczne w skali czasowej na poziomie dni.

Skala czasowa zmian cen w sklepach internetowych mieści się w przedziale, od godzin do dni. Podczas gdy zmiany cen akcji następują w przedziałach od nanosekund (handel wysokiej częstotliwości ang. High Frequency Trading HFT) do godzin dla mało płynnych walorów.

Obserwowane ceny można rozumieć, jako swego rodzaju stop klatka z filmu rejestrującego sytuacje rynkową. Rozkład cen w sklepach internetowych stanowi komplementarną daną rynkową w stosunku do zmian cen akcji.

Jako że analizowano rozkłady cen dla losowo wybranych produktów nasuwa się przypuszczenie, iż obserwowane rozkłady cen mają uniwersalny charakter i są wynikiem jakiejś permanentnej właściwości samego układu. Takie też sugestie wynikają z wcześniej prowadzonych badań (Zachara and Piskor-Ignatowicz 2010).

8. Rozkłady cen produktów generowane przez model

W rozdziale tym przedstawiono wyniki symulacji rozkładów cen generowanych z użyciem prezentowanego w pracy modelu Piskor-Ignatowicza-Zachary (Zachara and Piskor-Ignatowicz 2016), oraz przedstawiono rezultaty porównania otrzymanych z modelu rozkładów cen z rzeczywistymi rozkładami cen odpowiednich produktów zarejestrowanymi w sklepach skatalogowanych na portalu Ceneo.

Do modelowania procesów zachodzących w układach złożonych jak przykładowo procesy kształtowania cen w układach socjo-ekonomicznych, najbardziej adekwatne są systemy wieloagentowe. Układy te nie posiadają wcześniej już wspomnianych ograniczeń typowych dla przybliżenia średniego pola. Zatem nie zachodzi obawa, iż niewielka acz dobrze dostosowana do aktualnego stanu układu fluktuacja zostanie przez aparat matematyczny stłumiona.

Kolejnym istotnym czynnikiem jest łatwość implementacji, co ma ogromne znaczenie przy iteracyjnym podejściu do modelowania układu. W przeciwieństwie do prób opisu analitycznego układów złożonych (przy pomocy układów równań różniczkowych lub całkowo-różniczkowych (niestety jak dotychczas nieskutecznych)), zastosowanie systemów wieloagentowych pozwala na szybkie otrzymanie rezultatów.

W systemach wieloagentowych w naturalny sposób opis realizowany jest przy pomocy nisko poziomowych oddziaływaniach pomiędzy agentami, tak jak to ma miejsce w rzeczywistej sytuacji, gdzie oddziaływania zachodzą między elementami układu. Większość struktur pojawiających się w rzeczywistych układach pochodzi z niskopoziomowych oddziaływań element-element. Tak, więc, pomiędzy realnym światem a modelem istnieje pełna zgodność w tym aspekcie. Jedynym problemem

występującym w modelowaniu wieloagentowym jest brak danych, jaką tak naprawdę postać mają oddziaływania element-element w rzeczywistych układach złożonych.

8.1. Badanie zgodności dwóch rozkładów

Do porównania otrzymanych wyników symulacji z modelu Piskor-Ignatowicza-Zachary z rozkładami rzeczywistymi zastosowano test Kołmogorowa-Smirnowa (KS) w wersji dla dwóch prób (ang. two-sample Kolmogorov–Smirnov test). Konstrukcja testu jest następująca: dane z obydwu badanych populacji o liczebnościach odpowiednio N_{dane} , $N_{symulacja}$ porządkujemy w dwa szeregi niemalejące, a następnie tworzymy odpowiadające im dystrybuanty doświadczalne, przypisując kolejnym wartościom pomiarowym prawdopodobieństwa $\frac{1}{N_{dane}}, \frac{2}{N_{dane}}, \dots, \frac{N_{dane}}{N_{dane}}$ oraz dla danych z symulacji $\frac{1}{N_{symulacja}}, \frac{2}{N_{symulacja}}, \dots, \frac{N_{symulacja}}{N_{symulacja}}$. Wartość statystyki testowej D stanowi maksimum odległości między oboma dystrybuantami doświadczalnymi odpowiadającymi odpowiednio danym rzeczywistym i wynikiem symulacji:

$$D = \max_x (F^{dane}(x) - F^{symulacja}(x))$$

gdzie $F^{dane}(x)$, $F^{symulacja}(x)$ to odpowiednio dystrybuanty doświadczalne danych rzeczywistych cen ze sklepów internetowych i danych cen wysymulowanych.

Jeżeli wartość statystyki Kołmogorowa-Smirnowa D , jest mniejsza niż wartość krytyczna testu Kołmogorowa-Smirnowa dla danych liczebności na zadanym poziomie istotności, to nie ma podstaw twierdzić, iż próbki pochodzą z różnych populacji (innymi słowy oba badane rozkłady są takie same).

W przypadku gdy liczebność obu badanych populacji $N_{dane}, N_{symulacja} > 35$, to można stosować wzór aproksymacyjny (Darling 1957), na wartość krytyczną testu Kołmogorowa-Smirnowa oznaczaną jako $D_{critical}(N_1, N_2, \alpha)$ i mającą postać:

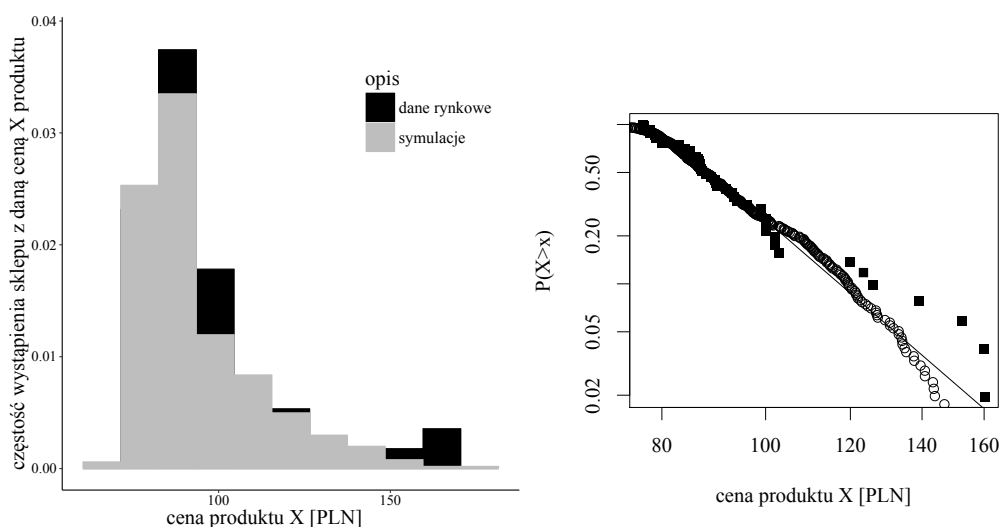
$$D_{critical}(N_1, N_2, \alpha) = c(\alpha) \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

gdzie N_1 to liczebność pierwszej z porównywanych populacji, N_2 to liczebność drugiej, α poziom istotności, $c(\alpha)$ współczynnik zależny od poziomu istotności.

Przy opisie kolejnych eksperymentów podane będą dwie wartości: przyjęta wartość stanowiąca cenę bazową produktu oraz numer cyklu symulacji, w którym

generowany przez model rozkład cen produktu najlepiej odpowiada rozkładowi cen obserwowanemu w sklepach internetowych obserwowanemu w sklepach internetowych

8.2. Porównanie danych i wyników symulacji dla czajnika „Philips HD 4646/00 Biały”



Rys. 8-1 Po lewej stronie rysunku przedstawiono histogram danych rynkowych cen czajnika “Philips HD 4646/00 Biały” (czarny kolor) wraz z nałożonym rozkładem cen pochodzących z symulacji (szary kolor). Po prawej stronie ilustracji umieszczono rozkłady $P(X > x)$ od ceny X wyrażonej w [PLN], czarne kwadraty odpowiadają danym doświadczenia, wraz z dopasowanym do nich rozkładem potęgowym (linia ciągła), otwarte kółka to wyniki symulacji. Opis dopasowania umieszczono poniżej w tekście.

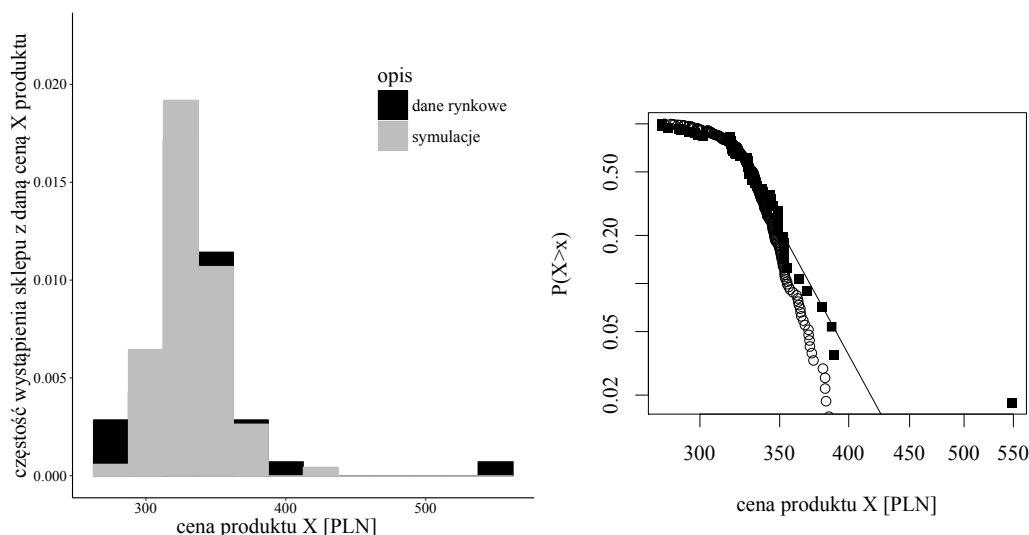
Powyżej na ilustracji Rys. 8-1 po lewej stronie przedstawiono dwa naniesione na siebie znormalizowane histogramy. Kolorem czarnym przedstawiono rynkowe ceny czajnika “Philips HD 4646/00 Biały” z 51 sklepów internetowych, a kolorem szarym wyniki symulacji, ilustrujące ceny dla 456 różnych sprzedawców. Ponieważ mamy możliwość znormalizowania obu histogramów, w symulacjach nie musimy się troszczyć by otrzymana ilość sprzedawców była równa ilości sprzedawców w danych rynkowych. A większa ilość sprzedawców uzyskana w symulacjach skutkuje mniejszymi fluktuacjami statystycznymi otrzymywanego rozkładu cen, tym samym zmniejszając szansę, iż otrzymany rozkład jest dziełem przypadku. Jak można zauważyć otrzymano dobrą zgodność pomiędzy wynikami symulacji a danymi rzeczywistymi. Jednakże wizualna ocena zgodności, jako dalece subiektywna jest mało przydatna. W celu jakościowej oceny zgodności symulacji z danymi

rzeczywistymi użyto testu Kołmogorowa-Smirnowa w wersji służącej do porównania dwóch prób (skrótowe przypomnienie teorii zamieszczono w podrozdziale 8.1).

Przyjęto jako cenę bazową 65 PLN. Cykl symulacji 7000. Otrzymano następujące wartości statystyk: Kołmogorowa-Smirnowa $D = 0.091$; p – wartość = 0.84. Wartość krytyczna testu Kołmogorowa-Smirnowa dla liczebności prób $N_{dane} = 51$; $N_{symulacje} = 456$; na poziomie istotności $\alpha = 0.1$; wynosi: $D_{critical}(51, 456, 0.1) = 0.18$.

Ponieważ $D < D_{critical}$ nie ma podstaw odrzucić hipotezy zerowej głoszącej, że oba badane rozkłady pochodzą z tej samej populacji na poziomie istotności 0.1.

8.3. Porównanie rozkładu cen monitora „BenQ 18,5 GL955A” z rezultatem symulacji



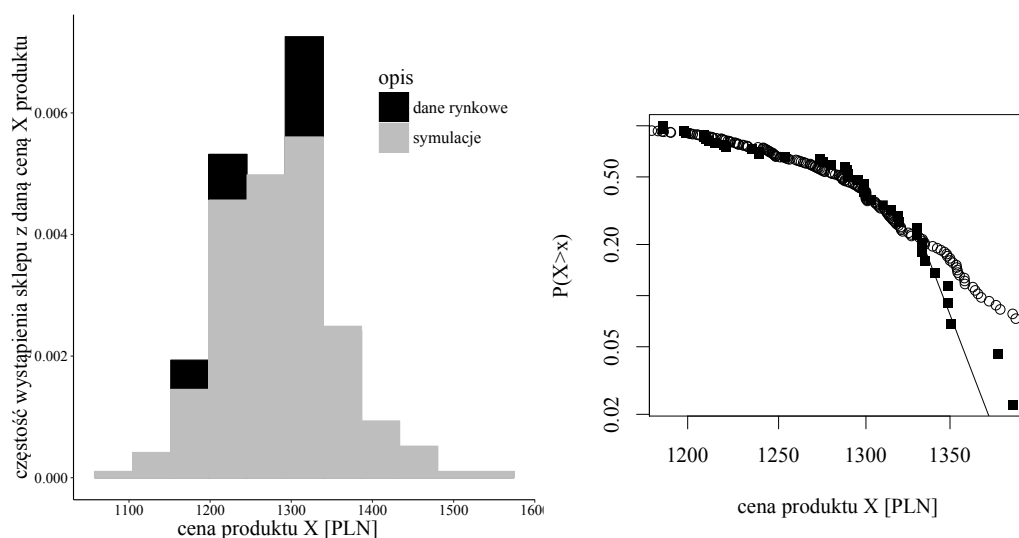
Rys. 8-2 Po lewej stronie ilustracji przedstawiono znormalizowane histogramy cen rynkowych monitora „BenQ 18,5 GL955A” (czarny kolor) wraz z nałożonym rozkładem cen pochodzących z symulacji (szary kolor). Po prawej stronie obrazka umieszczono rozkłady $P(X > x)$ od ceny X wyrażonej w [PLN], czarne kwadraty odpowiadają danym rynkowym, wraz z dopasowanym do nich rozkładem potęgowym (linia ciągła), otwarte kółka to wyniki symulacji. Opis dopasowania umieszczono poniżej w tekście.

Przyjęto jako cenę bazową 257 PLN. Cykl symulacji 3150. PLN. Otrzymano następujące wartości statystyk: Kołmogorowa-Smirnowa $D = 0.11$; p – wartość = 0.61. Wartość krytyczna testu Kołmogorowa-Smirnowa dla

liczebności prób $N_{dane} = 56$; $N_{symulacje} = 270$; na poziomie istotności $\alpha = 0.1$; wynosi: $D_{critical}(56,270,0.1) = 0.18$.

Ponieważ $D < D_{critical}(56,270,0.1)$, to na poziomie istotności $\alpha = 0.1$ nie ma podstaw odrzucić hipotezy, że oba rozkłady pochodzą z tej samej populacji.

8.4. Porównanie rozkładu cen pralki „Electrolux EWT1266TLW” z rezultatem symulacji



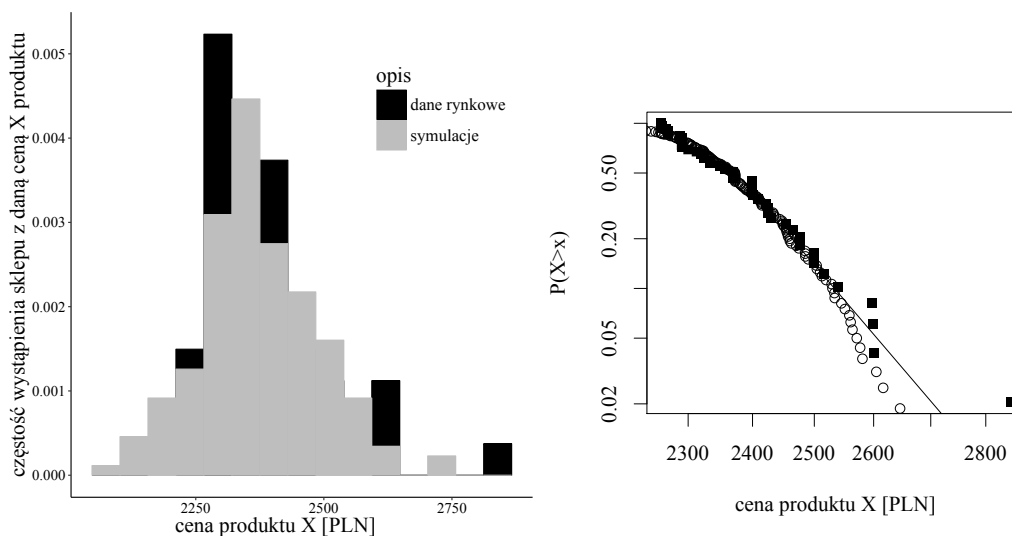
Rys. 8-3 Po lewej stronie ilustracji przedstawiono znormalizowane histogramy cen pralki „Electrolux EWT1266TLW” w 44 sklepach internetowych z portalu Ceneo.pl. (czarny kolor) wraz z nałożonym rozkładem cen pochodzących z symulacji (szary kolor). Po prawej stronie obrazka umieszczono rozkłady $P(X > x)$ od ceny X wyrażonej w [PLN], czarne kwadraty odpowiadają danym rynkowym, wraz z dopasowanym do nich rozkładem potęgowym (linia ciągła), otwarte kółka to wyniki symulacji. Opis dopasowania umieszczono poniżej w tekście.

Cykl symulacji 2050, cena bazowa 1000 PLN. Otrzymano następujące wartości statystyk: Kołmogorowa-Smirnowa $D = 0.11$; p – wartość = 0.77. Wartość krytyczna testu Kołmogorowa-Smirnowa dla liczebności prób $N_{dane} = 44$; $N_{symulacje} = 205$; na poziomie istotności $\alpha = 0.1$; wynosi $D_{critical}(44,205,0.1) = 0.20$.

Ponieważ $D < D_{critical}$ nie ma podstaw odrzucić hipotezy o zerowej na poziomie istotności 0.1. Otrzymana p – wartość jest większa od 0.1, co wskazuje, że

nie mamy podstaw odrzucić hipotezy, iż obie populacje pochodzą od jednego rozkładu prawdopodobieństwa na poziomie istotności 0.1.

8.5. Porównanie rozkładu cen telewizora „LG 50LF652V” z rozkładem cen z symulacji

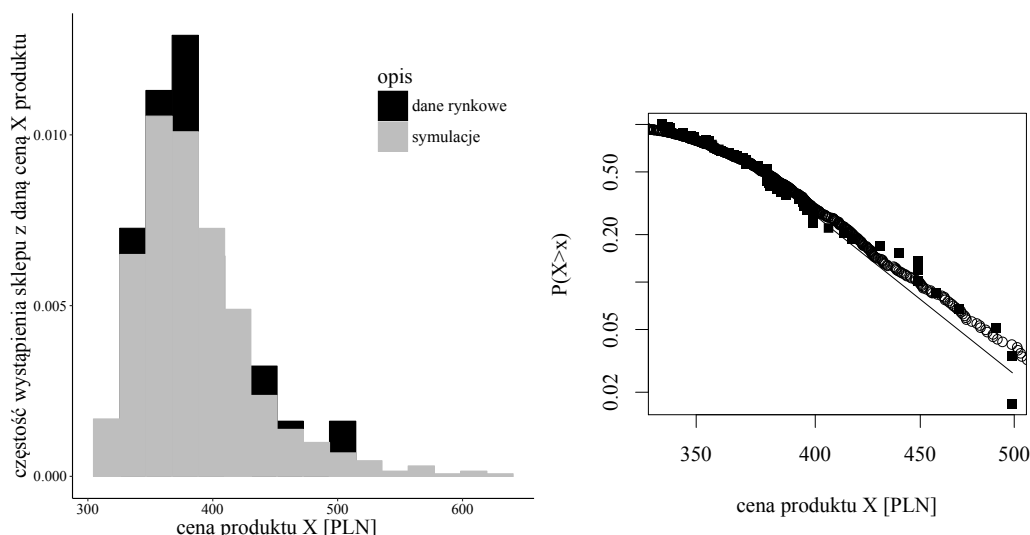


Rys. 8-4 Po lewej stronie ilustracji przedstawiono znormalizowane histogramy cen telewizora „LG 50LF652V” w 49 sklepach internetowych z portalu Ceneo.pl. (czarny kolor) wraz z nałożonym rozkładem cen pochodzących z symulacji (szary kolor). Po prawej stronie obrazka umieszczono rozkłady $P(X > x)$ od ceny X wyrażonej w [PLN], czarne kwadraty odpowiadają danym rynkowym, wraz z dopasowanym do nich rozkładem potęgowym (linia ciągła), otwarte kółka to wyniki symulacji. Opis dopasowania umieszczono poniżej w tekście

Cykl symulacji 1550, cena bazowa 1825 PLN. Otrzymano następujące wartości statystyk: Kołmogorowa-Smirnowa $D = 0.13$; p – wartość = 0.54. Wartość krytyczna testu Kołmogorowa-Smirnowa dla liczebności prób $N_{dane} = 49$; $N_{symulacje} = 160$; na poziomie istotności $\alpha = 0.1$; wynosi $D_{critical}(49,160,0.1) = 0.18$

Ponieważ $D < D_{critical}$ nie ma podstaw odrzucić hipotezy zerowej na poziomie istotności 0.1. Otrzymana p – wartość jest większa od 0.1, co wskazuje, że nie mamy podstaw odrzucić hipotezy, iż obie populacje pochodzą od jednego rozkładu prawdopodobieństwa na poziomie istotności 0.1.

8.6. Porównanie rozkładu cen odkurzacza „Zelmer Jupiter ZVC425HT” z rozkładem cen uzyskanym z symulacji

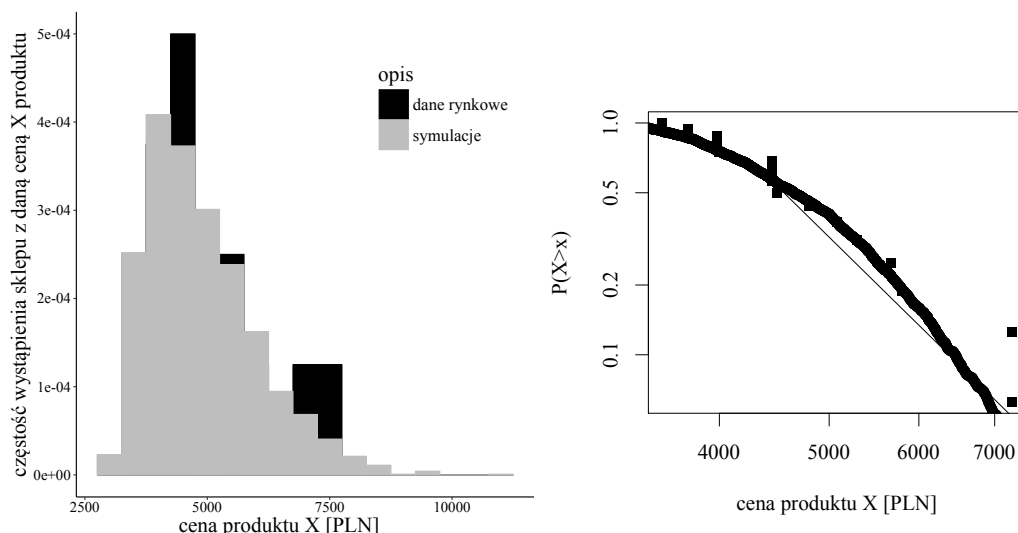


Rys. 8-5 Po lewej stronie ilustracji przedstawiono znormalizowane histogramy odkurzacza „Zelmer Jupiter ZVC425HT” w 59 sklepach internetowych z portalu Ceneo.pl. (czarny kolor) wraz z nałożonym rozkładem cen pochodzących z symulacji (szary kolor). Po prawej stronie obrazka umieszczono rozkłady $P(X > x)$ od ceny x wyrażonej w [PLN], czarne kwadraty odpowiadają danym rynkowym, wraz z dopasowanym do nich rozkładem potęgowym (linia ciągła), otwarte kółka to wyniki symulacji. Opis dopasowania umieszczono poniżej w tekście.

Przyjęto cenę bazową jako 300 PLN. Cykl symulacji 9150. Otrzymano następujące wartości statystyk: Kołmogorowa-Smirnowa $D = 0.092$; p – wartość = 0.76. Wartość krytyczna testu Kołmogorowa-Smirnowa dla liczebności prób $N_{dane} = 59$; $N_{symulacje} = 623$; na poziomie istotności $\alpha = 0.1$; wynosi $D_{critical}(59,623,0.1) = 0.146$

Ponieważ $D < D_{critical}$ nie ma podstaw odrzucić hipotezy zerowej na poziomie istotności 0.1. Otrzymana p – wartość jest większa od 0.1, co wskazuje, że nie mamy podstaw odrzucić hipotezy, iż obie populacje pochodzą od jednego rozkładu prawdopodobieństwa na poziomie istotności 0.1.

8.7. Porównanie rozkładu cen klimatyzatora „LG G09W1 Artcool Stylist” z wynikami symulacji



Rys. 8-6 Po lewej stronie ilustracji przedstawiono znormalizowane histogramy klimatyzatora "LG G09W1 Artcool Stylist" z 16 sklepów z portalu Ceneo.pl. (czarny kolor) wraz z nałożonym rozkładem cen pochodzących z symulacji (szary kolor). Po prawej stronie obrazka umieszczono rozkłady $P(X > x)$ od ceny X wyrażonej w [PLN], czarne kwadraty odpowiadają danym rynkowym, wraz z dopasowanym do nich rozkładem potęgowym (linia ciągła), otwarte kółka to wyniki symulacji. Opis dopasowania umieszczono poniżej w tekście.

Przyjęto jako cenę bazową 3050 PLN. Cykl symulacji 40000. PLN. Otrzymano następujące wartości statystyk: Kołmogorowa-Smirnowa $D = 0.12$; p – wartość = 0.98. Wartość krytyczna testu Kołmogorowa-Smirnowa dla liczebności prób $N_{dane} = 16$; $N_{symulacje} = 2365$; na poziomie istotności $\alpha = 0.1$; wynosi $D_{critical}(16, 2365, 0.1) = 0.27$

Ponieważ $D < D_{critical}$ nie ma podstaw odrzucić hipotezy zerowej na poziomie istotności 0.1. Otrzymana p – wartość jest większa od 0.1, co wskazuje, że nie mamy podstaw odrzucić hipotezy, iż obie populacje pochodzą od jednego rozkładu prawdopodobieństwa na poziomie istotności 0.1.

8.8. Podsumowanie otrzymanych rezultatów

W przypadku sześciu przebadanych rozkładów cen produktów z pośród dziesięciu losowo wybranych nie stwierdzono na poziomie istotności $\alpha = 0.1$ różnicy pomiędzy rozkładami cen generowanymi z modelu Piskor-Ignatowicza-Zachary a obserwowanymi w sklepach internetowych skatalogowanych na portalu Ceneo.pl. W pozostałych czterech przypadkach na wykresach $P(X > x)$ od ceny X wyrażonej w [PLN], widać charakterystyczne „przerwy” lub „uskoki”, których prezentowany w pracy model wydaje się iż nie może opisać. Na chwilę obecną pozostaje otwartym

pytanie o pochodzenie tych struktur? Autor rozważa dwa możliwe jego zdaniem scenariusze:

- albo struktury te są wynikiem fluktuacji – wówczas należy się spodziewać, że podczas dłuższej obserwacji rozkładów cen danego produktu będą one zanikać (w skalach czasowych znacznie krótszych niż czas dostępności produktu na rynku i pojawiać się na wykresach $P(X > x)$ innych produktów na „chwile” by znowu zaniknąć)

- albo struktury te są pozostałością/efektem pojawiania się na rynku danego produktu w innych warunkach rynkowych (przykładowo innej cenie waluty, za którą były nabywane hurtowo), wówczas obserwowany efekt powinien być bardziej trwały i dotyczyć stale tylko cen w tych sklepach, które nabyły dany towar w zmienionych warunkach, w takim wypadku nie powinno się obserwować „wędrowania” takiej struktury pomiędzy różnymi produktami jak w przypadku fluktuacji.

Jednakże pytanie o naturę obserwowanych struktur wychodzi poza ramy niniejszej pracy i będzie przedmiotem dalszych badań, których celem będzie analiza i modelowanie dynamiki zmian cen wybranych produktów w czasie. W przyszłych badaniach obserwowany będzie cały proces pojawiania się danego towaru w kolejnych sklepach internetowych jak również śledzony będzie w czasie cały proces zmian ceny danego produktu w różnych sklepach (w odróżnieniu od niniejszej pracy gdzie wszystkie ceny zostały pobrane jednorazowo w tym samym czasie, a celem pracy było zaproponowanie mechanizmów mogących prowadzić do powstania zaobserwowanych rozkładów cen).

W tym miejscu należy zwrócić uwagę na istotną cechę badań nad fizyką procesów obserwowanych w naturze. O ile matematyka pozwala na dowodzenie twierdzeń w oparciu o przyjęty zbiór aksjomatów o tyle w naukach przyrodniczych eksperymenty pozwalają jedynie na falsyfikację badanej teorii lub testowanego modelu. Oznacza to, że w naukach przyrodniczych nie ma czegoś takiego jak „dowód poprawności teorii”. A możliwe jest, co najwyżej stwierdzenie, że nie zaobserwowano różnic na zadanym poziomie istotności pomiędzy przewidywaniami teorii a wynikami eksperymentów prowadzonymi w zadanych warunkach (lub obserwacji – jak ma to miejsce przykładowo w astronomii).

W naukach przyrodniczych pojawienie się nowych danych eksperymentalnych lub obserwacyjnych stojących w sprzeczności z istniejącą teorią najczęściej w

początkowej fazie, prowadzi do wprowadzania kolejnych modyfikacji do samej teorii, (pozwalających opisać nowe fakty). A gdy ilość zebranych obserwacji przekroczy pewną wartość progową staje się możliwym sformułowanie nowej teorii, bardziej uniwersalnej.

Tak, więc otrzymany brak rozbieżności pomiędzy zaproponowanym w tej pracy modelem procesów kształtowania cen a pewną klasą rozkładów cen produktów w sklepach internetowych nie można traktować, jako „dowód poprawności” proponowanej teorii, a jedynie sugestię możliwego scenariusza procesów prowadzących do powstania obserwowanych rozkładów cen.

Niemniej zebrane dane obserwacyjne o rozkładach cen w sklepach internetowych negują poprawność postulowania w naukach ekonomicznych teorii jednej ceny (ang. law of one price), a także koncepcji Gaussowskiego charakteru rozkładu cen.

9. Podsumowanie

Rozprawa dotyczyła modelowania wybranych procesów socjo-ekonomicznych w układach złożonych, a konkretnie procesu kształtowania cen w układzie kupujących i sprzedawców.

Celem pracy było opracowanie dynamicznego modelu ze zindywidualizowanymi agentami - imitującego realne procesy decyzyjne prowadzące do powstania rozkładów cen produktów, takich jak obserwowane w sklepach internetowych.

Teza pracy składa się z dwóch wzajemnie powiązanych części i jest następująca.

Można wykazać, że obserwowane w sklepach internetowych rozkłady cen nie mają charakteru Gaussowskiego, lecz wskazują na ciężkoogonowy charakter rozkładu w kierunku wyższych cen, a także można stworzyć bliski rzeczywistości, wieloagentowy model procesów zakupowych i zweryfikować go poprzez porównanie rozkładów cen rzeczywistych i uzyskanych z modelu.

Zastosowana w pracy metodyka badawcza jest wieloetapowa i składa się z następujących elementów:

- stworzenia modelu sieci zindywidualizowanych agentów mogących w procesie „spotkań” nawiązywać ze sobą więzi,
- przebadania stabilności charakterystyk generowanych w modelu sieci społecznościowych przy ustalonych parametrach oddziaływania agentów w funkcji numeru cyklu symulacji,
- zbadania właściwości (takich jak średni współczynnik klasteryzacji i rozkład stopni wierzchołków) wynikowych sieci społecznościowych w funkcji zmian parametrów modelu,
- doboru parametrów modelu, przy których charakterystyki generowanych wynikowych sieci społecznych będą najbardziej zbliżone do obserwowanych w rzeczywistości,
- analizy statystycznej rozkładów cen konkretnych produktów ze sklepów internetowych,

- stworzenia modelu oddziaływania kupujący-kupujący (bezpośredni przekaz informacji o sprzedawcach w ramach tworzonej przez kupujących sieci społecznej o parametrach imitujących rzeczywistą sieć społeczną) oraz modelu oddziaływania kupujący-sprzedawcy (pośredni przekaz informacji – sprzedawca „widzi” efekty prowadzonej polityki cenowej za pośrednictwem wzrostu lub spadku generowanego zysku)
- porównanie otrzymanych w modelu rozkładów cen z rozkładami cen odpowiednich produktów w sklepach internetowych.

Zrealizowanie poszczególnych etapów oraz zamieszczone w pracy wnioski z kolejnych, przeprowadzonych badań eksperymentalnych potwierdziły prawdziwość tezy.

Za najważniejsze osiągnięcia w pracy uważam:

1. opracowanie nowego, parametrycznego, dynamicznego modelu sztucznej sieci społecznościowej, powstającej w wyniku niskopoziomowych oddziaływań agent-agent naśladującej powstawanie i zanikanie powiązań pomiędzy agentami bez wykorzystania informacji o globalnych parametrach sieci, a w szczególności

- a) stworzenie nowego, zindywidualizowanego modelu agentów, w którym uwzględniono cechy osobowości takie jak przyjaźliwość i profile zainteresowań wpływające na tworzenie przez agenta relacji znajomości (więzi),
- b) uwzględnienie w modelu faktu, że siła relacja znajomości pomiędzy agentami ewoluuje w czasie,
- c) bardziej szczegółowe zamodelowanie procesów spotkań wpływających na powstawanie dynamicznej sieci społecznościowej, poprzez wyróżnienie dwóch składowych spotkań losową i intencjonalną pozwalającą imitować procesy decyzyjne

2. opracowanie i zamodelowanie parametrycznego, dynamicznego modelu współdziałania procesów decyzyjnych agentów kupujących i sprzedających (zawierającego w sobie model sztucznej sieci społecznościowej), a szczególności

- a) opracowanie modelu zindywidualizowanych agentów „kupujących”, z których każdy posiada „osobowość” uwzględniającą preferencje zakupowe oraz własną pamięć różniącą się ilością możliwych do zapamiętania

informacji o sprzedawcach a także dynamiką zapamiętywania-zapominania zależną od „towarzyskiej aktywności” agenta,

- b) zamodelowanie mechanizmu decyzyjnego sprzedawców, uwzględniającego zmiany ich cen i jakości od dynamicznie zmieniającego się zysku
- c) opracowanie pośredniego przekazu informacji od kupujących do sprzedawców poprzez zamodelowanie procesu dynamiki zmian zysków sprzedawców i wzajemnego sprzężenia informacyjno-decyzyjnego pomiędzy oboma klasami agentów.

3. wykonanie analizy rozkładów cen dla dziesięciu losowo wybranych produktów w sklepach internetowych

4. przeprowadzenie weryfikacji generowanych w modelu rozkładów cen, przy pomocy rzeczywistych rozkładów cen obserwowanych w sklepach internetowych dla omawianych produktów.

5. przedstawienie swoich przemyśleń dotyczących tematyki układów złożonych

Autor planuję kontynuację prezentowanych w niniejszej pracy badań, w kierunku analizy i modelowania dynamiki zmian cen produktów, od etapu pojawienia się na rynku, poprzez wzrost sprzedaży aż po zniknięcie z pierwotnego rynku.

10. Spis literatury

- Albert, Reka, and Albert Laszlo Barabasi. 2002. "Statistical Mechanics of Complex Networks." *Reviews of Modern Physics* 74 (1): 47–97. <https://doi.org/10.1088/1478-3967/1/3/006>.
- Alderson, David, Lun Li, Walter Willinger, and John C. Doyle. 2005. "Understanding Internet Topology: Principles, Models, and Validation." *IEEE/ACM Transactions on Networking* 13 (6): 1205–18. <https://doi.org/10.1109/TNET.2005.861250>.
- Allen, Peter M. 2009. "COMPLEXITY IN SOCIO-ECONOMIC SYSTEMS." In *KNOWLEDGE MANAGEMENT, ORGANIZATIONAL INTELLIGENCE AND LEARNING, AND COMPLEXITY*. Vol. II. EOLSS Publications.
- Alstott, Jeff, Ed Bullmore, and Dietmar Plenz. 2014. "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." Edited by Fabio Rapallo. *PLoS ONE* 9 (1): e85777. <https://doi.org/10.1371/journal.pone.0085777>.
- Arlotti, L., E. De Angelis, L. Fermo, M. Lachowicz, and N. Bellomo. 2012. "On a Class of Integro-Differential Equations Modeling Complex Systems with Nonlinear Interactions." *Applied Mathematics Letters* 25 (3): 490–95. <https://doi.org/10.1016/j.aml.2011.09.043>.
- Babusci, D., D. Badoni, I. Balwierz-Pytko, G. Bencivenni, C. Bini, C. Bloise, F. Bossi, et al. 2013. "Limit on the Production of a Light Vector Gauge Boson in ϕ Meson Decays with the KLOE Detector." *Physics Letters B* 720 (1–3): 111–15. <https://doi.org/10.1016/j.physletb.2013.01.067>.
- Ball, Brain, and M.E.J. Newman. 2013. "Friendship Networks and Social Status." *Network Science* 1 (01): 16–30. <https://doi.org/10.1017/nws.2012.4>.
- Bandt, Christoph, and Bernd Pompe. 2002. "Permutation Entropy: A Natural Complexity Measure for Time Series." *Physical Review Letters* 88 (17): 174102. <https://doi.org/10.1103/PhysRevLett.88.174102>.
- Barabási, Albert-László, and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (October): 509–12. <https://doi.org/10.1126/science.286.5439.509>.
- Barabási, Albert-László, and Zoltán N Oltvai. 2004. "Network Biology: Understanding the Cell's Functional Organization." *Nature Reviews Genetics* 5 (2): 101–13. <https://doi.org/10.1038/nrg1272>.
- Berkes, Fikret, and Helen Ross. 2013. "Community Resilience: Toward an Integrated Approach." *Society & Natural Resources* 26 (1): 5–20.
- Blondel, Vincent D., Adeline Decuyper, and Gautier Krings. 2015. "A Survey of Results on Mobile Phone Datasets Analysis." *EPJ Data Science* 4 (1): 10. <https://doi.org/10.1140/epjds/s13688-015-0046-0>.
- Boccaletti, S., G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. 2014. "The Structure and Dynamics of Multilayer Networks." *Physics Reports* 544 (1): 1–122. <https://doi.org/10.1016/j.physrep.2014.07.001>.
- Boccaletti, Stefano, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. 2006. "Complex Networks: Structure and Dynamics." *Physics Reports* 424 (4–5): 175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>.
- Borgatta, Edgar F., and J. L. Moreno. 1954. "Who Shall Survive?" *American Sociological Review* 19 (3): 358. <https://doi.org/10.2307/2087774>.
- Bousquet, F, and C Le Page. 2004. "Multi-Agent Simulations and Ecosystem Management: A Review." *Ecological Modelling* 176 (3–4): 313–32. <https://doi.org/10.1016/j.ecolmodel.2004.01.011>.

- Breuer, H, and F Petruccione. 1995. "COntinuum Mechanics How to Build Master Equations for Complex Systems." *Continuum Mechanics and Thermodynamics* 7: 439–73.
- Caux, Robert De, Christopher Smith, Dominic Kniveton, Richard Black, and Andrew Philippides. 2014. "Dynamic, Small-World Social Network Generation through Local Agent Interactions." *Complexity* 19 (6): 44–53. <https://doi.org/10.1002/cplx.21528>.
- Cilliers, Paul. 1998. *Complexity and Postmodernism*.
- Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703. <https://doi.org/10.1137/070710111>.
- Cohen, Reuven, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. 2011. "Resilience of the Internet to Random Breakdowns." *The Structure and Dynamics of Networks* 9781400841 (3): 507–9.
- Corning, Peter A. 2007. "Synergy Goes to War: A Bioeconomic Theory of Collective Violence." *Journal of Bioeconomics* 9 (2): 109–44. <https://doi.org/10.1007/s10818-007-9017-2>.
- Costa, L da F, Francisco A Rodrigues, Gonzalo Travieso, and P R Villas Boas. 2007. "Characterization of Complex Networks: A Survey of Measurements." *Advances in Physics* 56 (1): 167–242. <https://doi.org/10.1080/00018730601170527>.
- Crucitti, Paolo, Vito Latora, and Sergio Porta. 2006. "Centrality in Networks of Urban Streets." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 16 (1): 015113. <https://doi.org/10.1063/1.2150162>.
- Dajda, Jacek, Roman Dębski, Marek Kisiel-Dorohinicki, and Kamil Piętak. 2014. "Multi-Domain Data Integration for Criminal Intelligence." In *Advances in Intelligent Systems and Computing*, 242:345–52. https://doi.org/10.1007/978-3-319-02309-0_37.
- Darling, D. a. 1957. "The Kolmogorov-Smirnov, Cramer-von Mises Tests." *The Annals of Mathematical Statistics* 28 (4): 823–38. <https://doi.org/10.1214/aoms/1177706788>.
- DeLellis, Pietro, Giovanni Polverino, Gozde Ustuner, Nicole Abaid, Simone Macrì, Erik M Bollt, and Maurizio Porfiri. 2014. "Collective Behaviour across Animal Species." *Scientific Reports* 4 (January): 3723. <https://doi.org/10.1038/srep03723>.
- Efron, B, and R Tibshirani. 1993. *An Introduction to the Bootstrap*. Macmillan Publishers Limited. All rights reserved.
- Eom, Young-ho, and Hang-hyun Jo. 2015. "Tail-Scope: Using Friends to Estimate Heavy Tails of Degree Distributions in Large-Scale Complex Networks." *Scientific Reports* 5 (1): 09752. <https://doi.org/10.1038/srep09752>.
- Erdős, Paul, and Alfréd Rényi. 1959. "On Random Graphs I." *Publicationes Mathematicae (Debrecen)* 6: 290–97.
- Euler, Leonhard. 1736. "Solutio Problematis Ad Geometrian Situs Pertinentis." *Comentarii Academiae Scientiarum Petropolitanae*.
- Garcimartín, A., I. Zuriguel, J.M. Pastor, C. Martín-Gómez, and D.R. Parisi. 2014. "Experimental Evidence of the 'Faster Is Slower' Effect." *Transportation Research Procedia* 2: 760–67. <https://doi.org/10.1016/j.trpro.2014.09.085>.
- Gillespie, Colin S. 2014. "Fitting Heavy Tailed Distributions: The PowerLaw Package." *ArXiv Preprint ArXiv:1407.3492* 64 (July): 1–16. <https://doi.org/http://dx.doi.org/10.18637/jss.v064.i02>.
- . 2015. "Fitting Heavy Tailed Distributions: The PowerLaw Package." *Journal of Statistical Software* 64 (2): 1–16. <https://doi.org/10.18637/jss.v064.i02>.
- Goh, K.-I., Eulsik Oh, Hawoong Jeong, Byungnam Kahng, and Doochul Kim. 2002. "Classification of Scale-Free Networks." *Proceedings of the National Academy of Sciences* 99 (20): 12583–88. <https://doi.org/10.1073/pnas.202301299>.
- Gong, Neil Zhenqiang, Wenchang Xu, Ling Huang, Prateek Mittal, Emil Stefanov, Vyas Sekar, and Dawn Song. 2012. "Evolution of Social-Attribute Networks." In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference - IMC '12*, 131. New York, New York, USA: ACM Press. <https://doi.org/10.1145/2398776.2398792>.

- Gong, Zhen, Xiaoyu Xu, and Guan-Zhu Han. 2017. "'Patient 0' and the Origin of HIV/AIDS in America." *Trends in Microbiology* 25 (1): 3–4. <https://doi.org/10.1016/j.tim.2016.11.005>.
- Heylighen, Francis. 2009. "Complexity and Self-Organization." In *Encyclopedia of Library and Information Sciences, Third Edition*, 13:1215–24. CRC Press. <https://doi.org/10.1081/E-ELIS3-120043869>.
- House, Thomas. 2012. "Modelling Epidemics on Networks." *Contemporary Physics* 53 (3): 213–25. <https://doi.org/10.1080/00107514.2011.644443>.
- Jin, Emily M., Michelle Girvan, and M. E. J. Newman. 2001. "Structure of Growing Social Networks." *Physical Review E* 64 (4): 046132. <https://doi.org/10.1103/PhysRevE.64.046132>.
- Jirsa, Viktor K., and Mingzhou Ding. 2004. "Will a Large Complex System with Time Delays Be Stable?" *Physical Review Letters* 93 (7): 1–4. <https://doi.org/10.1103/PhysRevLett.93.070602>.
- Krämer, Andreas, Martin Jung, and Thomas Burgartz. 2016. "A Small Step from Price Competition to Price War: Understanding Causes, Effects and Possible Countermeasures." *International Business Research* 9 (3): 1. <https://doi.org/10.5539/ibr.v9n3p1>.
- KSC. 2007. "Vehicle Assembly Building."
- Ladyman, James; James; Lambert, and Karoline Wiesner. 2011. "What Is a Complex System?" *European Journal for Philosophy of Science* 3 (1): 33–67. <https://doi.org/10.1007/s13194-012-0056-8>.
- Leng, Siyang, Wei Lin, and Jürgen Kurths. 2016. "Basin Stability in Delayed Dynamics." *Scientific Reports* 6 (1): 21449. <https://doi.org/10.1038/srep21449>.
- Lin, Hong. 2007. *Architectural Design of Multi-Agent Systems*. Edited by Hong Lin. IGI Global. <https://doi.org/10.4018/978-1-59904-108-7>.
- Liu, Yang-Yu, Jean-Jacques Slotine, and Albert-László Barabási. 2011. "Controllability of Complex Networks." *Nature* 473 (7346): 167–73. <https://doi.org/10.1038/nature10011>.
- M.E.J. Newman. 2003. "The Structure and Function of Complex Networks." *SIAM Review* 45 (2): 167–256.
- Malik, Nishant, and Peter J. Mucha. 2013. "Role of Social Environment and Social Clustering in Spread of Opinions in Coevolving Networks." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23 (4): 043123. <https://doi.org/10.1063/1.4833995>.
- Marković, Dimitrije, and Claudius Gros. 2014. "Power Laws and Self-Organized Criticality in Theory and Nature." *Physics Reports* 536 (2): 41–74. <https://doi.org/10.1016/j.physrep.2013.11.002>.
- McLuhan, Marshall. 1962. "The Gutenberg Galaxy." Toronto: University of Toronto Press.
- Milgram, Stanley. 1967. "The Small World Problem." *Psychology Today* 1 (January): 60–67. <https://doi.org/10.1037/e400002009-005>.
- Mislove, Alan, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. 2007. "Measurement and Analysis of Online Social Networks." In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement - IMC '07*, 29. New York, New York, USA: ACM Press. <https://doi.org/10.1145/1298306.1298311>.
- Mitzenmacher, Michael. 2004. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2): 226–51. <https://doi.org/10.1080/15427951.2004.10129088>.
- Muchnik, Lev, Sen Pei, Lucas C Parra, Saulo D S Reis, José S Andrade, Shlomo Havlin, and Hernán Makse. 2013. "Origins of Power-Law Degree Distribution in the Heterogeneity of Human Activity in Social Networks." *Scientific Reports* 3: 1783. <https://doi.org/10.1038/srep01783>.
- Newman, M. E. J. 2011. "Resource Letter CS–1: Complex Systems." *American Journal of Physics* 79 (8): 800–810. <https://doi.org/10.1119/1.3590372>.

- Newman, M. E. J., Stephanie Forrest, and Justin Balthrop. 2002. "Email Networks and the Spread of Computer Viruses." *Physical Review E* 66 (3): 035101. <https://doi.org/10.1103/PhysRevE.66.035101>.
- Newman, M. E. J., and Juyong Park. 2003. "Why Social Networks Are Different from Other Types of Networks." *Physical Review E* 68 (3): 036122. <https://doi.org/10.1103/PhysRevE.68.036122>.
- Newman, M. E.J. 2004. "Analysis of Weighted Networks." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 70 (5): 9. <https://doi.org/10.1103/PhysRevE.70.056131>.
- Newman, Mark E. J. 2002. "Random Graphs as Models of Networks." *Handbook of Graphs and Networks: From the Genome to the Internet*, no. 1: 34–68. <https://doi.org/10.1002/3527602755.ch2>.
- Nie, Norman H., D. Sunshine Hillygus, and Lutz Erbring. 2000. "Internet Use, Interpersonal Relations, and Sociability: A Time Diary Study." In *The Internet in Everyday Life*, 4:213–43. Oxford, UK: Blackwell Publishers Ltd. <https://doi.org/10.1002/9780470774298.ch7>.
- Ortblad, Katrina F., Rafael Lozano, and Christopher J.L. Murray. 2013. "The Burden of HIV." *Aids* 27 (13): 2003–17. <https://doi.org/10.1097/QAD.0b013e328362ba67>.
- Pais-Vieira, Miguel, Gabriela Chiuffa, Mikhail Lebedev, Amol Yadav, and Miguel A. L. Nicolelis. 2015. "Building an Organic Computing Device with Multiple Interconnected Brains." *Scientific Reports* 5 (July): 11869. <https://doi.org/10.1038/srep11869>.
- Parker, D C, S M Manson, M a Janssen, M J Hoffmann, and P Deadman. 2003. "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review." *Annals of the Association of American Geographers* 93 (2): 314–37. <https://doi.org/10.1111/1467-8306.9302004>.
- Pastor-Satorras, Romualdo, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. "Epidemic Processes in Complex Networks." *Reviews of Modern Physics* 87 (3): 925–79. <https://doi.org/10.1103/RevModPhys.87.925>.
- Piskor-Ignatowicz, C., J. Smyrski, P. Moskal, H.-H. Adam, A. Budzanowski, E. Czerwiński, R. Czyżykiewicz, et al. 2007. "NEAR THRESHOLD η MESON PRODUCTION IN Dp COLLISIONS." *International Journal of Modern Physics A* 22 (02n03): 528–32. <https://doi.org/10.1142/S0217751X07035768>.
- Pourbohloul, Babak, and Marie-Paule Kieny. 2011. "Complex Systems Analysis: Towards Holistic Approaches to Health Systems Planning and Policy." *Bulletin of the World Health Organization* 89 (4): 242–242. <https://doi.org/10.2471/BLT.11.087544>.
- Press, William H, S A Teukolsky, W T Vetterling, and B P Flannery. 1992. "Numerical Recipes in C: The Art of Scientific Computing, Second Edition." Cambridge Univ. Press, New York.
- R Core Team. 2017. "R: A Language and Environment for Statistical Computing." Vienna, Austria.
- Reis, Saulo D. S., Yanqing Hu, Andrés Babino, José S. Andrade Jr, Santiago Canals, Mariano Sigman, and Hernán A. Makse. 2014. "Avoiding Catastrophic Failure in Correlated Networks of Networks." *Nature Physics* 10 (10): 762–67. <https://doi.org/10.1038/nphys3081>.
- Rendón de la Torre, Stephanie, Jaan Kalda, Robert Kitt, and Jüri Engelbrecht. 2016. "On the Topologic Structure of Economic Complex Networks: Empirical Evidence from Large Scale Payment Network of Estonia." *Chaos, Solitons & Fractals* 90 (September): 18–27. <https://doi.org/10.1016/j.chaos.2016.01.018>.
- Roberts, Anthony John. 2014. "Model Emergent Dynamics in Complex Systems." SIAM.
- Saramäki, Jari, Mikko Kivelä, Jukka Pekka Onnela, Kimmo Kaski, and János Kertész. 2007. "Generalizations of the Clustering Coefficient to Weighted Complex Networks." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 75 (2): 2–5. <https://doi.org/10.1103/PhysRevE.75.027105>.

- Scheinkman, Jose A. 2008. "Social Interactions." In *The New Palgrave Dictionary of Economics* 2, 2:11. <https://doi.org/10.1146/annurev.economics.050708.143312>.
- Shapiro, S S, and M B Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52 (3/4): 591–611.
- Shenton, Andrew K. 2004. "Strategies for Ensuring Trustworthiness in Qualitative Research Projects." *Education for Information* 22: 63–75. <https://doi.org/10.1111/j.1744-618X.2000.tb00391.x>.
- Skibiński, Roman, Jacek Golak, Kacper Topolnicki, Henryk Witała, Evgeny Epelbaum, Hiroyuki Kamada, Hermann Krebs, Ulf-G. Meißner, and Andreas Nogga. 2017. "Modern Chiral Forces Applied to the Nucleon–Deuteron Radiative Capture." *Few-Body Systems* 58 (2): 28. <https://doi.org/10.1007/s00601-016-1190-1>.
- Sumpter, D. J. T. 2006. "The Principles of Collective Animal Behaviour." *Philosophical Transactions of the Royal Society B: Biological Sciences* 361 (1465): 5–22. <https://doi.org/10.1098/rstb.2005.1733>.
- Szuba, Tadeusz Ted. 2011. "Importance of the Chaos for Computational Processes of Collective Intelligence in Social Structures." In *Chaos Theory*, 414–22. WORLD SCIENTIFIC. https://doi.org/10.1142/9789814350341_0048.
- Ugander, Johan, Brian Karrer, Lars Backstrom, and Cameron Marlow. 2011. "The Anatomy of the Facebook Social Graph." *Arxiv Preprint ArXiv* abs/1111.4 (November): 1–17.
- Vázquez, Alexei, João Gama Oliveira, Zoltán Dezső, Kwang-Il Goh, Imre Kondor, and Albert-László Barabási. 2006. "Modeling Bursts and Heavy Tails in Human Dynamics." *Physical Review E* 73 (3): 036127. <https://doi.org/10.1103/PhysRevE.73.036127>.
- Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57 (2): 307–33.
- Wąs, Jarosław, and Robert Lubaś. 2014. "Towards Realistic and Effective Agent-Based Models of Crowd Dynamics." *Neurocomputing* 146 (September): 199–209. <https://doi.org/10.1016/j.neucom.2014.04.057>.
- Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p -Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129–33. <https://doi.org/10.1080/00031305.2016.1154108>.
- Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393 (6684): 440–42. <https://doi.org/10.1038/30918>.
- Weng, Lilian, Filippo Menczer, and Yong-Yeol Ahn. 2013. "Virality Prediction and Community Structure in Social Networks." *Scientific Reports* 3: 2522. <https://doi.org/10.1038/srep02522>.
- Xiao, Heng, Jian-xun Wang, and Roger G Ghanem. 2017. "A Random Matrix Approach for Quantifying Model-Form Uncertainties in Turbulence Modeling." *Computer Methods in Applied Mechanics and Engineering* 313 (January): 941–65. <https://doi.org/10.1016/j.cma.2016.10.025>.
- Ye, Cheng, Andrea Torsello, Richard C. Wilson, and Edwin R. Hancock. 2015. "Thermodynamics of Time Evolving Networks." In *International Workshop on Graph-Based Representations in Pattern Recognition.*, 4538:315–24. https://doi.org/10.1007/978-3-319-18224-7_31.
- Yu, Shuiyuan, Haitao Liu, and Chunshan Xu. 2011. "Statistical Properties of Chinese Phonemic Networks." *Physica A: Statistical Mechanics and Its Applications* 390 (7): 1370–80. <https://doi.org/10.1016/j.physa.2010.12.019>.
- Zachara, Marek, Ewa Majchrzyk-Zachara, and Cezary Piskor-Ignatowicz. 2013. "Simulation of Customers Behaviour as a Method for Explaining Certain Market Characteristic." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7770 LNCS:217–36. https://doi.org/10.1007/978-3-642-36815-8_10.
- Zachara, Marek, and Cezary Piskor-Ignatowicz. 2010. "Analysis of Product Price Distribution and Pricing Policy of Online Merchants." In *Materiały Krakowskiej Konferencji Młodych Uczonych*, 659–65.

- . 2016. “Agent-Based Creation and Simulation of Artificial Social Networks and the Analysis of Their Properties.” *Computing in Science & Engineering* 18 (4): 34–41. <https://doi.org/10.1109/MCSE.2016.46>.
- Zanin, Massimiliano, Luciano Zunino, Osvaldo A. Rosso, and David Papo. 2012. “Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review.” *Entropy* 14 (12): 1553–77. <https://doi.org/10.3390/e14081553>.
- Zygmunt, Anna, and Bogdan Gliwa. 2015. “The Comparison of Users Activity on the Example of Polish and American Blogosphere.” *Scientific Programming* 2015: 1–11. <https://doi.org/10.1155/2015/907547>.