

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Teresa Ponikowska

Nr albumu: 315587

**Sieci Bayesowskie i sieci
(samo)wspierania, teoria
i zastosowania do danych z systemu
USOS**

**Praca magisterska
na kierunku MATEMATYKA**

Praca wykonana pod kierunkiem
dra inż. Przemysława Biecka
Instytut Matematyki Stosowanej i Mechaniki

Wrzesień 2012

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

Streszczenie

Niniejsza praca przedstawia analizę danych historycznych systemu USOS Wydziału Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego w celu wyznaczenia kursów skupiających studentów o podobnych zainteresowaniach oraz znalezienia wśród kursów tych, które pozytywnie wpływają na pozostałe. Pierwszy problem zbadano przy użyciu sieci Bayesowskich, wykorzystując efektywność pakietów programu R z nimi związanych. Drugi problem zbadano przy użyciu własnej koncepcji sieci (samo)wspierania. Wszystkie jej komponenty zostały zaimplementowane również przy użyciu środowiska R.

Słowa kluczowe

sieć Bayesowska, sieć przyczynowa, sieć (samo)wspierania, USOS, wizualizacja danych, wykorzystanie R.

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.2 Statystyka

Klasyfikacja tematyczna

62-07 Data analysis
62-09 Graphical methods
62C12
62C10
60E05
60E99

Tytuł pracy w języku angielskim

Bayesian Networks and the (self)support networks, theory and applications.

Spis treści

Wprowadzenie	9
1. Informacje wstępne	11
1.1. Sieci Bayesowskie	11
1.1.1. Rachunek prawdopodobieństwa	11
1.1.2. Teoria grafów	12
1.2. Sieci (samo)wspierania	13
1.3. Dane z systemu USOS	14
1.3.1. Informacje ogólne o USOS	14
1.3.2. Opis danych	15
1.3.3. Wybór analizowanych kursów	15
2. Sieci Bayesowskie	19
2.1. Część teoretyczna	19
2.1.1. Sieci przyczynowe	19
2.1.2. Dokładna definicja sieci Bayesowskiej	22
2.1.3. Algorytm Junction Tree	23
2.1.4. Estymacja parametrów sieci Bayesowskich	26
2.1.5. Wyznaczenie struktury sieci Bayesowskich	26
2.1.6. Wybrane pakiety programu R dedykowane sieciom Bayesowskim	26
2.2. Wykorzystanie sieci Bayesowskich do danych USOS	28
2.2.1. Wyuczenie struktury <i>USOS</i> -sieci	28
2.2.2. Estymacja parametrów <i>USOS</i> -sieci	29
2.2.3. Rezultaty	29
3. Sieci (samo)wspierania	35
3.1. Przygotowanie danych	35
3.2. Analiza wsparcia	36
3.2.1. Koncepcja	37
3.2.2. Własności współczynnika wsparcia i zbiorów wspierających	38
3.2.3. Dobór parametrów	40
3.3. Analiza samowsparcia	40
3.3.1. Koncepcja	40
3.3.2. Własności współczynnika wsparcia i klasy samowspierającej	41
3.3.3. Dobór parametrów	41
3.4. Kierunek studiów	42
3.4.1. Metoda 1 - badanie przedmiotu rozróżniającego	42
3.4.2. Metoda 2 - analiza kodów kursów	43

3.4.3. Kierunek studiów - Porównanie metod	44
4. Wyniki	45
4.1. Analiza szczegółowa - przykład	45
4.2. Graficzne przedstawienie sieci (samo)wspierania	59
4.2.1. Konstrukcja	60
4.2.2. Prezentacja	60
Podsumowanie	67
A. Liczności zbiorów $W_K^{(\alpha, M)}$	69
B. Raport szczegółowy - wybrane elementy	77
C. Płyta CD	97
Bibliografia	99

Spis rysunków

1.1. Zależność puli kursów od specjalizacji	18
2.1. Rodzaje połączeń w sieciach przyczynowych	20
2.2. Sieć przyczynowa dla problemu uruchomienia samochodu	21
2.3. Sieć Bayesowska dla problemu uruchomienia samochodu	23
2.4. Struktura sieci Bayesowskiej nad zbiorem $\mathcal{U} = \{A_1, A_2, A_3, A_4, A_5, A_6\}$	24
2.5. Umoralnienie sieci i graf dziedziny	25
2.6. Mapa ciepła dla <i>USOS</i> -sieci.	31
2.7. Schemat <i>USOS</i> -sieci	32
2.8. Wybrane elementy <i>USOS</i> -sieci	33
3.1. Schemat przetworzenia danych	36
4.1. Rozkład ocen dla Procesów stochastycznych I	46
4.2. Diagram Venna dla pary (<i>pr.stoch1</i> , <i>jtrrz</i>)	48
4.3. Diagram Venna dla pary (<i>pr.stoch1</i> , <i>r.pr2a</i>)	50
4.4. Diagram Venna dla pary (<i>pr.stoch1</i> , <i>r.pr2b</i>)	52
4.5. Diagram Venna dla pary (<i>pr.stoch1</i> , <i>pr.stoch2</i>)	54
4.6. Rozkład ocen dla Procesów stochastycznych I - samowsparcie	57
4.7. Wykres ilości kursów wspierających	59
4.8. Wartości współczynników pk_1 i pk_0 dla <i>pr.stoch1</i>	61
4.9. Połączenie pary testowej w sieci (samo)wspierania	61
4.10. Sieć (samo)wspierania - sukces $s_{poz.ocena}$, studenci matematyki	63
4.11. Sieć (samo)wspierania - sukces $s_{poz.ocena}$, studenci informatyki	64
4.12. Sieć (samo)wspierania - sukces $s_{p*przecietna}$, studenci matematyki	65
4.13. Sieć (samo)wspierania - sukces $s_{p*przecietna}$, studenci informatyki	66

Spis tabel

1.1. Przykład tabel CPT i JPT	12
1.2. Atrybuty udostępnionych danych	15
1.3. Pula kursów	17
2.1. Zależność ilości możliwych BN od liczby zmiennych	27
2.2. Główne funkcje pakietu gRain	27
2.3. Główne funkcje pakietu bnlearn	28
3.1. Wartości różnicujące dla puli kursów	36
3.2. Schemat tablicy dwudzielnej dla analizy wsparcia	37
3.3. Schemat tablicy dwudzielnej dla problemu samowsparcia	39
3.4. Schemat tablicy dwudzielnej dla rozbicia $B_1 = \{\omega_1\}$, $B_0 = \{\omega_2, \dots, \omega_n\}$. . .	39
3.5. Schemat tablicy dwudzielnej wykorzystywanej w analizie samowsparcia	40
3.6. Liczności klas samowspierających \mathcal{S}_M^α - $s_{poz.ocena}$	42
3.7. Kodowanie kursów matematycznych i informatycznych na Wydziale MIMUW	44
4.1. Tablice dwudzielne analizy wsparcia sukcesu $s_{poz.ocena}$ - para testowa (<i>pr.stoch1, jtrrz</i>)	48
4.2. Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa (<i>pr.stoch1, jtrrz</i>)	48
4.3. Tablice dwudzielne dla analizy wsparcia i sukcesu s_{p*przt} - para testowa (<i>pr.stoch1, jtrrz</i>)	48
4.4. Wartości współczynników analizy wsparcia i sukcesu $s_{p*przecietna}$ - para testowa (<i>pr.stoch1, jtrrz</i>)	49
4.5. Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa (<i>pr.stoch1, r.pr2a</i>)	51
4.6. Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa (<i>pr.stoch1, r.pr2a</i>)	51
4.7. Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{p*przecietna}$ - para testowa (<i>pr.stoch1, r.pr2a</i>)	51
4.8. Wartości współczynników analizy wsparcia i sukcesu $s_{p*przecietna}$ - para testowa (<i>pr.stoch1, r.pr2a</i>)	51
4.9. Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa (<i>pr.stoch1, r.pr2b</i>)	53
4.10. Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa (<i>pr.stoch1, r.pr2b</i>)	53
4.11. Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{p*przecietna}$ - para testowa (<i>pr.stoch1, r.pr2b</i>)	53

4.12. Wartości współczynników analizy wsparcia i sukcesu $s_{p^*przecietna}$ - para testowa ($pr.stoch1, r.pr2b$)	53
4.13. Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa ($pr.stoch1, pr.stoch2$)	55
4.14. Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ - para testowa ($pr.stoch1, pr.stoch2$)	55
4.15. Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{p^*przecietna}$ - para testowa ($pr.stoch1, pr.stoch2$)	55
4.16. Wartości współczynników analizy wsparcia i sukcesu $s_{p^*przecietna}$ - para testowa ($pr.stoch1, pr.stoch2$)	55
4.17. Tablice dwudzielne dla samowsparcia i sukcesu $s_{poz.ocena}$ dla - kurs $pr.stoch1$	58
4.18. Wartości współczynników dla samowsparcia i sukcesu $s_{poz.ocena}$ - kurs $pr.stoch1$	58
4.19. Tablice dwudzielne dla samowsparcia i sukcesu $s_{p^*przecietna}$ - kurs $pr.stoch1$	58
4.20. Wartości współczynników dla samowsparcia i sukcesu $s_{p^*przecietna}$ - kurs $pr.stoch1$	58
4.21. Ilość kursów wspierających	60
4.22. Rodzaje krawędzi występujące w sieciach (samo)wspierania	62
A.1. Wykresy i tabele licznosci zbiorów $W_K^{(\alpha, M)}$ dla sukcesu $s_{poz.ocena}$	75

Wprowadzenie

Punktem wyjścia niniejszej pracy była praca dyplomowa mgr Filipa Grotkowskiego i związany z nią projekt *Ocenarium*. Jak określa autor w streszczeniu swojej pracy:

„Ocenarium jest systemem analitycznym związanym z USOS-em. Jego celem jest pokazanie różnych aspektów funkcjonowania Wydziału Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego. Oprócz tytułowych ocen są to ankiety, profile seminariów, rankingi prowadzących zajęcia czy analiza przeżycia studentów na studiach.”. [3]

Jedną z idei, jaką można przypisać do tego projektu, jest zwiększenie świadomości studentów podczas szeroko rozumianego procesu studiowania. Widać to w raportach, dzięki którym studenci mogą zapoznać się z różnymi aspektami studiów w oparciu o graficzne i tabelaryczne przedstawienie danych historycznych. Szczegółowe informacje dotyczące tego projektu można znaleźć w [3].

Jakkolwiek *Ocenarium* stanowi sam w sobie kompletny projekt, można próbować go rozwijać. Jedną z koncepcji rozwojowych jest zbadanie zależności przyczynowych, które hipotetycznie mogą występować wśród kursów oferowanych przez Wydział Matematyki, Informatyki i Mechaniki (Wydział MIMUW). Posiadając jedynie informacje o tym, na jakie kursy uczęszczali studenci, można skonstruować model obrazujący zależności pomiędzy wybieranymi kursami. Dodatkowo dane historyczne dotyczące kolejności, w jakiej studenci realizowali poszczególne kursy oraz jakie uzyskiwali z nich wyniki otwierają drogę do ciekawszych analiz. Jedną z nich jest identyfikacja tych kursów, które pomagają osiągnąć upragniony sukces, związany oczywiście z zaliczeniem kursu (choć niekoniecznie bezpośrednio). Wówczas studenci, którzy chcieliby zrealizować daną parę kursów (i którzy mogą to zrobić w dowolnej kolejności!), mieliby możliwość takiego rozplanowania swojej siatki godzin, która pozwoli im zoptymalizować swoje rezultaty.

Prezentowana praca magisterska dotyczy w/w zagadnień. Pierwszy problem, czyli próbę stworzenia siatki zależności, zbadano przy użyciu sieci Bayesowskich, a dokładniej zagadnienia rekonstrukcji struktury sieci w oparciu o pakiet **bnlearn** programu **R**. Próba znalezienia kursów, które wspierają inne nastąpiła w oparciu o własną koncepcję sieci (samo)wspierania. Składają się na nie *analiza wsparcia* i *analiza samowsparcia*. Dokładne określenie tych terminów znajduje się w drugiej części Rozdziału 1. Jego pierwsza część zawiera podstawowe definicje z teorii rachunku prawdopodobieństwa i teorii grafów, konieczne do nakreślenia ogólnej koncepcji sieci Bayesowskiej. Ostatnia to zaprezentowanie oryginalnych danych udostępnionych przez Wydział MIMUW do przeprowadzenia analizy.

Kolejne rozdziały są następujące:

W pierwszej części Rozdziału 2 nakreślono koncepcję sieci Bayesowskich w oparciu o pierwszą część [6]. Znajdują się w nim informacje m.in. o sieciach przyczynowych czy algorytmie propagowania informacji *Junction Tree*. Pokrótkę opisano również sposoby estymacji parametrów w sieciach Bayesowskich oraz sposoby wyznaczania ich struktury. Dodatkowo przed-

stawiono pakiety `gRain` i `bnlearn` programu R, które są dedykowane właśnie sieciom Bayesowskim. Druga część Rozdziału 2 to przykład wykorzystania techniki sieci Bayesowskich do danych USOS.

Rozdział 3 jest z kolei w całości poświęcony sieciom (samo)wspierania. Początek opisuje, w jaki sposób należy przygotować dane z systemu USOS, by można je było w nich wykorzystać. Dalsza część to opisanie metodologii kryjącej się za sieciami (samo)wspierania, uwzględniając podział na jej składowe. W końcowej części można znaleźć informacje o dwóch sposobach estymacji informacji o kierunku studiów, wybranych przez studentów MIMUW.

Rozdział 4 prezentuje wyniki wykonanej analizy (samo)wspierania. W jego pierwszej części przedstawiono przykładową analizę szczegółową, pomocną w doborze końcowych parametrów analizy samo(wspierania). Fragment ten stanowi jednocześnie wskazówkę do sposobu interpretacji danych zawartych w dołączonym Raporcie *Szczegóły wyznaczania sieci (samo)wspierania*. Druga część przedstawia już końcowe sieci (samo)wspierania, mające na celu poprawę uzyskiwanych przez studentów Wydziału MIMUW wyników.

Rozdział 1

Informacje wstępne

Rozdział ma na celu wprowadzenie w tematykę prezentowanej pracy. Został podzielony na trzy części związane z odpowiednio z sieciami Bayesowskimi, sieciami (samo)wspierania oraz analizowanymi danymi historycznymi. Definicje i twierdzenia z Rozdziału 1.1.1 zostały zaczerpnięte z [5, 6]. Rozdział 1.1.2 opracowano głównie w oparciu o [2, 11].

1.1. Sieci Bayesowskie

Ogólna koncepcja Sieci Bayesowskich została przedstawiona w Rozdziale 2. Poniżej zamieszczono podstawowe pojęcia z nimi związane.

1.1.1. Rachunek prawdopodobieństwa

Definicja 1.1. (Prawdopodobieństwo warunkowe) Niech A, B będą zdarzeniami, gdzie $P(B) > 0$. Wówczas prawdopodobieństwem zajścia zdarzenia A pod warunkiem zajścia zdarzenia B nazywa się liczbę:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1.1)$$

Definicja 1.2. (Niezależność zdarzeń) Zdarzenia A i B są nazywane niezależnymi, jeśli:

$$\begin{aligned} P(A|B) &= P(A), \\ P(B|A) &= P(B). \end{aligned} \quad (1.2)$$

Definicja 1.3. (Warunkowa niezależność zdarzeń) Zdarzenia A i B są warunkowo niezależne pod warunkiem zajścia zdarzenia C , jeśli:

$$\begin{aligned} P(A|B \cap C) &= P(A|C), \\ P(B|A \cap C) &= P(B|C). \end{aligned} \quad (1.3)$$

Symetryczność pojęć niezależności pozwala na zapisanie równań (1.2) i (1.3) w postaci:

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(A \cap B|C) &= P(A|C)P(B|C). \end{aligned} \quad (1.4)$$

Twierdzenie 1.1 (Wzór Bayesa). Niech $B_i, i \in \{1, \dots, N\}$ stanowią rozbięcie przestrzeni Ω na zdarzenia o dodatnich prawdopodobieństwach oraz niech $P(A) > 0$. Wówczas:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(\sum_{i=1}^N P(A|B_i)P(B_i))}. \quad (1.5)$$

Niech teraz A, B, C będą zmiennymi losowymi o przestrzeniach stanów $sp(A) = \{a_1, \dots, a_n\}$, $sp(B) = \{b_1, \dots, b_m\}$, $sp(C) = \{c_1, \dots, c_n\}$ odpowiednio. Ponieważ zbiory przyjmowanych wartości są skończone, informacje o rozkładach tych zmiennych mogą być przechowywane w postaci tabel rozkładów.

Przykładowo, dla zmiennych A, B ich rozkłady można przedstawić w tabeli o n i m wejściach, zaś rozkład łączny $P(A, B)$ oraz rozkłady warunkowe $P(A|B)$, $P(B|A)$ - w tabeli o $n \times m$ wejściach. Korzystając z następujących własności funkcji prawdopodobieństwa:

$$\sum_{i=1}^n P(A = a_i | B = b_j) = 1, \quad \forall b_j \in sp(B), \quad (1.6)$$

$$P(A, B) = \sum_{i=1}^n \sum_{j=1}^m P(A = a_i, B = b_j) = 1, \quad (1.7)$$

należy pamiętać, że kolumny tablic rozkładu warunkowego (CPT, ang. *Conditional Probability Table*) powinny sumować się do wartości 1 (1.6), podczas gdy tablice rozkładu łącznego (JPT, ang. *Joint Probability Table*) powinny sumować się w całości do tej wartości (1.7). Powyższą uwagę można zaobserwować na przykładzie Tabeli 1.1, w której przedstawiono przykładowe tablice CPT i JPT dla zmiennych A, B przyjmujących odpowiednio dwa i trzy stany.

Tabela CPT				Tabela JPT			
	b_1	b_2	b_3		b_1	b_2	b_3
a_1	0.4	0.3	0.6	a_1	0.16	0.12	0.12
a_2	0.6	0.7	0.4	a_2	0.24	0.28	0.08

Tabela 1.1: Przykład tabel JPT $P(A, B)$ i CPT $P(A|B)$ dla dyskretnych zmiennych A, B .
Źródło: Na podstawie [6].

Twierdzenie 1.2 (Reguła łańcucha). Rozważmy zbiór zmiennych losowych $\mathcal{U} = \{A_1, \dots, A_n\}$. Wówczas rozkład łączny $P(\mathcal{U})$ można wyrazić jako:

$$P(\mathcal{U}) = P(A_n | A_1, \dots, A_{n-1}) P(A_{n-1} | A_1, \dots, A_{n-2}) \cdots P(A_2 | A_1) P(A_1). \quad (1.8)$$

1.1.2. Teoria grafów

Definicja 1.4. (Graf skierowany, nieskierowany) Graf $\mathcal{G} = (V, E)$ jest strukturą składającą się ze skończonego zbioru wierzchołków V i zbioru krawędzi E stanowiącego zbiór par wierzchołków V .

Jeśli krawędzie są parami nieuporządkowanymi, wówczas mówi się o *grafach nieskierowanych*. Jeśli zaś krawędzie są parami uporządkowanymi - o *grafach skierowanych*.

Definicja 1.5. (Graf prosty) Grafem prostym nazywa się graf, który nie zawiera krawędzi postaci (v, v) , $\forall v \in V$ (pętli) oraz w którym każda krawędź (v, w) ze zbioru E występuje w nim dokładnie raz.

Definicja 1.6. (Ścieżka) Przez ścieżkę długości n w grafie $\mathcal{G} = (V, E)$ prowadzącą od wierzchołka u do wierzchołka v rozumie się ciąg różnych wierzchołków $(u = v_0, v_1, \dots, v_n = v)$ spełniający:

$$\{v_i, v_{i+1}\} \in E \quad \forall i = 0, 1, \dots, n-1. \quad (1.9)$$

Jeśli każda z par jest parą uporządkowaną, wówczas taką ścieżkę nazywa się *skierowaną*.

Definicja 1.7. (Cykl) Przez cykl długości n w grafie $\mathcal{G} = (V, E)$ rozumie się ścieżkę prowadzącą od wierzchołka u do wierzchołka u , dla $u \in V$.

Definicja 1.8. (Acykliczny graf skierowany (DAG)) Graf $\mathcal{G} = (V, E)$ jest acyklicznym grafem skierowanym, jeśli jest prostym grafem skierowanym i nie zawiera cykli skierowanych długości $m \geq 1$.

Definicja 1.9. (Klika) Przez klikę grafu \mathcal{G} rozumie się maksymalny podgraf, w którym każde dwa wierzchołki są ze sobą połączone.

Definicja 1.10. (Rodzice, dzieci) Niech Graf $\mathcal{G} = (V, E)$ będzie grafem skierowanym. Wówczas dla każdego wierzchołka $v \in V$ rozważa się zbiory ich bezpośrednich poprzedników i bezpośrednich następników, określanymi jako *rodziców* i *dzieci*:

$$pa(v) = \{w \in V : (w, v) \in E\}, \quad \forall v \in V, \quad (1.10)$$

$$ch(v) = \{w \in V : (v, w) \in E\}, \quad \forall v \in V. \quad (1.11)$$

Definicja 1.11. (Przodek, potomek) Niech Graf $\mathcal{G} = (V, E)$ będzie grafem skierowanym. Wówczas dla każdego wierzchołka $v \in V$ rozważa się zbiory ich wszystkich poprzedników i wszystkich następników, określanymi jako *przodków* i *potomków*:

$$an(v) = \{w \in V : \exists \tau = (v_0, \dots, v_n) : v_0 = w \wedge v_n = v \wedge \tau - \text{skierowana}\}, \quad (1.12)$$

$$de(v) = \{w \in V : \exists \tau = (v_0, \dots, v_n) : v_0 = v \wedge v_n = w \wedge \tau - \text{skierowana}\}. \quad (1.13)$$

Sieci Bayesowskie odnoszą się tylko do grafów prostych.

1.2. Sieci (samo)wspierania

Zaimplementowane na potrzeby niniejszej pracy sieci (samo)wspierania służą wskazaniu kursów, które mają istotny wpływ na osiągnięcie sukcesu w innym kursie realizowanym przez danego studenta. Dokładny opis algorytmu stanowiącego trzon sieci (samo)wspierania można znaleźć w Rozdziale 3. Jednak wcześniej należy uszczegółowić ogólną koncepcję z nimi związaną podaną we *Wprowadzeniu*.

W pierwszej kolejności należy nadać sens pojęciu *sukces*. Dla części studentów sukcesem może być uzyskanie pozytywnej oceny na koniec danego cyklu dydaktycznego. Nie ma znaczenia, czy odbędzie się to w I-szym, czy w II-gim terminie. Inni mogą z kolei ocenić, że brakiem sukcesu jest już konieczność ponownego przystąpienia do egzaminu. Studenci jeszcze bardziej rygorystyczni oceniają, że sukcesem nie jest samo osiągnięcie wyniku premiującego w I-szym terminie - istotne jest, by wynik ten pozwolił plasować się w czołówce.

Z powyższych uwag wynika, że sukces (s) należy traktować jako parametr. Należy zauważyć, że przedstawione powyżej rozważania nie wyczerpują wszystkich możliwych konfiguracji tego parametru. Nie mniej, od jego postaci zależy sposób przygotowania danych (roz. 3.1), a tym samym wyniki końcowe (roz. 4). W przedstawionej pracy rozważa się tylko dwie postaci s :

1. Sukces pozytywnej oceny ($s_{poz.ocena}$): uzyskanie pozytywnego wyniku w pierwszym podejściu do danego kursu;

2. Sukces oceny ponad*przeciętnej ($s_{p*przecietna}$): uzyskanie wynik pozwalającego sklasyfikować studenta w górnej części wyników z danego kursu podczas pierwszego podejścia do niego;

Z matematycznego punktu widzenia mechanizm konstruowania sieci (samo)wspierania będzie polegał na szukaniu uporządkowanych par kursów o pewnej własności. Dzięki niej będzie można określić, czy dany kurs stanowi *wsparcie* dla innego kursu. Dla ustalenia uwagi, wprowadza się następujące definicje:

Definicja 1.12. (Pula kursów) Pula kursów (\mathcal{K}) oznacza zbiór kursów poddane analizie (samo)wsparcia. Dowolny kurs z puli kursów zapisuje się jako K .

Definicja 1.13. (Para testowa) Przez parę testową rozumie się uporządkowaną parę kursów $(K_b, K_t) \in \mathcal{K} \times \mathcal{K}$ poddaną analizie (samo)wsparcia. Współrzędna K_b nosi nazwę kursu bazowego, zaś K_t - kursu testowego.

Definicja 1.14. (Kurs wspierający, zbiór wspierający) Dla pary testowej (K_b, K_t) kurs K_t nazwiemy kursem wspierającym (dla K_b), jeśli uczęszczanie na ten kurs nie zmniejsza prawdopodobieństwa osiągnięcia sukcesu s w kursie K_b .

Zbiór wszystkich kursów wspierających danego kursu bazowego K_b określa się jako zbiór wspierający i oznacza przez W_{K_b} .

Definicja 1.15. (Kurs samowspierający, klasa samowspierająca) Kursem samowspierającym nazwiemy każdy kurs z puli kursów, dla którego prawdopodobieństwo osiągnięcia sukcesu s w kolejnym podejściu do niego (tzn. w kolejnym cyklu dydaktycznym) nie pogarsza się.

Zbiór wszystkich kursów samowspierających określa się jako klasę samowspierającą i oznacza przez \mathcal{S} .

Definicja 1.16. (Analiza wsparcia) Przez analizę wsparcia rozumie się całość procesów algorytmu samo(wspierania) prowadzących do selekcji tych par testowych, w których kurs testowy jest jednocześnie kursem wspierającym.

Definicja 1.17. (Analiza samowsparcia, samowsparcie) Przez analizę samowsparcia (samowsparcie) rozumie się proces wyznaczenia klasy kursów samowspierających.

1.3. Dane z systemu USOS

Dane historyczne stanowiące przedmiot badań przedstawionych w niniejszej pracy dotyczą studentów Wydziału MIMUW. Poniżej przedstawiono krótko źródło danych oraz ich opis.

1.3.1. Informacje ogólne o USOS

Uniwersytecki System Obsługi Studentów (USOS) jest systemem informatycznym, który ułatwia zarządzanie sprawami związanymi ze studiowaniem. Pierwsze wdrożenie systemu rozpoczęło się w 2000 roku na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego oraz Wydziale Matematyki Uniwersytetu Łódzkiego. Obecnie do projektu USOS przystąpiły 34 placówki, których listę można znaleźć na stronie projektu:

<http://usos.edu.pl/>.

Jednym z obszarów wspieranych przez USOS jest zarządzanie tokiem studiów, w tym przechowywanie protokołów z ocenami i zaliczeniami. Analiza tych właśnie danych stanowi główny aspekt niniejszej pracy.

1.3.2. Opis danych

Zestaw danych możliwych do wykorzystania dotyczył kursów udostępnianych przez Uniwersytet Warszawski, na które uczęszczali studenci Wydziału MIMUW w latach 1999 - 2010. Dane były zgromadzone w 6262 plikach o 13 atrybutach, których opis jest przedstawiony w Tabeli 1.2.

Nazwa	Opis
os_id	Numer identyfikacyjny studenta.
data_ur	Data urodzenia studenta.
plec	Płeć studenta.
cdyd_kod	Kod cyklu dydaktycznego.
prz_kod	Kod kursu.
nazwa	Nazwa kursu.
tpro_kod	Możliwe rozliczenie kursu.
prot_id	Numer identyfikacyjny protokołu.
term_prot_id	Numer identyfikacyjny terminu protokołu.
toc_kod	Zastosowane rozliczenie kursu.
wartosc	Uzyskany przez studenta wynik kursu.
opis	Opis uzyskanego wyniku kursu.

Tabela 1.2: Zestawienie atrybutów udostępnionych danych.

Każdy z dostarczonych plików danych odpowiadał różnemu kodowi kursu, co nie przekładało się na zróżnicowanie samych kursów. Występowały sytuacje, kiedy np. w wyniku nieznacznej modyfikacji programu kursu, informacje o nim były rozbite na kilka plików. Dzięki pracy wykonanej przez Karola Kańskiego było możliwe scalenie dostarczonych plików w taki sposób, by każdy plik wynikowy odpowiadał już danemu kursowi. W [7] można znaleźć opis algorytmu odpowiadającego za grupowanie danych.

Ostatecznie otrzymano 4954 plików (a zarazem kursów) możliwych do analizy.

1.3.3. Wybór analizowanych kursów

Analiza wszystkich kursów jednocześnie uniemożliwiłaby czytelne przedstawienie uzyskanych później wyników. Dlatego pierwszym etapem pracy było wyselekcjonowanie części z nich, która została później poddana badaniu.

Kursy wybrane na tym etapie składają się na pulę kursów (def 1.12). Ich dobór nastąpił w oparciu o program studiów stacjonarnych drugiego stopnia Wydziału MIMUW. Każdy student rozpoczynający studia na kierunku *Matematyka* musi wybrać jedną z proponowanych specjalizacji z określonego programu studiów i zrealizować pulę przedmiotów z nią związanych. Lista wszystkich programów i specjalizacji jest dostępna pod adresem:

http://www.mimuw.edu.pl/dla_kandydata.

Do analizy zostały włączone przedmioty związane ze specjalizacjami:

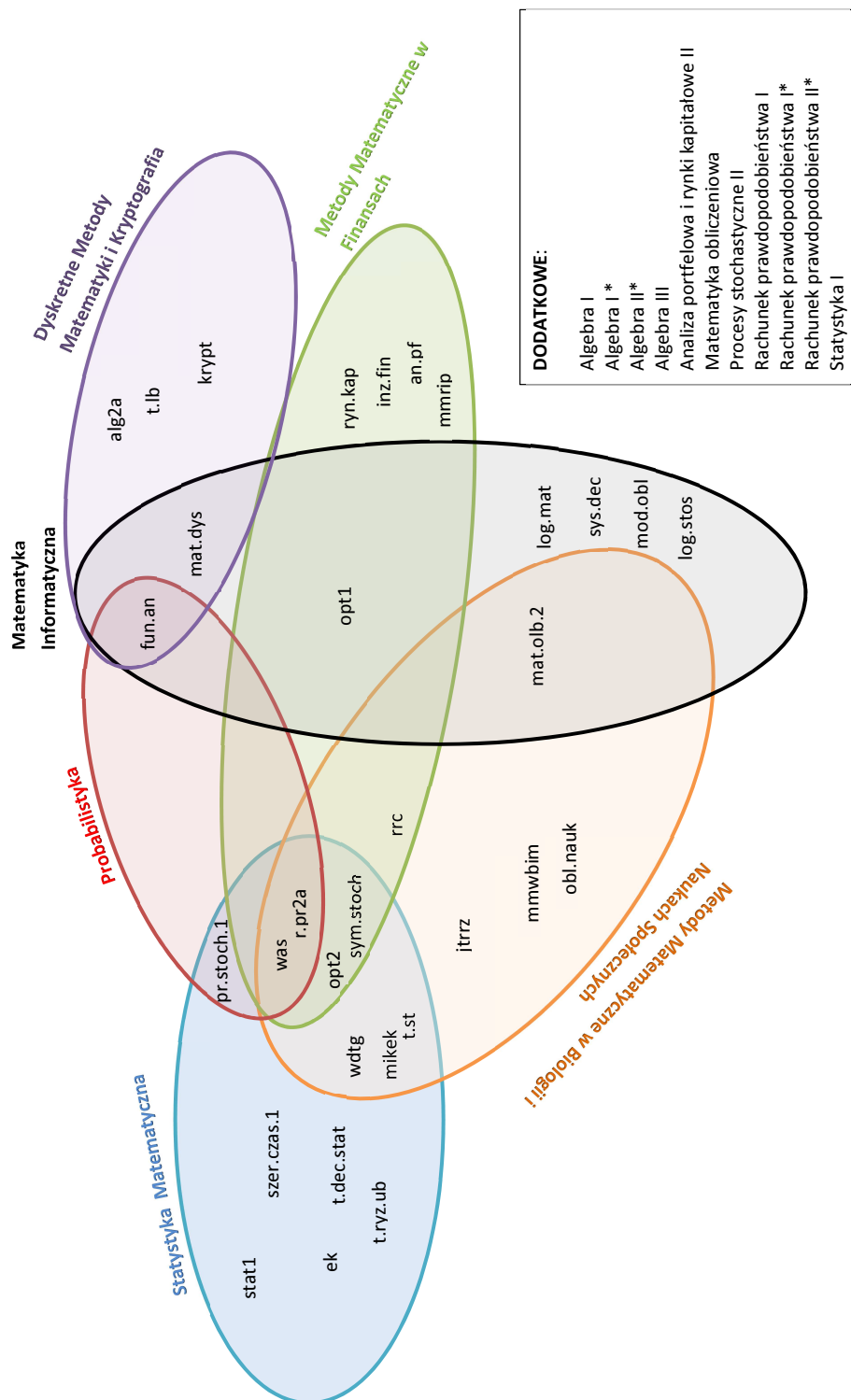
- Probabilistyka,
- Matematyka w Informatyce,
- Metody Matematyczne w Biologii i Naukach Społecznych,

- Statystyka Matematyczna,
- Metody Matematyczne w Finansach,
- Dyskretnie Metody Matematyki i Kryptografia,

a także kilka arbitralnie wybranych kursów, które w naturalny sposób je uzupełniają. Otrzymano w ten sposób 44 kursy, w których uczestniczyło łącznie 2103 studentów. Wykaz wszystkich przedmiotów przedstawiono w Tabeli 1.3. Rysunek 1.1 przedstawia przynależność wybranych kursów do danej specjalizacji.

Lp.	Nazwa kursu	Skrót	Plik źródłowy
1	Algebra I	alg1a	1000-113aAG1a.txt
2	Algebra I *	alg1b	1000-113aAG1_.txt
3	Algebra II	alg2a	1000-134AG2.txt
4	Algebra II *	alg2b	1000-134AG2_.txt
5	Algebra III	alg3	1000-135AG3.txt
6	Analiza portfelowa i rynki kapitałowe I	an.pfl	1000-135pk1.txt
7	Analiza portfelowa i rynki kapitałowe II	an.pf2	1000-1M00AP.txt
8	Ekonometria	ek	1000-135EKN.txt
9	Funkcje analityczne	fun.an	1000-134FAN.txt
10	Inżynieria finansowa	inz.fin	1000-135IFI.txt
11	Jakościowa teoria równań różniczkowych zwyczajnych	jtrrz	1000-135RRJ.txt
12	Kryptografia	krypt	1000-135KRG.txt
13	Logika matematyczna	log.mat	1000-135LOM.txt
14	Logika stosowana	log.stos	1000-135LOS.txt
15	Matematyka dyskretna	mat.dys	1000-611MAD.txt
16	Matematyka obliczeniowa	mat.obl1	1000-113aMOBa.txt
17	Matematyka obliczeniowa II	mat.obl2	1000-135MO2.txt
18	Mikroekonomia	miek	1000-135MIE.txt
19	Modele matematyczne rynków instrumentów pochodnych I	mmrip	1000-135IP1.txt
20	Modele matematyczne biologii i medycyny	mmwbim	1000-135MBM.txt
21	Modele obliczeń	mod.obl	1000-135MOL.txt
22	Obliczenia naukowe	obl.nauk	1000-135ONA.txt
23	Optymalizacja I	opt1	1000-134OP1.txt
24	Optymalizacja II	opt2	1000-135OP2.txt
25	Procesy stochastyczne I	pr.stoch1	1000-135PS1.txt
26	Procesy stochastyczne II	pr.stoch2	1000-135PS2.txt
27	Rachunek prawdopodobieństwa I	r.pr1a	1000-114aRP1a.txt
28	Rachunek prawdopodobieństwa I *	r.pr1b	1000-114aRP1_.txt
29	Rachunek prawdopodobieństwa II	r.pr2a	1000-135RP2.txt
30	Rachunek prawdopodobieństwa II*	r.pr2b	1000-115aRP2_.txt
31	Równania różniczkowe cząstkowe I	rrc	1000-135RC1a.txt
32	Rynki kapitałowe	ryn.kap	1000-135RKA.txt
33	Statystyka I	stat1	1000-135ST1.txt
34	Statystyka II	stat2	1000-135ST2.txt
35	Symulacje stochastyczne	sym.stoch	1000-135SST.txt
36	Systemy decyzyjne	sys.dec	1000-135SYD.txt
37	Szeregi czasowe I	szer.czas1	1000-135SC1.txt
38	Szeregi czasowe II	szer.czas2	1000-1M00SC.txt
39	Teoria decyzji statystycznych	t.dec.stat	1000-135TDS.txt
40	Teoria liczb	t.lb	1000-135TL1.txt
41	Teoria ryzyka w ubezpieczeniach	t.ryz.ub	1000-135TRU.txt
42	Teoria sterowania	t.st	1000-135TST.txt
43	Wstęp do analizy stochastycznej	was	1000-135WAS.txt
44	Wstęp do teorii gier	wdtg	1000-1M07TG.txt

Tabela 1.3: Lista wybranych kursów składających się na pulę kursów.



Rysunek 1.1: Przedstawienie zależności wybranych kursów od specjalizacji.

Rozdział 2

Sieci Bayesowskie

Pierwsza część Rozdziału 2.1 przedstawia ogólną koncepcję sieci Bayesowskich. Została opracowana głównie w oparciu o pierwszą część książki [6]. Ważne uzupełnienie stanowią pozycje [10, 8, 11]. Kolejna część (roz. 2.2) to praktyczne wykorzystanie sieci Bayesowskich do danych USOS w oparciu o pakiety `bnlearn` oraz `gRain` programu R - odpowiada na pytanie o możliwość znalezienia nowych zależności pomiędzy analizowanymi kursami. Zainteresowanych teorią i zastosowaniem sieci (samo)wspierających, które są niezależne od uzyskanych tutaj wyników, odsyła się do Rozdziałów 3 i 4.

2.1. Część teoretyczna

Sieci Bayesowskie (ang. *Bayesian Networks*) są graficznymi modelami służącymi do przedstawienia łącznego rozkładu prawdopodobieństwa w zbiorze zmiennych \mathcal{U} - wykorzystując regułę łańcucha można przedstawić ją w postaci grafu DAG (def. 1.8). Wierzchołki grafu reprezentują wówczas analizowane zmienne, a krawędzie bezpośrednie zależności pomiędzy nimi. Skierowanie krawędzi pozwala uchwycić kierunek tych zależności. Jeśli para wierzchołków jest ze sobą niepołączona, wówczas odpowiadające im zmienne są (warunkowo) niezależne (def. 1.2, 1.3). Dokładna definicja sieci Bayesowskich (def. 2.3) wywodzi się z koncepcji sieci przyczynowych, dlatego poświęcono im osobny Rozdział 2.1.1.

2.1.1. Sieci przyczynowe

Sieci przyczynowe (ang. *Causal Networks*) stanowią graficzny opis relacji pomiędzy zmiennymi w postaci grafu skierowanego i stanowią podstawę do pełnego zdefiniowania sieci Bayesowskich.

Wyróżnia się trzy podstawowe sposoby połączenia zmiennych w sieciach przyczynowych:

1. Połączenie łańcuchowe (ang. *Serial Connection*);

Połączenie łańcuchowe jest widoczne na Rysunku 2.1 a). W tym rodzaju połączenia zmienna A wpływa na zmienną B , która wpływa na zmienną C . Wobec powyższego uzyskanie informacji o konkretnym stanie pierwszej zmiennej (poprzez zmienną środkową) wpłynie na tę ostatnią. Taki rodzaj rozumowania nazywa się często w sieciach przyczynowych *wnioskowaniem przyczynowym*. Jest to rozumowanie zgodne z kierunkiem połączenia. Jeśli początkowa informacja będzie dotyczyć zmiennej C , będzie można przeprowadzić rozumowanie wsteczne, które również poprzez zmienną środkową wpłynie ostatecznie na zmienną A . Taki rodzaj wnioskowania określa się jako *diagnostyczne*, czyli przeciwnie do kierunku połączenia.

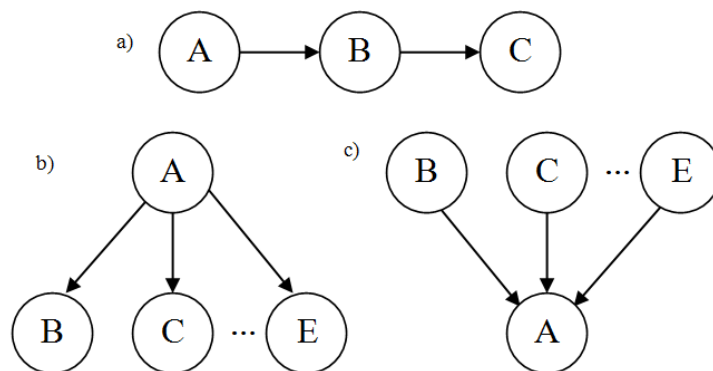
Pomimo braku bezpośredniego połączenia pomiędzy zmiennymi A i C , nie można o nich mówić, że są zmiennymi niezależnymi. Sytuacja zmieni się, gdy zmienna im pośrednicząca przyjmie określoną wartość - przepływ informacji zostanie zablokowany i wówczas staną się one zmiennymi (warunkowo) niezależnymi.

2. Połączenie zbieżne (ang. *Diverging Connection*);

Połączenie zbieżne jest widoczne na Rysunku 2.1 b). W tym rodzaju połączenia przepływ informacji pomiędzy zmiennymi B, C, \dots, E może odbywać się za pośrednictwem ich wspólnego rodzica - zmiennej A (def. 1.10). Przepływ będzie możliwy tylko wtedy, gdy ów zmienna nie przyjmie konkretnej wartości. W przeciwnym razie, podobnie jak w przypadku połączenia łańcuchowego, nastąpi (warunkowe) uniezależnienie się bezpośrednich następników zmiennej A (def. 1.10).

3. Połączenie rozbieżne (ang. *Converging Connection*);

Połączenie rozbieżne jest widoczne na Rysunku 2.1 c). Tutaj dopóty, dopóki nie będzie dostępna wiedza o wspólnym dziecku (zmienna A) rodzice (zmienne B, C, \dots, E) pozostaną zmiennymi niezależnymi. Wiedza dotycząca jednej z możliwych przyczyn danego zdarzenia nie mówi bowiem nic o pozostałych. Nie mniej, jeśli informacja o skutku będzie znana, wówczas pogląd na przyczyny może ulec zmianie.



Rysunek 2.1: Rodzaje połączeń występujące w sieciach przyczynowych - a) połączenie łańcuchowe b) połączenie zbieżne c) połączenie rozbieżne. Źródło: Na podstawie [6].

Sposób przepływu informacji w połączeniu rozbieżnym dobrze jest przeanalizować na rzeczywistym przykładzie - poniższy został zaczerpnięty z [6].

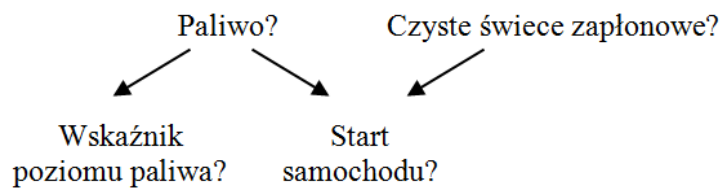
Przykład - problem uruchomienia samochodu

Poranna próba uruchomienia samochodu zakończyła się niepowodzeniem. Wykluczono jednak problem akumulatora, gdyż podczas przekręcania kluczyka w stacyjce można było usłyszeć dźwięk rozrusznika. Wobec tego za równie prawdopodobne przyczyny uznaje się:

- nocną kradzież paliwa,
- zanieczyszczenie świec zapłonowych.

Wykluczając chwilowo inne przyczyny, należy podjąć decyzję, której z nich należy przeciwdziałać w pierwszej kolejności. W tym celu można sprawdzić poziom wskaźnika paliwa - jeśli poziom jest wysoki, wówczas racjonalniej wydaje się zająć świecami zapłonowymi, gdyż prawdopodobieństwo kradzieży paliwa mocno spada (choć nie możemy jej całkowicie wykluczyć, z uwagi na przykład na zawieszenie się wskaźnika paliwa).

Sieć przyczynowa adekwatna do przedstawionego powyżej problemu znajduje się na Rysunku 2.2.



Rysunek 2.2: Sieć przyczynowa dla problemu uruchomienia samochodu. Źródło: Na podstawie [6].

Widać więc, że w odróżnieniu od połączenia łańcuchowego i rozbieżnego, w połączeniu zbieżnym informacje są przekazywane tylko wtedy, gdy jest dostępna wiedza o zmiennej i jednym z jej potomków (def. 1.11). Brak dodatkowej informacji dotyczącej wspólnego potomka blokuje przepływ informacji.

Powyższe informacje motywują wprowadzenie następujących definicji:

Definicja 2.1. (d-separacja, d-połączenie¹) Dwa wierzchołki A, B w sieci przyczynowej są d-separowane, jeśli dla każdej ścieżki pomiędzy A i B istnieje pośredniczący im wierzchołek $V \notin \{A, B\}$ t.ż.:

- połączenie pomiędzy nimi jest łańcuchowe lub rozbieżne i stan wierzchołka V jest znany
- lub
- połączenie pomiędzy nimi jest zbieżne oraz nie ma dodatkowych informacji o wierzchołku V oraz żadnym jego potomku.

Wierzchołki A i B , które nie są d-separowane nazywa się d-połączonymi.

Definicja 2.2. (Otoczka Markowa) Otoczka Markowa (ang. *Markov blanket*) dla zmiennej A to zbiór rodziców A , dzieci A i ich pozostałych rodziców.

Otoczka Markowa dla zmiennej A skupia zmienne, które przy ustalonych wartościach oddzielają tę zmienną od reszty sieci.

¹Pojawiające się w definicji 2.1 d pochodzi od ang. *directed graph*

2.1.2. Dokładna definicja sieci Bayesowskiej

W sieciach przyczynowych określa się własność d-separacji i d-połączenia (def. 2.1). W przypadku d-połączeń, naturalnym wydaje się być dodanie do tychże połączeń wartości liczbowych, obrazujących jakość wiązania. Wartości te są wyznaczone przez prawdopodobieństwa warunkowe dołączone do każdego wierzchołka. Wszystkie te informacje składają się na następującą definicję sieci Bayesowskich:

Definicja 2.3. (Sieć Bayesowska) Na sieć Bayesowską składają się następujące aspekty:

- zbiór zmiennych i zbiór krawędzi skierowanych pomiędzy zmiennymi,
- każda zmienna posiada skończoną, wzajemnie się wykluczającą przestrzeń stanów,
- zmienne razem z krawędziami tworzą graf DAG (def. 1.8),
- do każdej zmiennej A o zbiorze rodziców $pa(A) = \{B_1, \dots, B_n\}$ jest dołączona informacja o rozkładzie warunkowym $P(A|B_1, \dots, B_n)$.

W przypadku wierzchołków A , dla których $pa(A) = \emptyset$ informacje o rozkładzie warunkowym redukują się do informacji o rozkładzie danej zmiennej $P(A)$.

Należy zauważyć, że mimo iż definicja sieci Bayesowskich była motywowana rozważanymi wcześniej sieciami przyczynowymi, nie wymaga się, by połączenia odpowiadały związkowi przyczynowemu. Dlatego należy skonfrontować, czy występujące w sieci własności d-separacji są zgodne z analizowanym problemem - model nie powinien zawierać warunkowych niezależności, które nie znajdują odzwierciedlenia w rzeczywistości.

Główną zaletą sieci Bayesowskich jest fakt, że reguła łańcucha (def. 1.8) dla zbioru zmiennych \mathcal{U} przyjmuje skompresowaną postać (tw. 2.1). Dzięki temu wyznaczanie rozkładów brzegowych i warunkowych można znacząco uprościć.

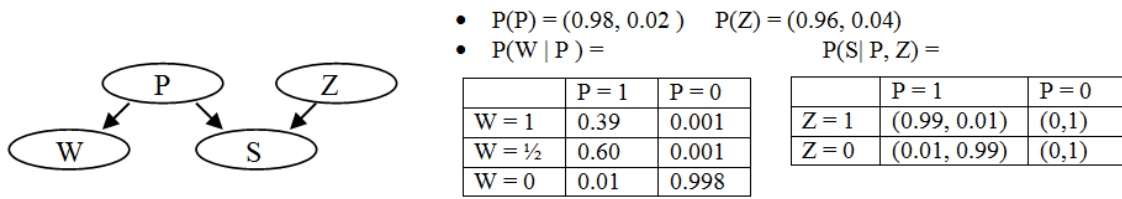
Twierdzenie 2.1 (Reguła łańcucha dla sieci Bayesowskich). Rozważmy sieć Bayesowską nad $\mathcal{U} = \{A_1, \dots, A_n\}$. Sieć ta określa łączny rozkład prawdopodobieństwa $P(\mathcal{U})$ jako produkt wszystkich rozkładów warunkowych w niej występujących:

$$P(\mathcal{U}) = \prod_{i=1}^n P(A_i|pa(A_i)),$$

gdzie $pa(A_i)$ określają zbiór rodziców A_i oraz \mathcal{U} oddaje własności rozważanej sieci Bayesowskiej, tzn:

1. Prawdopodobieństwa warunkowe w $P(\mathcal{U})$ muszą być określone w sieci;
2. Jeśli zmienne A, B są d-separowane w sieci pod warunkiem \mathcal{C} , to zmienne A, B są warunkowo niezależne pod warunkiem \mathcal{C} .

Dowód powyższego Twierdzenia można znaleźć w [6, s.37].



Rysunek 2.3: Sieć Bayesowska dla problemu uruchomienia samochodu - graf DAG i tabele CPT. Źródło: Na podstawie [6].

Przykład cd..

W Rozdziale 2.1.1 został opisany problem związany z uruchomieniem samochodu. Rysunek 2.2 przedstawiał odpowiadającą mu sieć przyczynową. W celu przedstawienia wspomnianego problemu w postaci sieci Bayesowskich należy określić odpowiednie tabele CPT (def. 1.1). Dla ustalenia wprowadza się następujące zmienne: $P \sim$ Paliwo, $W \sim$ Wskaźnik poziomu paliwa, $S \sim$ Start samochodu i $Z \sim$ Czyste świece zapłonowe, o przestrzeniach stanów $sp(P) = \{0, 1\}$, $sp(W) = \{0, \frac{1}{2}, 1\}$, $sp(S) = \{0, 1\}$ oraz $sp(Z) = \{0, 1\}$ odpowiednio. Rysunek 2.3 przedstawia komponenty pełnej sieci Bayesowskiej dla tego problemu.

Zgodnie z Twierdzeniem 2.1 rozkład łączny można obliczyć ze wzoru:

$$P(P, W, Z, S) = P(P)P(Z)P(W|P)P(S|P, Z). \quad (2.1)$$

2.1.3. Algorytm Junction Tree

Silą sieci Bayesowskich nie jest tylko możliwość przedstawienia rozkładu łącznego w zwartej postaci i przy dobrej wizualizacji. Reguła łańcucha dla sieci Bayesowskich (tw. 2.1) daje duże możliwości wyznaczania konfiguracji rozkładu w sytuacji, gdy pojawia się nowa informacja o danej zmiennej w sieci, tzw. ustalenie (ang. *finding*). Jednym z algorytmów przekazywania ustaleń jest algorytm *Junction Tree*. Algorytm skupia się na tym, by podczas procesu wyznaczania danego prawdopodobieństwa wybrać tę ścieżkę, w której łączne dziedziny potencjałów² rozważanych w danym kroku będą miały najmniejsze dziedziny.

Przykład

Rozważmy sieć Bayesowską nad zbiorem $\mathcal{U} = \{A_1, A_2, A_3, A_4, A_5, A_6\}$ o strukturze jak na Rysunku 2.4.

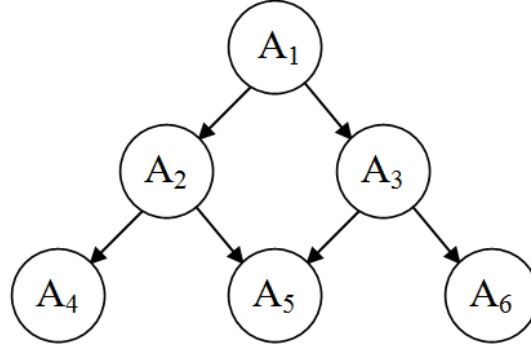
Zgodnie z Definicją 2.3 przedstawiona sieć wymaga określenia potencjałów przy każdej zmiennej ze zbioru \mathcal{U} , tzn. potencjałów $\phi_1(A_1)$, $\phi_2(A_2, A_1)$, $\phi_3(A_3, A_1)$, $\phi_4(A_4, A_2)$, $\phi_5(A_5, A_2, A_3)$, $\phi_6(A_6, A_3)$. Załóżmy, że chcemy wyznaczyć $P(A_4)$. Zauważmy najpierw, że z reguły łańcucha dla sieci Bayesowskich otrzymuje się, że łączny rozkład zbioru zmiennych dla rozważanego zbioru \mathcal{U} można zapisać jako:

$$P(\mathcal{U}) = \phi_1 \phi_2 \phi_3 \phi_4 \phi_5 \phi_6. \quad (2.2)$$

Szukane prawdopodobieństwo $P(A_4)$ otrzymuje się przez zsumowanie prawej strony równania 2.2 w ustalonej kolejności.

Rozważmy następujące kolejności sumowania:

²Definiując sieci Bayesowskie określa się prawdopodobieństwa warunkowe $P(A|pa(A))$, które można rozumieć jako funkcje $\phi : pa(A) \cup \{A\} \rightarrow [0; 1]$ nazywane *potencjałami*.



Rysunek 2.4: Struktura sieci Bayesowskiej nad zbiorem $\mathcal{U} = \{A_1, A_2, A_3, A_4, A_5, A_6\}$. Źródło: Na podstawie [6].

1. Kolejność sumowania A_1, A_2, A_3, A_5, A_6 :

$$\begin{aligned}
 P(A_4) &= \sum_{A_1, A_2, A_3, A_5, A_6} \phi_1(A_1) \phi_2(A_2, A_1) \phi_3(A_3, A_1) \phi_4(A_4, A_2) \phi_5(A_5, A_2, A_3) \phi_6(A_6, A_3) = \\
 &= \sum_{A_1} \phi_1(A_1) \sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \sum_{A_3} \phi_3(A_3, A_1) \underbrace{\sum_{A_5} \phi_5(A_5, A_2, A_3)}_{\phi'_5(A_2, A_3)} \underbrace{\sum_{A_6} \phi_6(A_6, A_3)}_{\phi'_6(A_3)} = \\
 &= \sum_{A_1} \phi_1(A_1) \sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \underbrace{\sum_{A_3} \phi_3(A_3, A_2) \phi'_5(A_2, A_3) \phi'_6(A_3)}_{\phi'_3(A_1, A_2)} = \\
 &= \sum_{A_1} \phi_1(A_1) \underbrace{\sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \phi'_3(A_1, A_2)}_{\phi'_2(A_1, A_4)} = \sum_{A_1} \phi_1(A_1) \phi'_2(A_1, A_2).
 \end{aligned}$$

2. Kolejność sumowania A_1, A_2, A_5, A_6, A_3 :

$$\begin{aligned}
 P(A_4) &= \sum_{A_1, A_2, A_5, A_6, A_3} \phi_1(A_1) \phi_2(A_2, A_1) \phi_3(A_3, A_1) \phi_4(A_4, A_2) \phi_5(A_5, A_2, A_3) \phi_6(A_6, A_3) = \\
 &= \sum_{A_1} \phi_1(A_1) \sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \underbrace{\sum_{A_5} \sum_{A_6} \sum_{A_3} \phi_3(A_3, A_1) \phi_5(A_5, A_2, A_3) \phi_6(A_6, A_3)}_{\phi'_3(A_1, A_2, A_5, A_6)} = \\
 &= \sum_{A_1} \phi_1(A_1) \sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \underbrace{\sum_{A_5} \sum_{A_6} \phi'_3(A_1, A_2, A_5, A_6)}_{\phi'_6(A_1, A_2, A_5)} =
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{A_1} \phi_1(A_1) \sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \underbrace{\sum_{A_5} \phi'_6(A_1, A_2, A_5)}_{\phi'_5(A_1, A_2)} = \\
&\sum_{A_1} \phi_1(A_1) \underbrace{\sum_{A_2} \phi_2(A_2, A_1) \phi_4(A_4, A_2) \phi'_5(A_1, A_2)}_{\phi'_2(A_1, A_4)} = \sum_{A_1} \phi_1(A_1) \phi'_2(A_1, A_4).
\end{aligned}$$

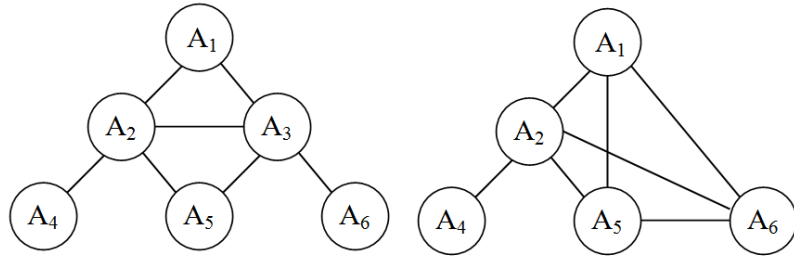
Obliczając rozkład $P(A_4)$ w 1.kolejności maksymalny rozmiar sumy potencjałów w poszczególnych krokach wynosi 3, podczas gdy w 2.kolejności wynosi on 4. Wybór odpowiedniej kolejności obliczeń, tzw. *porządku eliminacji*, może spowodować redukcję rozmiaru tablic niezbędnych do ich wykonania.

Różne wymiary łącznych dziedzin potencjałów rozważanych w danych krokach wyznaczania rozkładu brzegowego można wytłumaczyć się w ujęciu *grafu dziedzinowego*.

Definicja 2.4. (Graf dziedzinowy) Niech $\Phi = \{\phi_1, \dots, \phi_n\}$ będzie zbiorem potencjałów zdefiniowanych nad zbiorem zmiennych \mathcal{U} o dziedzinach $\text{dom}(\phi_i) = \mathcal{D}_i$. Grafem dziedzinowym dla Φ nazywa się graf prosty nieskierowany o wierzchołkach odpowiadających \mathcal{U} i krawędziach pomiędzy każdą parą zmiennych z pochodzących z tej samej dziedziny \mathcal{D}_i .

W kontekście sieci Bayesowskich dziedziny potencjałów są postaci $\mathcal{D}_\phi = \text{pa}(A) \cup \{A\}$. Wobec Definicji 2.4 zbudowanie grafu dziedzinowego ogranicza się więc do połączenia niepołączonych do tej pory zmiennych mających wspólne dziecko oraz zamiany krawędzi skierowanych na nieskierowane. Dlatego proces wyznaczania grafów dziedzinowych w kontekście sieci Bayesowskich nazywa się często *umoralnianiem*.

Podczas eliminacji zmiennej X pracuje się na potencjałach, które w swoich dziedzinach zawierają tę zmienną - dziedzina iloczynu tych potencjałów zawiera zmienną X i jej sąsiadów ($\text{nb}(X)$) z grafu dziedzinowego. Kiedy zmienna zostanie wyeliminowana otrzymuje się nowy potencjał, którego dziedziną jest zbiór $\text{nb}(X)$. Potencjały użyte do jego wyznaczenia (czyli zawierające w swojej dziedzinie X) zostają usuwane. Graf dziedzinowy nowego zbioru potencjałów powstaje przez usunięcie z grafu wierzchołka X wraz z jego krawędziami i dodaniu krawędzi pomiędzy wszystkimi jego sąsiadami. Rysunek 2.5 przedstawia umoralnienie grafu z rozważanego wyżej przykładu i graf dziedzinowy powstały po usunięciu wierzchołka A_3 .



Rysunek 2.5: Umoralnienie sieci Bayesowskiej z rysunku 2.4 i graf dziedzinowy powstały w wyniku usunięcia wierzchołka A_3 . Źródło: Na podstawie [6].

Jak widać, drugi graf nie stanowi podgrafu pierwszego z nich - musiały zostać dołączone do niego krawędzie uzupełniające (ang. *fill-ins*). W ujęciu wyznaczania prawdopodobieństwa

$P(A_4)$ zgodnie z 2. porządkiem eliminacji powoduje to zwiększenie dziedziny potencjałów. Optymalnym porządkiem eliminacji byłby taki porządek, który nie spowodowałby wprowadzenia krawędzi uzupełniających. Ciąg eliminacji o takiej własności nazywa się *doskonałym ciągiem eliminacyjnym*.

Wspomniane zbiory $nb(X)$ odpowiadają sumie dziedzin potencjałów wykorzystywanych podczas eliminacji zmiennej X . Dziedziny odpowiadające eliminacji w ustalonym porządku eliminacyjnym są elementami tzw. *zbioru dziedzinowego*³. Okazuje się, że doskonale ciągi eliminacyjne związane ze zbiorem potencjałów Φ posiadają te same zbiory dziedzinowe, które odpowiadają zbiorowi klik (def. 1.9) wyznaczonych dla grafu dziedzinowego Φ . Algorytm *Junction Tree* wyznacza doskonały ciąg eliminacyjny właśnie w oparciu o kliki grafu umoralnionego. Kluczowym elementem algorytmu jest zorganizowanie klik w tzw. *Join Tree*, z którego łatwo odczytuje się, które wierzchołki można usunąć bez pojawienia się krawędzi uzupełniających. Następnie, w oparciu o *Join Tree* konstruuje się *Junction Tree*, które przechowuje informacje o sposobie przekazywania dodatkowych ustaleń o zmiennych. Szczegółowy opis algorytmu można znaleźć w [6].

2.1.4. Estymacja parametrów sieci Bayesowskich

Sieci Bayesowskie są wykorzystywane również w sytuacjach, gdy znany jest graf zależności pomiędzy modelowanymi zmiennymi (skonstruowany np. w oparciu o wiedzę ekspercką), lecz nie są wyznaczone rozkłady warunkowe (parametry sieci). Wówczas w oparciu o bazę danych korespondującą z badanym problemem parametry te mogą zostać estymowane.

Po więcej informacji odsyła się do rozdziału 6. I części [6] lub do 5. rozdziału [10]. Można tam znaleźć rozważania dotyczące zarówno sytuacji, w których dane użyte do estymacji parametrów są kompletne, jak i w których występują braki danych.

2.1.5. Wyznaczenie struktury sieci Bayesowskich

Innym zadaniem związanym z sieciami Bayesowskimi jest wyznaczenie struktury sieci Bayesowskiej na podstawie zbioru danych.

Należy zauważyć, że liczba możliwych do wyznaczenia sieci Bayesowskich zależnych od n zmiennych rośnie szybciej niż eksponentalnie. Przykładowe wyniki zostały przedstawione w Tabeli 2.1.

Do wyznaczania struktur sieci Bayesowskich stosuje się różne metody, które można podzielić na:

1. **Metody *constraint-based*** - struktura sieci jest wyznaczona zgodnie z warunkowymi niezależnościami, jakie występują w zbiorze danych;
2. **Metody *score-based*** - struktura sieci jest wyznaczona w oparciu o szukanie najlepszego kandydata w oparciu o wybraną *funkcję oceny*, która sprawdza stopień dopasowania danych w stosunku do rozważanej struktury.

2.1.6. Wybrane pakiety programu R dedykowane sieciom Bayesowskim

Sieci Bayesowskie cieszą się w ostatnich latach sporym zainteresowaniem. Skutkiem tego jest zwiększająca się lista oprogramowań ułatwiających pracę z nimi na różnym poziomie

³Jeśli ciąg eliminacyjny skutkowałby wyznaczeniem dziedzin \mathcal{D}_i , \mathcal{D}_j , t.ż. $\mathcal{D}_i \subsetneq \mathcal{D}_j$, wówczas zbiór \mathcal{D}_i nie zostanie uwzględniony w zbiorze dziedzinowym.

n	# DAG	n	# DAG	n	# DAG
1	1	9	$1.2 \cdot 10^{15}$	17	$6.3 \cdot 10^{52}$
2	3	10	$4.2 \cdot 10^{18}$	18	$9.3 \cdot 10^{58}$
3	25	11	$3.2 \cdot 10^{22}$	19	$3.3 \cdot 10^{65}$
4	543	12	$5.2 \cdot 10^{26}$	20	$2.35 \cdot 10^{72}$
5	29281	13	$1.9 \cdot 10^{31}$	21	$3.5 \cdot 10^{79}$
6	$3.8 \cdot 10^6$	14	$1.4 \cdot 10^{36}$	22	$1.1 \cdot 10^{87}$
7	$1.1 \cdot 10^9$	15	$2.4 \cdot 10^{41}$	23	$7.0 \cdot 10^{94}$
8	$7.8 \cdot 10^{11}$	16	$8.4 \cdot 10^{46}$	24	$9.4 \cdot 10^{102}$

Tabela 2.1: Zależność ilości struktur sieci Bayesowskich (# DAG) od liczby zmiennych (n).
Źródło: Na podstawie [6].

- począwszy od propagowania informacji w sparametryzowanej już sieci, poprzez uczenie parametrów, aż do wyszukiwania samej struktury sieci. Część z nich to pakiety programu R. Można tutaj wymienić m.in.:

- **gRain** - pakiet autorstwa Søren Højsgaard. Związany przede wszystkim z sieciami, co do których posiadana jest wiedza ekspercka dotycząca jej struktury. Zaimplementowany algorytm rozpowszechniania informacji jest zbliżony do algorytmu *Junction Tree* opisanego pokrótce w Rozdziale 2.1.3. Zestawienie głównych funkcji zostało przedstawione w Tabeli 2.2. Szczegółowy opis pakietu można znaleźć w [4].

Nazwa	Opis
cptable	Tworzy tabele CPT.
compileCPT	Wstępnie przetwarza listę tabel CPT w celu uzyskania obiektu klasy cptspec , wykorzystywanego później do utworzenia sieci.
grain	Tworzy sieć Bayesowską w oparciu o obiekt klasy cptspec .
querygrain	Daje możliwość zapytania sieci o prawdopodobieństwo łączne, brzegowe bądź warunkowe dla danych wierzchołków sieci. Wynikiem jest lista odpowiednich tabel CPT.
setFinding	Koduje w sieci informację o tym, że zmienne przyjęły ustalone wartości.
getFinding	Informuje użytkownika, jaka informacja została zakodowana w sieci, łącznie z prawdopodobieństwem jej wystąpienia.
pFinding	Zwraca wyłącznie prawdopodobieństwo wystąpienia zakodowanej w sieci informacji.
compile	Tworzy reprezentację grafu w postaci klik i łączy je w drzewo <i>Junction Tree</i> .

Tabela 2.2: Główne funkcje pakietu **gRain**.

- **bnlearn** - pakiet autorstwa Marco Scutari. Służy głównie do wyznaczenia struktury sieci. Autor zaimplementował w nim wybrane algorytmy zarówno z metod constraint-based jak i score-based. Dokładną dokumentację tego pakietu można znaleźć w [15]. Zestawienie głównych funkcji zostało przedstawione w Tabeli 2.3.

Nazwa	Opis
<code>hc</code>	Wyznacza strukturę sieci w oparciu o algorytm <code>hill-climbing</code> zwracając obiekt <code>bn</code> . Opis pozostałych zaimplementowanych algorytmów można znaleźć w [15]. Każdy z nich w wyniku zwraca obiekt <code>bn</code> .
<code>bn.fit</code>	Wyznacza parametry sieci Bayesowskiej w oparciu o zadaną strukturę w postaci obiektu <code>bn</code> i zbiór danych.
<code>compare</code>	Porównuje dwie struktury sieci Bayesowskich zakodowane w obiektach <code>bn</code> .
<code>set.arc</code>	Pozwala na eksperycką korektę modelu poprzez dodanie/usunięcie/zmianę kierunku krawędzi.
<code>modelstring</code>	Zwraca postać struktury modelu w przypadku struktury skierowanej.
<code>directed</code>	Zwraca informację o tym, czy struktura stanowi graf skierowany.
<code>acyclic</code>	Zwraca informację o tym, czy struktura stanowi graf acykliczny.
<code>directed.arcs/ undirected.arcs</code>	Zwraca skierowane/nieskierowane krawędzie z danej struktury w postaci obiektu <code>matrix</code> .
<code>parents/children</code>	Dla ustalonego wierzchołka sieci generuje wektor ich rodziców/ dzieci.
<code>mb</code>	Dla ustalonego wierzchołka sieci generuje wektor stanowiący jego otoczkę Narkova.
<code>as.grain</code>	Funkcja umożliwia konwertowanie obiektów klasy <code>bn</code> na obiekty klasy <code>grain</code> .

Tabela 2.3: Główne funkcje pakietu `bnlearn`.

Obydwa pakiety są dostępne w repozytorium CRAN programu R:

<http://cran.r-project.org/>.

2.2. Wykorzystanie sieci Bayesowskich do danych USOS

Podstawową informacją zawartą w danych USOS (roz. 1.3) jest wiedza o kursach realizowanych przez poszczególnych studentów Wydziału MIMUW. Można ją zakodować w postaci binarnej macierzy $M \in M_{4345 \times 44}$, gdzie 4345 odpowiada liczbie studentów uwzględnionych w bazie. Analiza macierzy M może doprowadzić do wyznaczenia nowych bloków tematycznych, obok tych związanych ze specjalizacjami oferowanych w ramach studiów (roz. 1.3.3). W tym celu można skonstruować sieć Bayesowską wykorzystując funkcje pakietów `bnlearn` i `grain` opisanych w Rozdziale 2.1.6. Dla ustalenia uwagi, sieć Bayesowską dla danych USOS nazywa się dalej *USOS-siecią*.

2.2.1. Wyczenie struktury *USOS-sieci*

Pierwszym krokiem związanym z wykorzystaniem sieci Bayesowskich do danych zgromadzonych w macierzy M jest dobór odpowiedniej struktury, która najlepiej odda jej charakter. W tym celu został wykorzystany algorytm zachłanny zaimplementowany w funkcji

`bnlearn::hc`. Poniżej przedstawiono podsumowanie zaproponowanej przez algorytm struktury dla *USOS*-sieci:

Bayesian network learned via Score-based methods

```

model:
  [alg1b][r.pr1b|alg1b] ... [mat.dys|krypt:log.mat:r.pr1a]
nodes:                                44
arcs:                                 116
  undirected arcs:                     0
  directed arcs:                       116
average markov blanket size:          7.41
average neighbourhood size:           5.27
average branching factor:             2.64

learning algorithm:                   Hill-Climbing
score:                                Bayesian Information Criterion
penalization coefficient:             4.188391
tests used in the learning procedure: 107199
optimized:                            TRUE

```

Z powyższego komunikatu można odczytać, że struktura grafu wynikowego zawiera 116 skierowanych krawędzi. Na każdy wierzchołek przypada średnio 5,27 sąsiadów. Struktura została wyznaczona po przeprowadzeniu 107199 testów - ich liczba jest związana z ilością iteracji algorytmu - jednym z możliwych do ustawienia parametrów funkcji `hc` jest `restart`, który określa liczbę losowych startów programu; jako że `hc` jest algorytmem szukającym najlepszych rozwiązań w ujęciu lokalnym począwszy od wygenerowanego grafu losowego zwiększenie parametru `restart` pozwala na odcięcie się lokalności rozwiązania.

2.2.2. Estymacja parametrów *USOS*-sieci

Po wyuczeniu struktury *USOS*-sieci nastąpiła estymacja jej parametrów, z wykorzystaniem funkcji `bnlearn::bn.fit`. Dzięki temu uzyskana sieć Bayesowska dla macierzy M stanowi kompletną strukturę w myśl Definicji 2.3. Z użyciem funkcji `bnlearn::as.grain` przekonwertowano ją na obiekt klasy `grain`, aby można było sprawnie komunikować się z siecią po wprowadzeniu odpowiednich ustaleń dla zadanych zmiennych.

2.2.3. Rezultaty

Jednym ze sposobów wyznaczenia wspomnianych bloków tematycznych jest skonstruowanie mapy ciepła w oparciu o *macierz ciepła* $MC \in M_{44 \times 44}([0; 1])$, wyznaczoną na podstawie *USOS*-sieci. Postać macierzy MC jest następująca:

$$MC_{kurs_i, kurs_j} = \begin{cases} P(kurs_j = 1 | kurs_i = 1), & \text{gdy } i \neq j, \\ NA, & \text{wp.p.} \end{cases} \quad (2.3)$$

Zgodnie z równaniem 2.3, każdy wiersz macierzy ciepła koduje informację o tym, z jakim prawdopodobieństwem wybierane były pozostałe kursy, w sytuacji gdy ogranicza się rozważania do podprzestrzeni związanej z ustalonym kursem, związanym z tym wierszem.

Wprowadzając odpowiednie ustalenia do wyznaczonej *USOS*-sieci, macierz *MC* można automatycznie wygenerować wykorzystując pakiet *grain*.

Rysunek 2.6 przedstawia ostateczną mapę ciepła - białe kolory odpowiadają wartościom *NA* z macierzy *MC*. Na załączonym Rysunku można łatwo odnaleźć kursy realizowane przez większość studentów każdego z rozważanych przedmiotów - są to w większości kursy stanowiące podstawę kształcenia na Wydziale MIMUW, jak *Funkcje analityczne*, *Równania Różniczkowe Cząstkowe*, *Algebra I*, czy *Rachunek prawdopodobieństwa I,II*. Trzy ostatnie przedmioty są rzadziej wybierane przez słuchaczy kursów nurtu * - jest to związane z faktem, że studenci są zobligowani do wyboru tylko jednego nurtu w ramach przedmiotu. Pozostając przy nurtach z * widać, że studenci, którzy już raz zdecydowali się uczestniczyć bardziej zaawansowany kursie, decydują na to również przy okazji innych kursów.

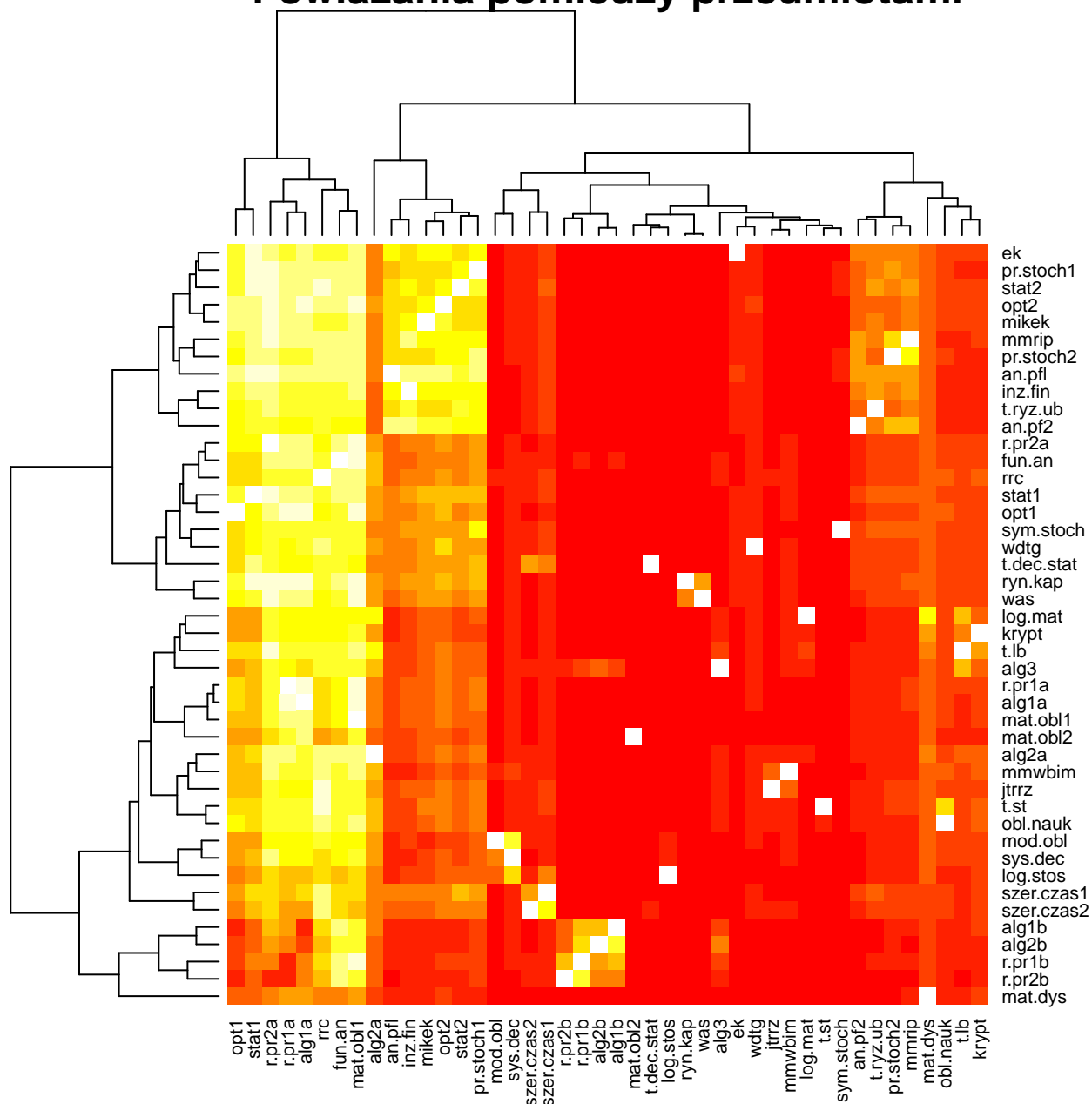
Należy zauważyć, że powyższy rezultat odnosi się do sytuacji, w której każdorazowo do *USOS*-sieci było wprowadzane ustalenie odnoszące się do konkretnego kursu. Sieci Bayesowskie dają możliwość ustalenia wartości kilku zmiennych jednocześnie, co wychodzi na przeciw realnym zapotrzebowaniom studentów. Propagując mnogie ustalenia w sieci można w sposób dynamiczny obserwować, z jakim prawdopodobieństwem były wybierane przedmioty wśród studentów o podobnych zainteresowaniach. Odnosząc się do podziału na specjalizacje (rys. 1.1), można zauważyć, że studenci specjalizacji Probabilistyka nie mają narzuconej dużej ilości obowiązkowych do realizacji kursów. Wobec tego ta grupa studentów mogłaby być szczególnie zainteresowana wskazaniem kursów, na których spotkają studentów o podobnym profilu. Dobrą ilustrację do tego stanowią grafy przedstawione na Rysunkach 2.7 i 2.8.

Pierwszy graf przedstawia strukturę sieci Bayesowskiej wyuczonej dla *USOS*-sieci. Dołączone do wierzchołków grafu wykresy słupkowe stanowią wizualizację ich rozkładów brzegowych. Dzięki temu widać, że obserwacja wysnuta podczas analizy mapy ciepła z Rysunku 2.6 znajduje swoje potwierdzenie - kursy takie jak *Algebra I*, *Rachunek Prawdopodobieństwa I* czy *Równania Różniczkowe Cząstkowe* są zdecydowanie częściej wybierane w stosunku do pozostałych kursów. Należy tutaj przypomnieć, że występujące połączenia rodzic-dziecko (def. 1.10) nie niosą za sobą informacji, że realizując zmienną będącą rodzicem prawdopodobieństwo uczestniczenia w zmiennej będącej dzieckiem wzrośnie. Informacje o sposobie zaburzenia początkowego rozkładu są zakodowane w niewidocznych tutaj tablicach CPT. Aby przeanalizować wpływ danej grupy kursów na pozostałe należy wprowadzić do sieci odpowiednie ustalenia, a następnie obliczyć ponownie parametry sieci. Drugi z załączonych grafów odnosi się właśnie do sytuacji, w których poinformowano sieć, że zostały już zrealizowane kursy związane ze specjalizacją Rachunku Prawdopodobieństwa, tzn: *Wstęp do Analizy Stochastycznej*, *Rachunek Prawdopodobieństwa II*, *Procesy Stochastyczne I* i *Funkcje Analityczne*. Zakładając, że studenci są zainteresowani realizacją kursów z zamieszczonej na grafie podpuli, mogą zwiększyć świadomość swojego wyboru. Przykładowo widać, większość studentów związanych z tą specjalizacją realizuje również *Statystykę I* czy *Mikroekonomię*. Jakkolwiek pierwszy z wymienionych kursów stanowi podstawę programową dla studiów I-go stopnia dla Wydziału MIMUW, tak drugi z nich jest w tym kontekście przedmiotem typowo fakultatywnym. Warto więc rozważyć jego realizację.

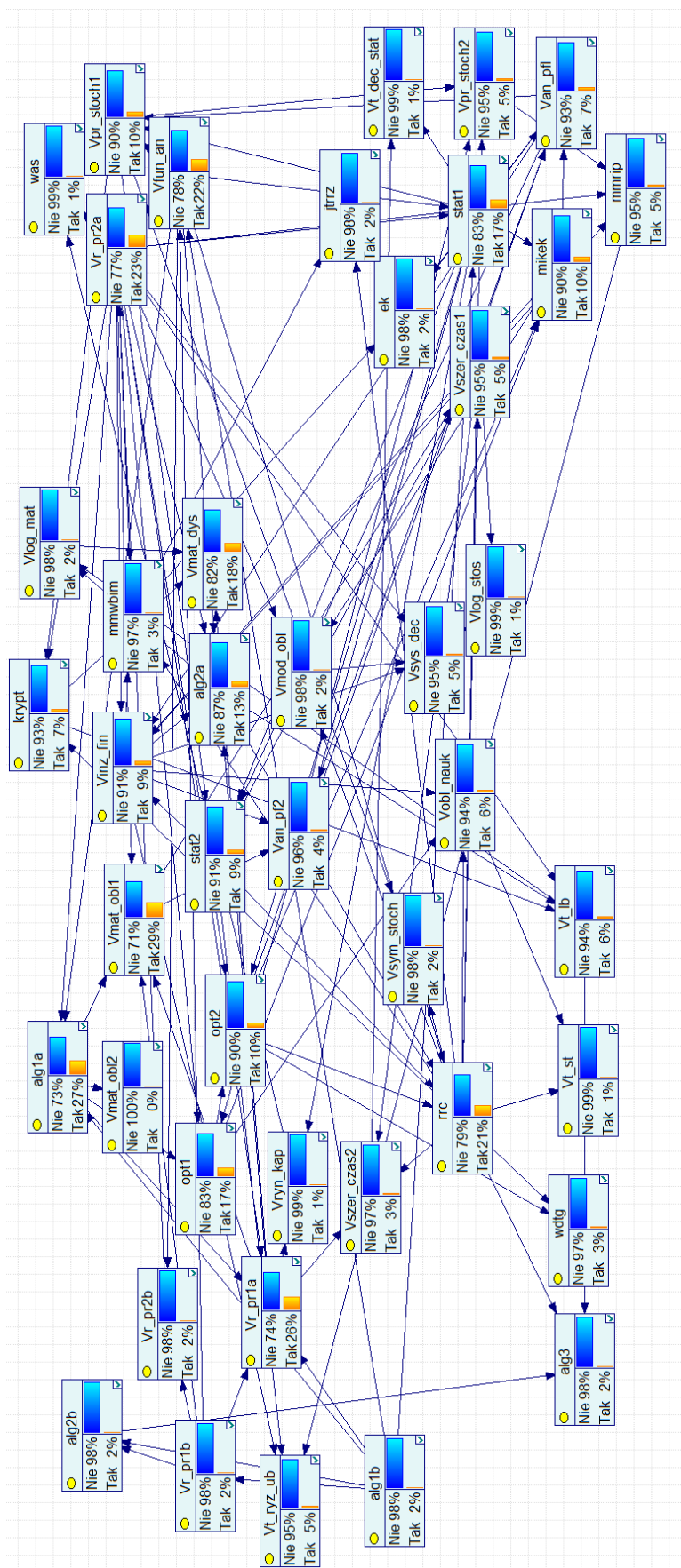
Grafy z Rysunków 2.7 i 2.8 wykonano w programie *GeNIe*, opartym na licencji *The Decision Systems Laboratory of the University of Pittsburgh*. Program jest dostępny pod adresem:

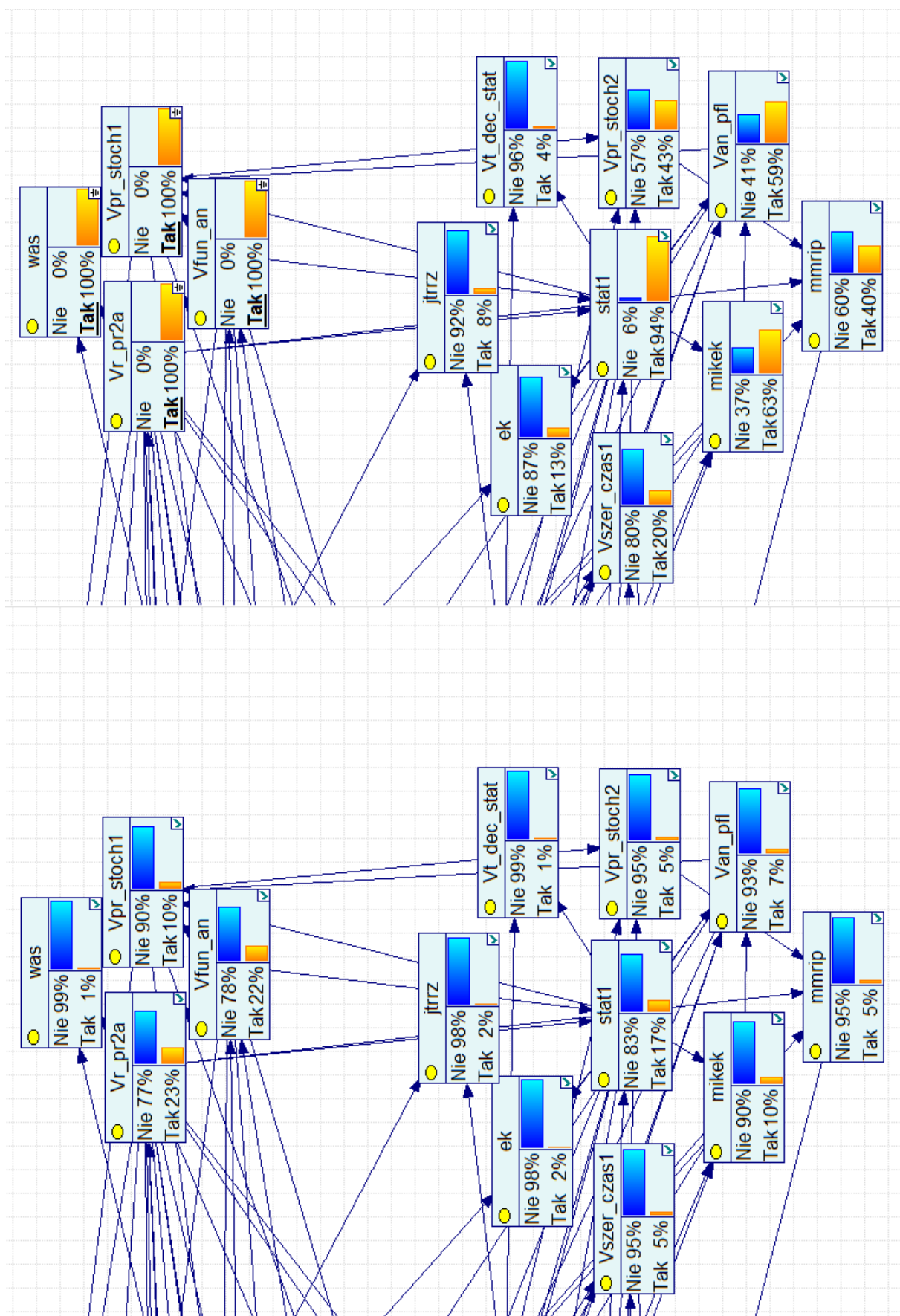
<http://genie.sis.pitt.edu/>.

Powiązania pomiędzy przedmiotami



Rysunek 2.6: Mapa ciepła dla *USOS*-sieci.





Rysunek 2.8: Wybrane elementy *USOS*-sieci. Schemat wykonany w programie GeNIe.

Rozdział 3

Sieci (samo)wspierania

W tym rozdziale przedstawiono teorię związaną z sieciami (samo)wspierania, których koncepcja została opisana w Rozdziale 1.2. Algorytm stanowiący trzon mechanizmu wyznaczania sieci (samo)wspierania został podzielony na dwie główne składowe: analizę wsparcia i samowsparcia. Zostały im poświęcone Rozdziały 3.2 i 3.3 odpowiednio. Początek niniejszego rozdziału (roz. 3.1) odnosi się do sposobu przetworzenia oryginalnych danych do postaci wejściowej algorytmu. Zakończenie rozdziału (roz. 3.4) opisuje metodę podziału studentów na studentów matematyki bądź informatyki. Informacja ta została wykorzystana podczas konstrukcji ostatecznych sieci (samo)wspierających. Uzyskane wyniki zostały zaprezentowane w Rozdziale 4.

3.1. Przygotowanie danych

Opis danych dostarczonych do analizy został przedstawiony w Rozdziale 1.3. Dane składały się z 12 atrybutów (tab. 1.2). Z punktu widzenia sieci (samo)wspierających istotne były tylko 4 z nich:

- *os_id* - dzięki niemu była możliwa identyfikacja studentów,
- *dyd.kod* - kluczowy atrybut z punktu widzenia analizy - dzięki niemu możliwe było uwzględnienie następstwa czasowego podczas analizy ustalonej pary testowej,
- *nazwa* - dzięki niemu była możliwa identyfikacja kursu,
- *wartosc* - kluczowy atrybut z punktu widzenia analizy - dzięki niemu możliwe było ustalenie, czy sukces *s* został osiągnięty przez danego studenta.

Z uwagi na binarny charakter rozważanych sukcesów, sukcesu pozytywnej oceny i sukcesu oceny ponad*przeciętnej, (roz. 1.2), atrybut *wartosc* należało przekodować na odpowiednie zmienne zero-jedynkowe. W obrębie ustalonego kursu odbyło się to za pomocą następującego przekształcenia:

$$s_{poz.ocena}(os) = \begin{cases} 0, & \text{gdy } w(os) = 2, \\ 1, & \text{gdy } w(os) \in \{3, 3.5, 4, 4.5, 5\}. \end{cases} \quad (3.1)$$

$$s_{p*przecietna}(os) = \begin{cases} 0, & \text{gdy } w(os) \leq \hat{w}, \\ 1, & \text{gdy } w(os) > \hat{w}. \end{cases} \quad (3.2)$$

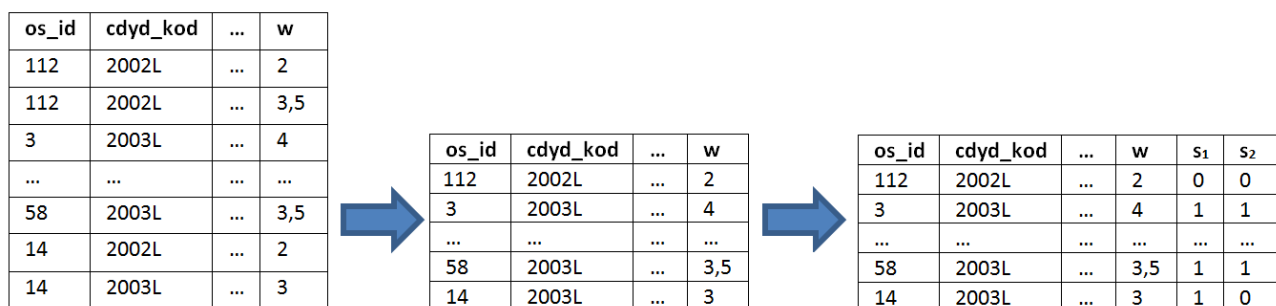
Wartość \hat{w} różnicuje uzyskane przez studentów oceny w celu uzyskania zbiorów zbliżonych do równolicznych. Jej określenie odbywa się już na etapie, w którym odfiltrowano obserwacje niezwiązane z danym sukcesem. W Tabeli 3.1 przedstawiono otrzymane wartości różnicujące dla kursów z puli kursów.

kurs	\hat{w}	kurs	\hat{w}	kurs	\hat{w}	kurs	\hat{w}
alg1a	3.00	krypt	3.50	opt1	3.00	stat2	3.00
alg1b	4.00	log.mat	3.00	opt2	3.50	sym.stoch	4.00
alg2a	3.00	log.stos	3.50	pr.stoch1	3.00	sys.dec	4.50
alg2b	4.00	mat.dys	3.00	pr.stoch2	3.00	szer.czas1	4.00
alg3	3.50	mat.obl1	3.00	r.pr1a	3.00	szer.czas2	4.00
an.pfl	3.50	mat.obl2	4.00	r.pr1b	4.00	t.dec.stat	3.50
an.pf2	3.50	mikek	4.00	r.pr2a	3.00	t.lb	3.50
ek	3.00	mmrip	3.50	r.pr2b	4.50	t.ryz.ub	3.00
fun.an	3.00	mmwbim	3.50	rrc	3.00	t.st	3.50
inz.fin	3.00	mod.obl	4.00	ryn.kap	3.00	was	3.00
jtrrz	3.50	obl.nauk	3.50	stat1	3.00	wdtg	3.50

Tabela 3.1: Wartości różnicujące dla puli kursów.

Powyższe przekształcenia danych wystarczają do wykonania analizy wsparcia (def. 1.16).

W myśl Definicji 1.15 analiza samowsparcia wymaga przygotowania dodatkowych zbiorów danych zawierających informacje o osiągnięciu sukcesu w kolejnym cyklu dydaktycznym. W stosunku do poprzednich modyfikacji danych, należy zmienić sposób odfiltrowywania danych. Określenie wartości sukcesu pozytywnej ocen i sukcesu oceny ponad*przeciętnej odbywa się w ten sam sposób, przy czym wartości różnicujące nie są wyznaczane ponownie. Proces przygotowania danych został przedstawiony na Rysunku 3.1.



Rysunek 3.1: Schemat przetworzenia danych dotyczących kursu bazowego: oryginalne dane → dane przefiltrowane → dane końcowe.

3.2. Analiza wsparcia

Zgodnie z Definicją 1.16, analiza wsparcia polega na zidentyfikowaniu zbiorów wspierających (def. 1.14) dla każdego kursu należącego do puli kursów. Algorytm, który to tego prowadzi jest algorytmem iteracyjnym względem kursów bazowych rozważanych par testowych. W celu opisanie ogólnej koncepcji analizy wsparcia, ustala się chwilowo kurs $K \in \mathcal{K}$. Będzie on pełnił rolę kursu bazowego.

3.2.1. Koncepcja

W oczywisty sposób wszystkich studentów próbujących swoich sił w kursie K można podzielić na grupy związane z osiągnięciem sukcesu s . Z drugiej strony, uwzględniając inny kurs $K_t \in \mathcal{K}$, studentów uczestniczących w kursie K można podzielić na tych, którzy przed przystąpieniem do egzaminu z kursu K mieli styczność z kursem K_t . Innymi słowy, każdemu studentowi kursu K można przypisać etykietę $Kursant$, gdzie:

$$Kursant = \begin{cases} 1, & \text{gdy w momencie przystąpienia do egzaminu student wysłuchał kursu } K_t, \\ 0, & \text{w p.p.} \end{cases}$$

W nawiązaniu do danych, wartość etykiety $Kursant$ jest ustalana przez porównanie atrybutu $cdyd_kod$ dla kursów K i K_t . Powyższe podziały oraz poszczególne licznosci można zestawzić w tablicy dwudzielnej. Jej schemat dla ustalonej pary testowej jest przedstawiony w Tabeli 3.2:

Kurs testowy \ Kurs bazowy	$sukces = 1$	$sukces = 0$
	a	b
$Kursant = 1$	a	b
$Kursant = 0$	c	d

Tabela 3.2: Schemat tablicy dwudzielnej dla analizy wsparcia.

Dodatkowo tablica dwudzielna niesie informację o liczbie unikalnych studentów, którzy kiedykolwiek zapisali się na kurs bazowy. W Rozdziale 4.1 tablice dwudzielne są poprzedzane zmodyfikowanymi diagramami Venna. Dzięki nim można było uzyskać ogólne informacje o wzajemnym zainteresowaniu analizowaną parą testową.

Na podstawie tablic dwudzielnych można określić następujące współczynniki:

Definicja 3.1. (pk_1, pk_0) Dla pary testowej (K_b, K_t) określa się współczynniki pk_1, pk_0 dane wzorami:

$$pk_1(K_b, K_t) = \frac{a}{a+b}, \quad (3.3)$$

$$pk_0(K_b, K_t) = \frac{c}{c+d}. \quad (3.4)$$

Współczynniki pk_1 i pk_0 interpretuje się w terminach prawdopodobieństw warunkowych związanych z odpowiadającą im parą testową (K_b, K_t) .

Rozważymy dyskretną przestrzeń probabilistyczną (Ω, \mathcal{F}, P) związaną z kursem K_b :

$$\begin{aligned} \Omega &= \{\omega_1, \omega_2, \dots, \omega_n\}, \quad \text{gdzie } n = \#K_b = a + b + c + d, \\ \mathcal{F} &= 2^\Omega, \\ P &= \frac{1}{n}. \end{aligned}$$

Zdarzenia A, B_1, B_0 zdefiniowane następująco:

$$\begin{aligned} A &= \{\omega_i : \omega_i \text{ ukończył kurs } K_b \text{ z sukcesem}\}, \\ B_1 &= \{\omega_i : \omega_i \text{ był Kursantem } K_t\}, \\ B_0 &= \{\omega_i : \omega_i \text{ nie był Kursantem } K_t\}. \end{aligned}$$

Wówczas pk_1 i pk_0 są prawdopodobieństwami warunkowymi dla zdarzenia A względem zdarzeń B_1 i B_0 odpowiednio, tzn:

$$pk_0 = P(A|B_0), \quad pk_1 = P(A|B_1). \quad (3.5)$$

Zależności pomiędzy tymi współczynnikami stanowią punkt odniesienia do wyselekcjonowania kursów wspierających. W tym celu wprowadza się następujące pojęcia:

Definicja 3.2. (Współczynnik wsparcia) Współczynnikiem wsparcia dla pary testowej (K_b, K_t) nazywamy wielkość:

$$wsp(K_b, K_t) = pk_1(K_b, K_t) - pk_0(K_b, K_t). \quad (3.6)$$

Definicja 3.3. (Współczynnik proporcji) Współczynnikiem proporcji dla pary testowej (K_b, K_t) nazywamy wielkość:

$$prop(K_b, K_t) = \frac{pk_1(K_b, K_t)}{pk_0(K_b, K_t)}. \quad (3.7)$$

Zbiory wspierające dla danego kursu bazowego (def. 1.14) określa się wówczas jako:

$$W_{K_b} = \{K_t \in \mathcal{K} : wsp(K_b, K_t) \geq 0\}. \quad (3.8)$$

3.2.2. Własności współczynnika wsparcia i zbiorów wspierających

Wprowadzony powyżej współczynnik wsparcia stanowi ostatecznie punkt wyjścia do określenia zbiorów wspierających, a tym samym do przeprowadzenia analizy wsparcia. Poniżej zestawiono jego główne własności:

1. Wartości współczynnika wsparcia:

Współczynnik wsparcia może przyjmować wartości z odcinka $[-1; 1]$, o ile odpowiednie wartości pk_0 i pk_1 przyjmują wartości liczbowe - sytuacja przeciwna ma miejsce, gdy jeden ze zbiorów B_0 , B_1 jest zbiorem pustym. Mówimy wówczas, że współczynnik wsparcia dla danej pary testowej pozostaje nieokreślony i przyjmuje wartość NA, tak samo jak współczynnik pk_0/pk_1 .

2. Problem samowsparcia w terminach analizy wsparcia:

Samowsparcie (def. 1.17) w terminach analizy wsparcia odpowiada na pytanie, czy $K \in W_K$? Okazuje się, że dla ustalonego kursu K , wartość współczynnika $wsp(K, K)$ pozostaje nieokreślona. W tym przypadku nie ma bowiem żadnego studenta, któremu można przypisać etykietę *Kursant* równą 0. Tablica dwudzielna dla tego przypadku została przedstawiona w Tabeli 3.3. Zgodnie z Definicją 3.1 współczynnik pk_0 pozostanie nieokreślony, co skutkuje nieokreślonością głównego współczynnika wsparcia.

Taki rezultat jest wynikiem sposobu przygotowania danych (roz. 3.1), a dokładniej - filtracji. Aby nie tracić informacji o klasie samowspierającej (def. 1.15) opracowano drugą część algorytmu (samo)wsparcia, który skupia się jedynie na tej klasie. Szczegóły stanowią treść Rozdziału 3.3.

Kurs bazowy \ Kurs bazowy	$sukces = 1$	$sukces = 0$
$Kursant = 1$	a	b
$Kursant = 0$	0	0

Tabela 3.3: Schemat tablicy dwudzielnej dla problemu samowsparcia.

3. Wrażliwość na wartości a, b, c, d tablic dwudzielnych:

Rozważmy sytuację, kiedy rozbieżność kursu K_b na zbiory B_0 i B_1 wygląda następująco:

$$B_1 = \{\omega_1\},$$

$$B_0 = \{\omega_2, \dots, \omega_n\},$$

gdzie n oznacza liczbę kursu K . Odpowiada to sytuacji, kiedy tylko 1 student, Iksiński, zdecydował się realizować parę kursów w odwrotnej kolejności niż wszyscy jego koledzy (tab. 3.4).

a) sukces Iksińskiego				b) porażka Iksińskiego			
Kurs bazowy \ Kurs bazowy	$s = 1$	$s = 0$		Kurs bazowy \ Kurs bazowy	$s = 1$	$s = 0$	
$Kursant = 1$	1	0		$Kursant = 1$	0	1	
$Kursant = 0$	c	d		$Kursant = 0$	c	d	

Tabela 3.4: Schemat tablicy dwudzielnej dla rozbieżności $B_1 = \{\omega_1\}$, $B_0 = \{\omega_2, \dots, \omega_n\}$.

W tej sytuacji otrzymuje się następującą wartość współczynnika wsparcia:

$$wsp(K_b, K_t) = \begin{cases} 1 - \frac{c}{c+d} \geq 0, & \text{gdy Iksiński osiągnie sukces,} \\ -\frac{c}{c+d} \leq 0, & \text{w p.p.} \end{cases} \quad (3.9)$$

Nierówności słabe są osiągnięte tylko w sytuacji, gdy $c = 0 \vee d = 0$. Zatem, jeśli tylko wśród pozostałych studentów będzie co najmniej jeden, który z punktu widzenia sukcesu osiągnie wynik inny od pozostałych, decyzja o sklasyfikowaniu kursu testowego jako wspierającego będzie zdeterminowana przez wynik Iksińskiego.

Analogiczne rozumowanie można przeprowadzić dla rozbieżności $B_0 = \{\omega_1\}$, $B_1 = \{\omega_2, \dots, \omega_n\}$, kiedy wartości (c, d) są zero-jedynkowe.

Przykład Iksińskiego pokazuje, że współczynnik wsparcia jest wrażliwy na małe wartości a, b, c, d . Dlatego kursy wspierające powinny zależeć od dodatkowego parametru, który określi możliwe minimalne wartości występujące w tabeli dwudzielnej, dla których współczynnik wsparcia jest określany. Dodatkowo można wprowadzić parametr, dzięki któremu zbiory wspierające będą gromadzić tylko te kursy, których wspieranie (mierzone wartością współczynnika wsparcia) będzie odpowiednio silne.

Powyższe uwagi prowadzą do następującej definicji:

Definicja 3.4. (Zbiór (α, M) -wspierający) Przez zbiór (α, M) -wspierający $W_{K_b}^{(\alpha, M)}$ rozumiemy podzbiór puli przedmiotów o następującej własności:

$$W_{K_b}^{(\alpha, M)} = \{K_t \in \mathcal{K} : wsp(K_b, K_t) \geq \alpha \wedge a, b, c, d \geq M\}, \quad (3.10)$$

gdzie: $\alpha \in [0; 1]$, $M \in \mathbb{N}$, a, b, c, d - wartości tablicy dwudzielnej.

3.2.3. Dobór parametrów

Mając w pamięci powyższe definicje i uwagi, należy ustalić ostateczne wartości parametrów (α, M) . W tym celu przeprowadzono analizę wsparcia dla następujących wartości parametrów w obrębie każdej pary testowej:

- $\alpha \in \{0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$,
- $M \in \{0, 1, 3, 5, 7, 10\}$.

Analiza została przeprowadzona oddzielnie dla sukcesu pozytywnej oceny oraz sukcesu oceny ponad*przeciętnej. Dodatkowo uwzględniono podział studentów na studentów matematyki bądź informatyki. Sposób ich etykietowania został opisany w Rozdziale 3.4. Z kolei w Rozdziale 4.1 można znaleźć częściową analizę wsparcia na przykładzie *Procesów Stochastycznych I*.

Ostatecznie wybrano następujące wartości parametrów:

1. Sukces $s_{poz.ocena}$: Para parametrów (α, M) ustalona na poziomie $(0.15, 3)$,
2. Sukces $s_{p*przecietna}$: Para parametrów (α, M) ustalona na poziomie $(0.25, 3)$.

Parametry dla studentów matematyki i studentów informatyki w obrębie rozważanego aktualnie sukcesu zostały ustalone na tych samych poziomach, by ułatwić ich późniejsze porównanie.

3.3. Analiza samowsparcia

Jak pokazano w Rozdziale 3.2.2, klasy samowspierającej \mathcal{S} (def. 1.15) nie można wyznaczyć w ramach analizy wsparcia. Można natomiast dokonać korekty tego mechanizmu niwelując problem zerowania się jednego z wierszy tablic dwudzielnych, który uniemożliwiał wykorzystanie algorytmu dla pary testowej (K, K) .

3.3.1. Koncepcja

Próbując wyznaczyć klasę \mathcal{S} nie ma potrzeby rozważania następstwa czasowego, gdyż, jak pokazano w Rozdziale 3.2.2 jest ono jednoznaczne. Istotą jest porównanie dwóch prawdopodobieństw odnoszących się do rozważanego kursu K - jednym związanym osiągnięciem sukcesu s podczas pierwszej realizacji kursu, drugim - związanym z osiągnięciem sukcesu s przez studentów, którzy musieli realizować kurs ponownie. Można więc zdefiniować nowe tabele dwudzielne, które będą adekwatne do problemu samowsparcia. Ich schemat został przedstawiony w Tabeli 3.5.

Kurs K		
	$sukces = 1$	$sukces = 0$
cykl.dyd.2	a	b
cykl.dyd.1	c	d

Tabela 3.5: Tablica dwudzielna wykorzystywana w analizie samowsparcia.

Zdefiniowane na podstawie tablic dwudzielnych dla analizy wsparcia współczynniki pk_1 , pk_0 , (def. 3.1) mają swoje zastosowanie w analizie samowsparcia, lecz zmienia się ich interpretacja. Nie są one teraz związane z prawdopodobieństwami warunkowymi, lecz z prawdopodobieństwami osiągnięcia sukcesu s podczas drugiego i pierwszego cyklu dydaktycznego

odpowiednio. Współczynniki $wsp(K, K)$ i $prop(K, K)$ (def. 3.2, 3.3) stanowią z kolei odniesienia w wyznaczaniu klasy \mathcal{S} , którą można zapisać jako:

$$\mathcal{S} = \{K \in \mathcal{K} : wsp(K, K) \geq 0\}. \quad (3.11)$$

W celu odróżnienia, że współczynniki wsparcia i proporcji odnoszą się do analizy samowsparcia stosuje się skróconą notację $wsp(K)$ i $prop(K)$.

3.3.2. Własności współczynnika wsparcia i klasy samowspierającej

Ogólne własności współczynnika wsparcia $wsp(K)$ są identyczne co własności współczynnika wsparcia $wsp(K_t, K_b)$. Może on przyjmować wartości z przedziału $[-1; 1]$ bądź pozostać nieokreślony, co umownie zapisuje się jako wartość NA. Nie zmienia się również jego wrażliwość na wartości a, b, c, d pochodzące z tabeli dwudzielnej, na podstawie której został wyznaczony. Szczegóły można znaleźć w Rozdziale 3.2.2.

Powyższe uwagi prowadzą do wprowadzenia ograniczeń w wyznaczaniu klasy \mathcal{S} :

Definicja 3.5. (Klasa (α, M) -samowspierająca) Przez klasę (α, M) -samowspierającą \mathcal{S}_M^α rozumiemy podzbiór puli przedmiotów o następującej własności:

$$\mathcal{S}_M^\alpha = \{K_t \in \mathcal{K} : wsp(K) \geq \alpha \wedge a, b, c, d \geq M\}, \quad (3.12)$$

gdzie: $\alpha \in [0; 1]$, $M \in \mathbb{N}$, a, b, c, d - wartości tablicy dwudzielnej dla problemu samowsparcia.

3.3.3. Dobór parametrów

Samowsparcie również jest sparametryzowane przez s , M i α . Przed prezentacją ostatecznych wyników należało wybrać odpowiedni zestaw parametrów dla rozważanych wartości s , który najlepiej oddaje badany problem. Zbiór początkowo rozważanych wartości parametrów jest identyczny jak w przypadku analizy wsparcia:

- $\alpha \in \{0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$,
- $M \in \{0, 1, 3, 5, 7, 10\}$,

zarówno dla $s_{poz.ocena}$ jak i $s_{p*przecietna}$. Podział studentów na studentów matematyki i informatyki (roz. 3.4) w dalszym ciągu uwzględniano.

Pierwszym krokiem w ustaleniu parametrów była analiza liczności odpowiednich klas samowspierających. Zostały one przedstawione w Tabeli 3.6. Można zauważyć, że odcinając się w tym przypadku od wartości 0 w tabelach dwudzielnych, klasa \mathcal{S} będzie zbiorem pustym dla studentów informatyki i zbiorem co najwyżej jednoelementowym dla studentów matematyki. Należy się jednak spodziewać, że w przypadku samowsparcia będzie zachodził związek:

$$a + b << c + d. \quad (3.13)$$

Uzasadnieniem dla związku 3.13 jest fakt, że wartość $a + b$ jest prawie zawsze mniejsza od wartości d , gdyż część studentów mogła uzyskać zaliczenie podczas egzaminu poprawkowego, jeszcze podczas pierwszego cyklu dydaktycznego. Wówczas ci studenci nie będą już wliczani w wartość $a + b$. Wrażliwość na wartości a, b, c, d , jakkolwiek dalej obecna, nie odgrywa już tak istotnej roli.

Ostatecznie wybrano następujące wartości parametrów:

	M-0	M-1	M-3	M-5	M-7	M-10
α 0	8-4	1-0	1-0	1-0	1-0	1-0
α 0.1	5-4	0-0	0-0	0-0	0-0	0-0
α 0.15	2-3	0-0	0-0	0-0	0-0	0-0
α 0.2	2-2	0-0	0-0	0-0	0-0	0-0
α 0.25	1-1	0-0	0-0	0-0	0-0	0-0
α 0.3	0-1	0-0	0-0	0-0	0-0	0-0
α 0.35	0-1	0-0	0-0	0-0	0-0	0-0
α 0.4	0-0	0-0	0-0	0-0	0-0	0-0
α 0.45	0-0	0-0	0-0	0-0	0-0	0-0
α 0.5	0-0	0-0	0-0	0-0	0-0	0-0

Tabela 3.6: Liczności klas samowspierających \mathcal{S}_M^α - $s_{poz.ocena}$ dla studentów matematyki i informatyki.

1. Sukces $s_{poz.ocena}$: Para parametrów (α, M) ustalona na poziomie $(0.10, 0)$,
2. Sukces $s_{p*przecietna}$: Para parametrów (α, M) ustalona na poziomie $(0.10, 0)$.

Parametry dla studentów matematyki i studentów informatyki w obrębie rozważanego aktualnie sukcesu zostały ustalone na tych samych poziomach.

3.4. Kierunek studiów

Wydział MIMUW w swojej ofercie posiada pięć kierunków studiów I stopnia, tj. *bioinformatykę*, *informatykę*, *JSEM*, *JSIM* i *matematykę* oraz trzy kierunki studiów II stopnia, tj. *bioinformatykę*, *informatykę* i *matematykę*. Z uwagi na to analiza wsparcia powinna uwzględniać zróżnicowanie studentów o te kierunki. Jednak informacja o tym nie była zawarta bezpośrednio w danych, które zostały udostępnione do analizy. Dlatego istotnym krokiem była próba przypisania studentów do specjalizacji w oparciu o posiadaną wiedzę dotyczącą kursów i studentów. Analizując siatki zajęć dla poszczególnych kierunków oraz sposób kodowania przedmiotów, można było wyróżnić dwie główne metody określenia specjalizacji dla studentów, pozwalające podzielić ich jako *studentów informatyki* lub *studentów matematyki*.

3.4.1. Metoda 1 - badanie przedmiotu rozróżniającego

Metoda 1 opiera się na tym, iż podczas I semestru studiów I stopnia studenci mają z góry określony plan zajęć i kursy do zrealizowania. Przedmiotami, które istotnie wyznaczają w/w podział wśród studentów są:

- *Wstęp do informatyki* - przedmiot obowiązkowy dla kierunku Matematyka, JSEM, Bioinformatyka,
- *Wstęp do programowania* - przedmiot obowiązkowy dla kierunku Informatyka, JSIM.

Można sprawdzić, na który spośród tych dwu przedmiotów uczęszczała dana osoba podczas swojego pierwszego semestru studiów i na tej podstawie określić ją jako studenta matematyki bądź studenta informatyki odpowiednio.

Użycie powyższej metody dla studentów realizujących co najmniej jeden kurs z puli wybranych kursów daje następujące rezultaty:

```
> table(factor(infoStudenci$k1, levels=c("Inf","Mat","nk"),
+ labels = c("studenci informatyki","studenci matematyki","niesklasyfikowani")))
```

```
studenci informatyki  studenci matematyki  niesklasyfikowani
                613                814                676
```

Mimo, iż metoda ta jest oparta na programie studiów, nie pozwala ona sklasyfikować aż 32% studentów. Powodem, dla którego odsetek niesklasyfikowanych studentów jest tak duży może być fakt, iż nie wszyscy studenci uwzględnieni w danych USOS musieli rozpoczynać studia właśnie na Wydziale MIMUW, co spowodowało, że nie uczęszczali na kluczowe dla tej metody przedmioty. Poza tym, wymagana jest tutaj znajomość daty rozpoczęcia studiów przez danego studenta - informacja ta również nie występuje *explicite*. Została ona uzyskana z danych jako minimalna wartość atrybutu *cdyd_kod* dla danego studenta, już na podstawie wszystkich 4954 plików danych. Poniżej podsumowanie tak wyliczonego roku rozpoczęcia studiów:

```
> table(factor(infoStudenci$data_rozp2))
```

```
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
      3  650  177  187  163  171  159  148  184  174   87
```

Z powyższego podsumowania widać, że w oparciu o dane, rok 2000 stanowi początkową datę studiów dla blisko czterokrotnie większej liczby studentów niż w latach 2001-2008. Jest to spowodowane tym, że właśnie w roku 2000 rozpoczęło się wdrażanie systemu USOS na Wydziale MIMUW (roz. 1.3.1). Pojawiające się trzy obserwacje z roku 1999 są najprawdopodobniej wynikiem przyswajania systemu - kursy, na które chodzili studenci w 1999 roku to m.in. *Analiza matematyczna*, *Algebra liniowa*, czy *Wprowadzenie do ekonomii*¹. Z perspektywy kształcenia na Wydziale MIMUW są one z dużo bardziej oblegane.

Ta obserwacja tłumaczy wysoką zawodność przedstawionej metody wyznaczenia kierunku studiów. Należy zaznaczyć, że wraz z upływem czasu odsetek sklasyfikowanych studentów będzie wzrastał. Nie mniej, studenci rozpoczynający swoją edukację akademicką poza Wydziałem MIMUW w dalszym ciągu nie będą mogli zostać sklasyfikowani.

3.4.2. Metoda 2 - analiza kodów kursów

Metoda 2 bazuje na całej historii dobierania kursów przez studentów. Analizując kody kursów oferowanych w ramach poszczególnych grup przedmiotów oferowanych przez Wydział MIMUW można część z nich można pogrupować na kursy matematyczne i informatyczne. Szczegóły znajdują się w Tabeli 3.7. Strategia doboru kierunku studiów może opierać się wówczas na ilości kursów matematycznych i informatycznych - jeśli spośród tych dwu rodzajów przeważają kursy matematyczne klasyfikuje się studenta jako studenta matematyki, jeśli informatyczne - studenta informatyki. W przypadku równej ilości obydwu rodzajów kursu student pozostaje niesklasyfikowany. Warto odnotować, że w przypadku powtarzania przez studenta danego kursu, przedmiot ten będzie wliczany tylko raz.

Powyższa metoda prowadzi do następujących wyników:

```
> table(factor(infoStudenci$k3, levels=c("Inf","Mat","nk"),
+ labels = c("studenci informatyki","studenci matematyki", "nk")))
```

¹Sumarycznie pojawiło się 9 kursów, w których występują obserwacje z 1999 roku.

kod kursu	rodzaj kursu
1000-1...	kursy matematyczne
1000-5...	
1000-7...	
1000-2...	kursy informatyczne
1000-4...	
1000-6...	

Tabela 3.7: Zestawienie kodu kursów oferowanych przez Wydział MIMUW.

```

studenci informatyki  studenci matematyki  niesklasyfikowani
                        805                      1287                      11
>

```

3.4.3. Kierunek studiów - Porównanie metod

Pierwsza z zaproponowanych metod wykazuje zawodność na poziomie 32%, druga - na poziomie 0.5%. Dodatkową zaletą drugiej z metod jest jej niezależność od rodzaju studiów - zarówno studenci studiów jednolitych, pierwszego czy drugiego stopnia mogą zostać nią sklasyfikowani. Z kolei wyniki uzyskane pierwszą metodą można stosować tylko dla studentów studiów jednolitych i pierwszego stopnia. Przed ostateczną decyzją wyboru metody klasyfikacyjnej uzyskane wyniki zostały porównane.

```

> table("Metoda 1" = factor(infoStudenci$k1, levels=c("Inf","Mat","nk"),
+   labels = c("Informatycy","Matematycy","nk")),
+   "Metoda 2" = factor(infoStudenci$k3, levels=c("Inf","Mat","nk"),
+   labels = c("Informatycy","Matematycy","nk")))

```

	Metoda 2		
Metoda 1	Informatycy	Matematycy	nk
Informatycy	568	39	6
Matematycy	50	762	2
nk	187	486	3

Powyższy fragment kodu programu R przedstawia macierz kontyngencji uzyskanych wyników. Okazuje się, że na 1427 studentów, którzy zostali sklasyfikowani w oparciu o przedmiot rozróżniający, 93% otrzymało te same etykiety przy wykorzystaniu metody związanej z analizą kodów kursów. To pokazuje, że istotnie uczęszczanie na dany kurs rozróżniający skutkuje w przyszłości wyborem przedmiotów z nim związanych. Dlatego ostatecznie wykorzystano wyniki uzyskane za pomocą drugiej metody. Informacje o 11 studentach, którzy w tym przypadku nie zostali sklasyfikowani zostały pominięte.

Rozdział 4

Wyniki

Pierwsza część niniejszego rozdziału (roz. 4.1) opisuje fragment przeprowadzonej analizy szczegółowej, która prowadziła do wyboru odpowiednich parametrów dla problemu wsparcia i samowsparcia. Z uwagi na dużą liczbę rozważanych przypadków, zaprezentowane rozważania odnoszą się do arbitralnie ustalonych par testowych (def. 1.13), w których rolę kursu bazowego odgrywają zawsze *Procesy Stochastyczne I*. Odniesiono się zarówno do problemu wsparcia z sukcesem pozytywnej oceny, oceny ponad*przeciętnej, jak i do problemu samowsparcia.

W drugiej części (roz. 4.2) przedstawiono uzyskane sieci (samo)wspierania. Sposób konstruowania ich graficznej reprezentacji stanowi treść Rozdziału 4.2.1. Ostateczne postaci sieci, które wykonano w programie **Cytoscape**, zostały przedstawione w Rozdziale 4.2.2. Sam program jest ogólnodostępny pod adresem [16]. Można znaleźć wiele publikacji ułatwiających jego efektywne wykorzystanie. Podczas tworzenia przedstawionych grafów szczególnie pomocna okazała się pozycja [12].

4.1. Analiza szczegółowa - przykład

Pierwszym etapem wykonania analizy wsparcia i samowsparcia był dobór odpowiednich parametrów M i α , regulujących licznosc poszczególnych grup studentów oraz siłę wiązania pomiędzy analizowaną parą testową. W Rozdziałach 3.2.3 i 3.3.3 można znaleźć dokładne informacje na ten temat. Duża liczba przypadków składająca się na analizę szczegółową uniemożliwiła jej przedstawienie w całości. W celu zilustrowania algorytmu zamieszczono poniżej jej przykład w oparciu o kurs bazowy *Procesy stochastyczne I* (pr.stoch1) oraz o następujące przedmioty testowe:

- Jakościowa teoria równań różniczkowych zwyczajnych - jtrrz,
- Rachunek prawdopodobieństwa II - r.pr2a,
- Rachunek prawdopodobieństwa II* - r.pr2b,
- Procesy Stochastyczne II - pr.stoch2.

Przypomina się, że zgodnie z Tabelą 3.1 wartością różnicującą (roz. 3.1) dla *Procesów stochastycznych I* jest 3.

Całość analizy szczegółowej stanowi treść Raportu *Szczegóły wyznaczania sieci (samo)wspierania* znajdującego się na płycie CD dołączonej do niniejszej pracy magisterskiej. Jego wybrane elementy można znaleźć w załączniku B.

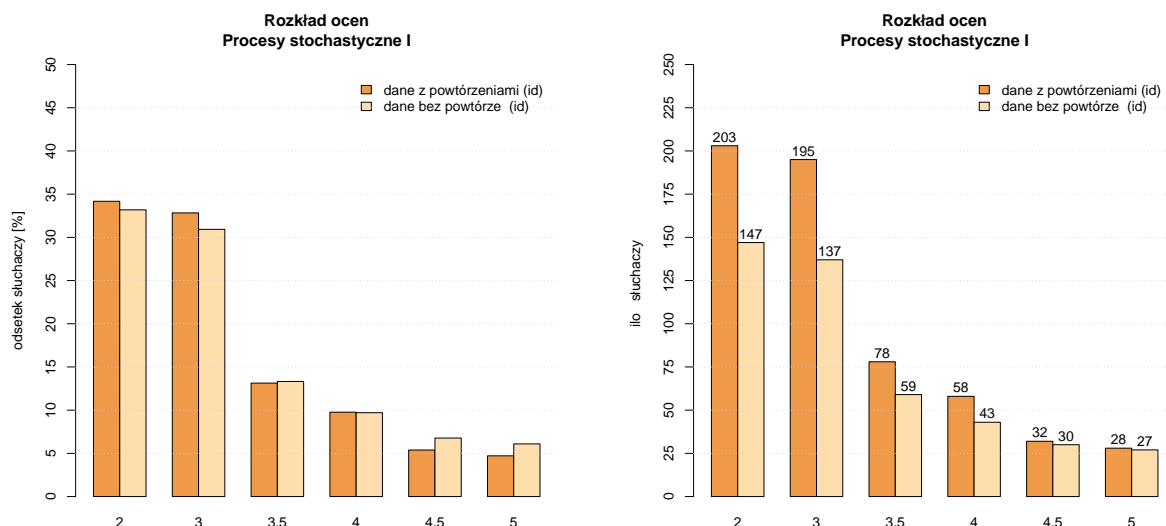
Poznanie kursu bazowego

Pierwszym etapem analizy szczegółowej jest ogólne zapoznanie się z przedmiotem bazowym.

Na Rysunku 4.1 przedstawiono rozkład uzyskanych ocen dla badanego przedmiotu bazowego - *Procesów Stochastycznych I* - w ujęciu ilościowym oraz unormowany. Wprowadzone pogrupowanie nawiązuje do wstępnego przetwarzania danych opisanego w Rozdziale 3.1, gdzie ciemniejsze słupki to dane przed odfiltrowaniem, a jaśniejsze - po odfiltrowaniu.

Widać, że co trzecia osoba nie uzyskuje zaliczenia z badanego kursu przystępując pierwszy raz do egzaminu. Co więcej, dla blisko 50 osób drugi termin również nie wystarcza do uzyskania progu zdawalności. Jeśli już studenci uzyskują zaliczenie podczas pierwszego terminu egzaminu, wówczas w dla 46% z nich jest to najniższy z możliwych wyników. Wykonanie analizy wsparcia ma na celu wskazanie kursów, których wcześniejsza realizacja ma szansę pozytywnie wpłynąć na powyższe statystyki.

Należy zaznaczyć, że ogólne zapoznanie się z kursem nie uwzględnia zróżnicowania na studentów danego typu. Informacje o tym pojawiają się dopiero podczas analizy ustalonych par testowych, która stanowi kolejny etap analizy szczegółowej.



Rysunek 4.1: Rozkład ocen dla Procesów stochastycznych I: a) unormowany b) w ujęciu ilościowym.

Badanie par testowych

1. Para testowa (*pr.stoch1, jtrrz*):

Rysunek 4.2 przedstawia wzajemne zainteresowanie badaną aktualnie parą testową. Mimo, iż blisko połowa studentów uczęszczających na *Jakościową Teorię Równań Różniczkowych Zwyczajnych* zrealizowała podczas swoich studiów również badane aktualnie *Procesy Stochastyczne I*, to stanowią oni tylko 9% nich. Uwzględniając następstwo czasowe oraz przynależność do odpowiedniej specjalizacji (tab. 4.1) można zauważyć dodatkowo, że zdecydowana większość słuchaczy kursu bazowego to studenci matematyki. Widać również, że sumarycznie kurs testowy był realizowany przez studentów obydwu specjalizacji równomiernie w stosunku do kursu bazowego - 46% realizowało go nie później niż kurs bazowy. Jednak wśród studentów informatyki tylko 1 student zdecydował się na taką kolejność. W Tabeli 4.2 przedstawiono wartości współczynników analizy wsparcia, zdefiniowane w Rozdziale 3.2.1.

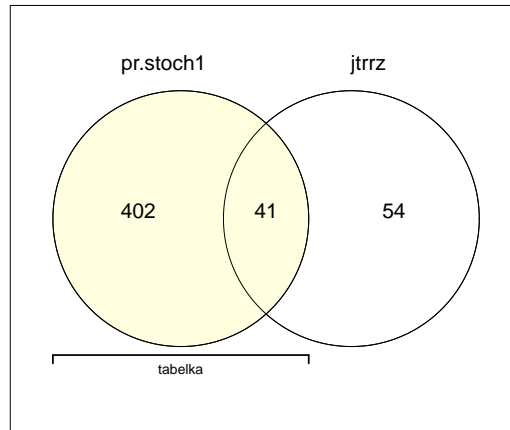
Konsekwencją tego, iż wspomniany student informatyki osiągnął sukces w myśl sukcesu $s_{poz.ocena}$ jest pozytywna wartość współczynnika wsp . Przez wysoki odsetek zdawalności kursu bazowego wśród studentów informatyki, którzy nie mieli jeszcze do czynienia z kursem testowym (77%) uplasował się on na poziomie 23%. Jednak z uwagi na pojawiającą się wartość 0 w tablicy dwudzielnej (tab. 4.2b)) *Jakościowa Teoria Równań Różniczkowych Zwyczajnych* będzie wchodzić w skład zbiorów $W_{p.stoch1}^{(\alpha, M)}$ tylko dla ograniczenia $M = 0$ (gdy $\alpha \in \{0, 0.10.15, 0.20\}$).

W przypadku tej pary testowej dla studentów matematyki współczynnik wsp został ustalony na zbliżonym poziomie, tj. 22%. Jednak tutaj *Jakościowa Teoria Równań Różniczkowych Zwyczajnych* będzie zasilać zbiór $W_{p.stoch1}^{(\alpha, M)}$ dodatkowo dla parametru $M \in \{1, 2, 3\}$, ze względu na wyższe wartości tablicy dwudzielnej.

Rozważmy teraz tablice dwudzielne dla sukcesu związanego z uzyskaniem oceny ponadprzeciętnej, tzn. oceny ≥ 3.5 (tab. 4.3). W dalszym ciągu większość studentów matematyki, która zrealizowała nie później kurs testowy uzyskuje sukces, plasując się tym samym w górnej części uzyskanych wyników. Nie można tego powiedzieć o pozostałych studentach - skutkuje to tym, iż proporcje w tabeli dwudzielnej wśród studentów o etykiecie $Kursant = 0$ odwróciły się, w porównaniu do rozważanego wcześniej sukcesu. Ostatecznie współczynnik wsparcia wynosi aż 31%.

W kontekście studentów informatyki współczynnik wsparcia zwiększył się aż do 54%, z uwagi na ponadprzeciętną ocenę studenta, który zrealizował *Jakościową Teorię Równań Różniczkowych Zwyczajnych* nie później niż *Procesy Stochastyczne I* oraz ponowne zaburzenie proporcji wśród pozostałych studentów. Widać również, że współczynnik proporcji pierwszy raz przekroczył wartość 2. Jednak pomimo dużej wartości współczynnika wsparcia, w większości przypadków kurs testowy nie będzie stanowił jednego z kursów (α, M) -wspierających.

Słuchacze



Rysunek 4.2: Diagram Venna dla pary $(pr.stoch1, jtrrz)$.

studenci matematyki			studenci informatyki		
jtrrz \ pr.stoch1	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$	jtrrz \ pr.stoch1	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$
	Kursant - 1	19	3	1	0
Kursant - 0	246	135	Kursant - 0	30	9

Tabela 4.1: Tablice dwudzielne dla problemu wsparcia i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, jtrrz)$.

	studenci matematyki	studenci informatyki
pk_1	0.86	1
pk_0	0.65	0.77
wsp	0.22	0.23
$prop$	1.37	1.30

Tabela 4.2: Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, jtrrz)$.

studenci matematyki			studenci informatyki		
jtrrz \ pr.stoch1	$s_{p*przt} = 1$	$s_{p*przt} = 0$	jtrrz \ pr.stoch1	$s_{p*przt} = 1$	$s_{p*przt} = 0$
	Kursant - 1	14	8	1	0
Kursant - 0	126	255	Kursant - 0	18	21

Tabela 4.3: Tablice dwudzielne dla analizy wsparcia i sukcesu s_{p*przt} dla pary $(pr.stoch1, jtrrz)$.

	studenci matematyki	studenci informatyki
pk_1	0.64	1
pk_0	0.33	0.46
wsp	0.31	0.54
$prop$	1.92	2.17

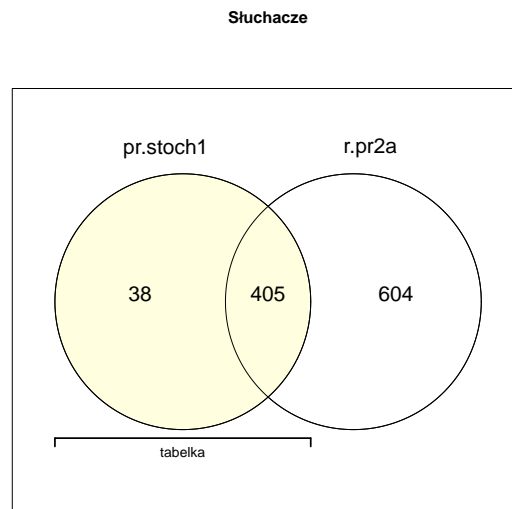
Tabela 4.4: Wartości współczynników analizy wsparcia i sukcesu $s_{p.*przecietna}$ dla pary $(pr.stoch1, jtrrz)$.

2. Para testowa (*pr.stoch1, r.pr2a*):

Wzajemnie zainteresowanie badaną parą kursów przedstawia diagram Venna (rys. 4.3). W tym przypadku jest 405 studentów, którzy realizowali obydwie kursy współtworzące parę testową, co stanowi 91% wszystkich studentów *Procesów Stochastycznych I*. Informacje o zróżnicowaniu na specjalizacje oraz o związku czasowym zostały pośrednio przedstawione w tablicach dwudzielnych z Tabeli 4.5. W odróżnieniu od poprzedniej pary testowej mamy teraz do czynienia z sytuacją, w której kurs testowy był w większości przypadków realizowany nie później niż kurs bazowy (99%). Przewagę tych studentów stanowią studenci matematyki. W Tabeli 4.6 zostały przedstawione wartości współczynników wsparcia dla rozważanych kursów.

Bez względu na rodzaj specjalizacji studenci *Procesów Stochastycznych I*, którzy w momencie przystąpienia do egzaminu posiadali już w wiedzę uzyskaną podczas realizacji *Rachunku Prawdopodobieństwa II*, rzadziej osiągal sukces *s_{poz.ocena}*, co skutkowało ujemnymi wartościami współczynnika *wsp*. Skutkuje to tym, iż dla żadnej rozważanej pary parametrów (α, M) kurs bazowy nie będzie stanowił kursu wspierającego.

W tym przypadku sukces związany z osiągnięciem oceny ponad*przeciętnej prowadzi do podobnych wniosków. Zarówno studenci matematyki, jak i informatyki częściej uzyskiwali lepsze rezultaty w sytuacji, gdy kurs bazowy realizowali po kursie testowym (tab. 4.5). To powoduje ustalenie ujemnych wartości współczynników wsparcia (tab. 4.6).



Rysunek 4.3: Diagram Venna dla pary (*pr.stoch1, r.pr2a*).

studenci matematyki			studenci informatyki		
<div>pr.stoch1</div> <div>r.pr2a</div>	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$	<div>pr.stoch1</div> <div>r.pr2a</div>	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$
Kursant - 1	240	130	Kursant - 1	26	8
Kursant - 0	25	8	Kursant - 0	5	1

Tabela 4.5: Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, r.pr2a)$.

	studenci matematyki	studenci informatyki
pk_1	0.65	0.76
pk_0	0.76	0.83
wsp	-0.11	-0.07
$prop$	0.86	0.92

Tabela 4.6: Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, r.pr2a)$.

studenci matematyki		
<div>pr.stoch1</div> <div>r.pr2a</div>	$s_{p*przt} = 1$	$s_{p*przt} = 0$
Kursant - 1	124	246
Kursant - 0	16	17

studenci informatyki		
<div>pr.stoch1</div> <div>r.pr2a</div>	$s_{p*przt} = 1$	$s_{p*przt} = 0$
Kursant - 1	15	19
Kursant - 0	4	2

Tabela 4.7: Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{p*przecietna}$ dla pary $(pr.stoch1, r.pr2a)$.

	studenci matematyki	studenci informatyki
pk_1	0.34	0.44
pk_0	0.48	0.67
wsp	-0.15	-0.23
$prop$	0.69	0.66

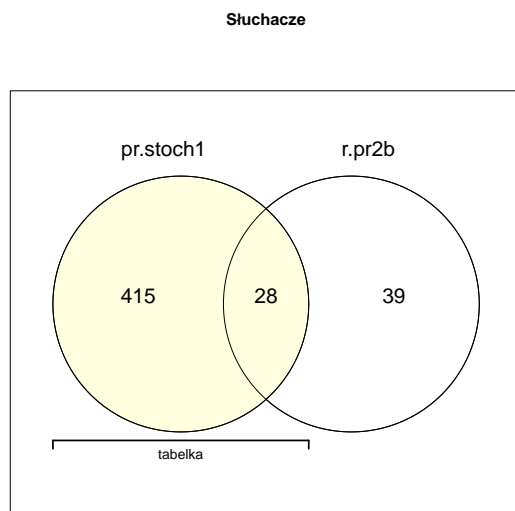
Tabela 4.8: Wartości współczynników analizy wsparcia i sukcesu $s_{p*przecietna}$ dla pary $(pr.stoch1, r.pr2a)$.

3. Para testowa (*pr.stoch1*, *r.pr2b*):

Studenci uczestniczący w którymkolwiek z kursów badanej pary testowej zostali przedstawieni sumarycznie na rysunku 4.4. Podobnie jak wyżej, 40% kursu testowego uczestniczyło w rozważanym kursie bazowym. Jednak w odróżnieniu od sytuacji poprzedniej, stanowią oni tylko 6% studentów zmagających się z *Procesami Stochastycznymi I*. Uwzględniając wiedzę o specjalizacji studentów i następstwie czasowym przedstawioną w tabeli 4.10 dowiadujemy się, że żaden ze studentów realizujących obydwa kursy nie zdecydował uczestniczyć najpierw w kursie bazowym, a dopiero kolejnych semestrach w kursie testowym. Studenci informatyki nie mieli później żadnych problemów z osiągnięciem sukcesu $s_{poz.ocena}$ z kursu bazowego. Wysoką skuteczność zanotowali również studenci matematyki, co zostało przedstawione w Tabeli 4.10.

Współczynnik wsparcia dla studentów informatyki jest wyższy niż dla studentów matematyki (odpowiednio 26%, 22%). Dodatkowo etykietę *Kursant* można przypisać dla 14% studentów informatyki, podczas gdy dla studentów matematyki jest to tylko 6%. Mimo to, *Rachunek Prawdopodobieństwa II** będzie kursem wspierającym dla *Procesów Stochastycznych I* dla studentów informatyki tylko dla 5 rozważanych par parametrów, podczas gdy dla studentów matematyki - dla 12.

Analizując tablice dwudzielne dla sukcesu $s_{p*przecietna}$ oraz wartości współczynników wsparcia (tab. 4.11 i tab. 4.12) widać, że współczynniki wsparcia wzrosły w stosunku do sukcesu pozytywnej oceny. W szczególności dla studentów informatyki osoba, która uzyskała ocenę dostateczną i była wcześniej słuchaczem kursu testowego spowodowała, że *Rachunek Prawdopodobieństwa II** będzie stanowił jeden z kursów testowych również dla ograniczenia $M = 1$. W przypadku studentów matematyki wartość tego ograniczenia wzrosła aż do 10.



Rysunek 4.4: Diagram Venna dla pary (*pr.stoch1*, *r.pr2b*).

studenci matematyki			studenci informatyki		
r.pr2b \ pr.stoch1	pr.stoch1		r.pr2b \ pr.stoch1	pr.stoch1	
	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$		$s_{poz.oc} = 1$	$s_{poz.oc} = 0$
Kursant - 1	20	3	Kursant - 1	5	0
Kursant - 0	245	135	Kursant - 0	26	9

Tabela 4.9: Tablice dwudzielne dla analizy i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, r.pr2b)$.

	studenci matematyki	studenci informatyki
pk_1	0.879	1
pk_0	0.67	0.74
wsp	0.22	0.26
$prop$	1.35	1.35

Tabela 4.10: Wartości współczynników analizy wsparcia i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, r.pr2b)$.

studenci matematyki			studenci informatyki		
r.pr2b \ pr.stoch1	pr.stoch1		r.pr2b \ pr.stoch1	pr.stoch1	
	$s_{p*przt} = 1$	$s_{p*przt} = 0$		$s_{p*przt} = 1$	$s_{p*przt} = 0$
Kursant - 1	13	10	Kursant - 1	4	1
Kursant - 0	127	253	Kursant - 0	15	20

Tabela 4.11: Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{p*przecietna}$ dla pary $(pr.stoch1, r.pr2b)$.

	studenci matematyki	studenci informatyki
pk_1	0.57	0.8
pk_0	0.33	0.43
wsp	0.23	0.37
$prop$	1.69	1.87

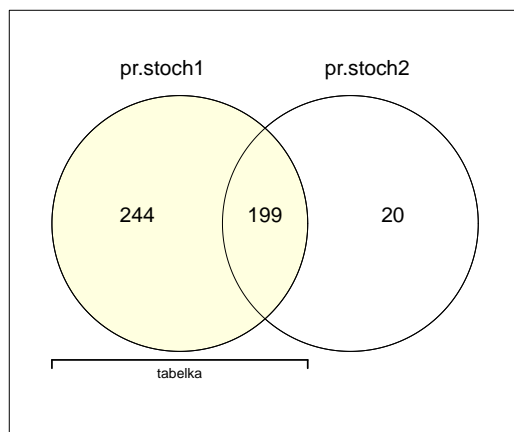
Tabela 4.12: Wartości współczynników analizy wsparcia i sukcesu $s_{p*przecietna}$ dla pary $(pr.stoch1, r.pr2b)$.

4. Para testowa ($pr.stoch1, pr.stoch2$):

Ostatnia para testowa składa się z kursów, co do których kolejność realizacji w większości przypadków jest określona - kurs testowy stanowi kontynuację kursu bazowego. Można to zauważyć na diagramie Venna z Rysunku 4.5 oraz w tablicach dwudzielnych z Tabeli 4.13. Okazuje się, że 99% studentów, którzy zdecydowali się zrealizować *Procesy Stochastyczne II* uczestniczyło również w *Procesach Stochastycznych I*, a tylko 3% z nich realizowało je w późniejszych cyklach dydaktycznych - byli to wyłącznie studenci matematyki. Taka postać tabeli dwudzielnej skutkuje nieokreślonością większości współczynników wsparcia dla studentów informatyki, co jest widoczne w Tabeli 4.14. Studenci matematyki, którzy zdecydowali się zrealizować rozważaną parę kursów w odwrotnej kolejności, niż większość studentów, osiągnęli sukces $s_{poz.ocena}$ tylko w 50% przypadków. Wynik ten nie przełożył się na pozytywną wartość współczynnika wsparcia. Skutkuje to brakiem przynależności kursu testowego do zbiorów $W_{p.stoch1}^{(\alpha, M)}$ dla każdego z rozważanych par parametrów (α, M) .

Tutaj analiza sukcesu oceny ponad*przeciętnej dla studentów informatyki jest analogiczna jak dla sukcesu pozytywnej - możliwa jest migracja danych jedynie w obrębie wierszy tablic dwudzielnych, dlatego współczynniki wsparcia i proporcji pozostają w dalszym ciągu nieokreślone. Widać dodatkowo, że wynik ponad*przeciętny uzyskuje co drugi student informatyki i co 3 student matematyki (w ujęciu globalnym). Współczynnik wsparcia dla studentów matematyki wynosi -18% , przez co *Procesy Stochastyczne II* nie są przedmiotem wspierającym dla poprzedzającego go kursu *Procesy Stochastyczne I* dla żadnego z rozważanych parametrów (α, M) .

Słuchacze



Rysunek 4.5: Diagram Venna dla pary ($pr.stoch1, pr.stoch2$).

studenci matematyki			studenci informatyki		
pr.stoch2 \ pr.stoch1	pr.stoch1		pr.stoch2 \ pr.stoch1	pr.stoch1	
	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$		$s_{poz.oc} = 1$	$s_{poz.oc} = 0$
Kursant - 1	3	3	Kursant - 1	0	0
Kursant - 0	262	135	Kursant - 0	31	9

Tabela 4.13: Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{poz.ocena}$ dla pary $(pr.stoch1, pr.stoch2)$.

	studenci matematyki	studenci informatyki
pk_1	0.5	NA
pk_0	0.66	0.775
wsp	-0.16	NA
$prop$	0.76	NA

Tabela 4.14: Wartości współczynników i sukcesu $s_{poz.ocena}$ analizy wsparcia dla pary $(pr.stoch1, pr.stoch2)$.

studenci matematyki			studenci informatyki		
pr.stoch2 \ pr.stoch1	pr.stoch1		pr.stoch2 \ pr.stoch1	pr.stoch1	
	$s_{p*przt} = 1$	$s_{p*przt} = 0$		$s_{p*przt} = 1$	$s_{p*przt} = 0$
Kursant - 1	1	5	Kursant - 1	0	0
Kursant - 0	139	258	Kursant - 0	19	21

Tabela 4.15: Tablice dwudzielne dla analizy wsparcia i sukcesu $s_{p*przecietna}$ dla pary $(pr.stoch1, pr.stoch2)$.

	studenci matematyki	studenci informatyki
pk_1	0.17	NA
pk_0	0.35	0.48
wsp	-0.18	NA
$prop$	0.48	NA

Tabela 4.16: Wartości współczynników analizy wsparcia i sukcesu $s_{p*przecietna}$ dla pary $(pr.stoch1, pr.stoch2)$.

Powyższe obserwacje prowadzą do następujących zbiorów wspierających $W_{p.stoch1}^{(\alpha,M)}$, ograniczając się jedynie do w/w kursów testowych:

Sukces pozytywnej oceny

Studenci matematyki:

$$W_{p.stoch1}^{(\alpha,M)} = \begin{cases} \{\text{jtrrz, r.pr2b}\}, & \text{dla } (\alpha, M) \in \{(0, 0), (0.1, 0), (0.15, 0), (0.2, 0), (0, 1), \\ & (0.1, 1), (0.15, 1), (0.2, 1), (0, 3), (0.1, 3), \\ & (0.15, 3), (0.2, 3)\}, \\ \emptyset, & \text{w p.p. .} \end{cases}$$

Studenci informatyki:

$$W_{p.stoch1}^{(\alpha,M)} = \begin{cases} \{\text{jtrrz, r.pr2b}\}, & \text{dla } (\alpha, M) \in \{(0, 0), (0.1, 0), (0.15, 0), (0.2, 0), \\ & (0.25, 0)\}, \\ \{\text{r.prb2}\}, & \text{dla } (\alpha, M) \in \{(0.25, 0)\}, \\ \emptyset, & \text{w p.p. .} \end{cases}$$

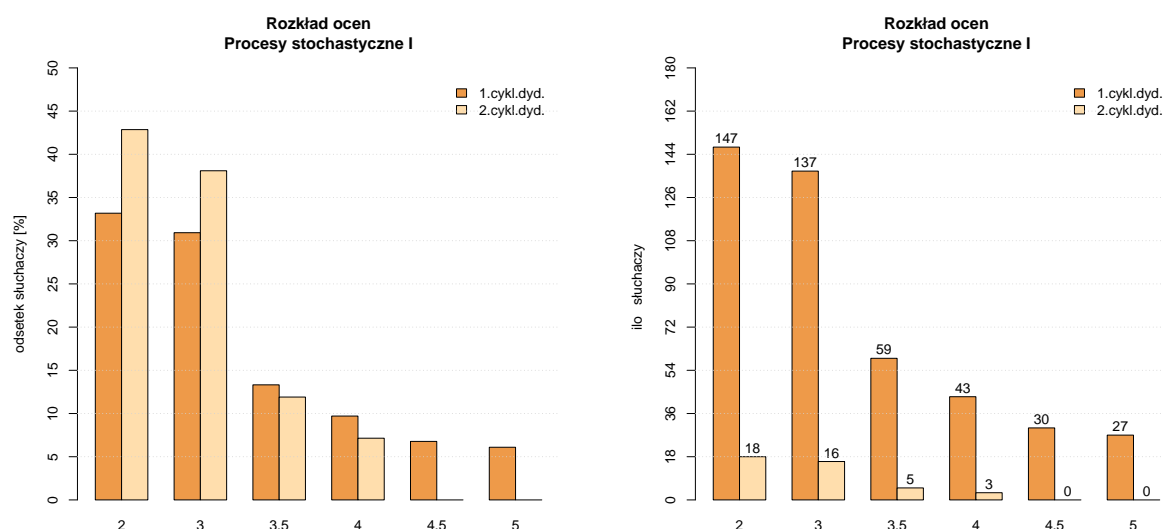
Sukces oceny ponad*przeciętnej

Studenci matematyki:

$$W_{p.stoch1}^{(\alpha,M)} = \begin{cases} \{\text{jtrrz, r.pr2b}\}, & \text{dla } (\alpha, M) \in \{(0, 0), (0.1, 0), (0.15, 0), (0.2, 0), (0, 1), \\ & (0.1, 1), (0.15, 1), (0.2, 1), (0, 3), (0.1, 3), \\ & (0.15, 3), (0.2, 3), (0, 5), (0.1, 5), (0.15, 5), \\ & (0.20, 5), (0, 7), (0.1, 7), (0.15, 7), \\ & (0.20, 7)\}, \\ \{\text{jtrrz}\}, & \text{dla } (\alpha, M) \in \{(0.25, 0), (0.25, 1), (0.25, 3), (0.25, 5), \\ & (0.25, 7), (0.3, 0), (0.3, 1), (0.3, 3), (0.3, 5), \\ & (0.3, 7)\}, \\ \{\text{r.prb2}\}, & \text{dla } (\alpha, M) \in \{(0, 10), (0.1, 10), (0.15, 10), (0.20, 10)\}, \\ \emptyset, & \text{w p.p. .} \end{cases}$$

Studenci informatyki:

$$W_{p.stoch1}^{(\alpha,M)} = \begin{cases} \{\text{jtrrz, r.pr2b}\}, & \text{dla } (\alpha, M) \in \{(0, 0), (0.1, 0), (0.15, 0), (0.2, 0), (0.25, \\ & 0), (0.3, 0), (0.35, 0)\}, \\ \{\text{jtrrz}\}, & \text{dla } (\alpha, M) \in \{(0.40, 0), (0.45, 0), (0.50, 0)\}, \\ \{\text{r.prb2}\}, & \text{dla } (\alpha, M) \in \{(0, 1), (0.1, 1), (0.15, 1), (0.2, 1), (0.25, \\ & 1), (0.30, 1), (0.35, 1)\}, \\ \emptyset, & \text{w p.p. .} \end{cases}$$



Rysunek 4.6: Rozkład ocen dla Procesów stochastycznych I - samowsparcie: a) unormowany b) w ujęciu ilościowym.

Samowsparcie

Ostatnim etapem analizy szczegółowej jest zbadanie kursu bazowego w celu określenia klas (α, M) -samowspierających, do których ten kurs należy (def. 3.5). Pierwszym krokiem jest zbadanie odpowiednich rozkładów ocen (rys. 4.6). Serie danych odnoszą się do numeru podejścia do *Procesów Stochastycznych I*. Niepokojący jest fakt, że odsetek studentów, którzy nie zdołali uzyskać zaliczenia podczas ponownej realizacji przedmiotu wzrósł do blisko 45%. Należy więc przypuszczać, że w tym przypadku ponownie uczestniczenie w kursie nie wpływa korzystnie na osiągnięcie zarówno sukcesu $s_{poz.ocena}$ jak i $s_{p*przecietna}$. Aby zweryfikować to przypuszczenie, należy odwołać się do Tabel 4.17 i 4.19. Pierwsza z nich przedstawia tabelicę dwudzielną dla sukcesu pozytywnej oceny. W przypadku *Procesów Stochastycznych I* żaden z 9 studentów informatyki, którzy nie uzyskali zaliczenia podczas swojego pierwszego podejścia w pierwszym cyklu dydaktycznym nie realizował kursu ponownie. Skutkuje to nieokreślonością głównego współczynnika wsparcia (tab. 4.18). Warto odnotować, że wartości przedstawiające rozkład ocen w kolejnym cyklu dydaktycznym dla studentów matematyki można teraz odczytać wprost z zamieszczonego wyżej histogramu. Istotnie okazało się, że ponowne uczestniczenie w kursie bazowym nie wpływa korzystnie na uzyskanie zaliczenia w pierwszym terminie. W przypadku sukcesu ponad*przeciętnej dysproporcje zwiększyły się jeszcze bardziej - podczas drugiego cyklu dydaktycznego tylko 19% studentów matematyki uzyskało wynik wyższy niż 3, podczas gdy w pierwszym cyklu dydaktycznym udało się to 35% (tab. 4.20).

studenci matematyki			studenci informatyki		
pr.stoch1			pr.stoch1		
	$s_{poz.oc} = 1$	$s_{poz.oc} = 0$		$s_{poz.oc} = 1$	$s_{poz.oc} = 0$
cykl.dyd.2	24	18	cykl.dyd.2	0	0
cykl.dyd.1	265	138	cykl.dyd.1	31	9

Tabela 4.17: Tablice dwudzielne dla problemu samowsparcia i sukcesu $s_{poz.ocena}$ dla kursu pr.stoch1.

	st. matematyki	st. informatyki
pk_1	0.57	NA
pk_0	0.66	0.78
wsp	-0.09	NA
$prop$	0.87	NA

Tabela 4.18: Wartości współczynników dla problemu samowsparcia i sukcesu $s_{poz.ocena}$ dla kursu pr.stoch1.

studenci matematyki			studenci informatyki		
pr.stoch1			pr.stoch1		
	$s_{p*przt} = 1$	$s_{p*przt} = 0$		$s_{p*przt} = 1$	$s_{p*przt} = 0$
cykl.dyd.2	8	34	cykl.dyd.2	0	0
cykl.dyd.1	140	263	cykl.dyd.1	19	21

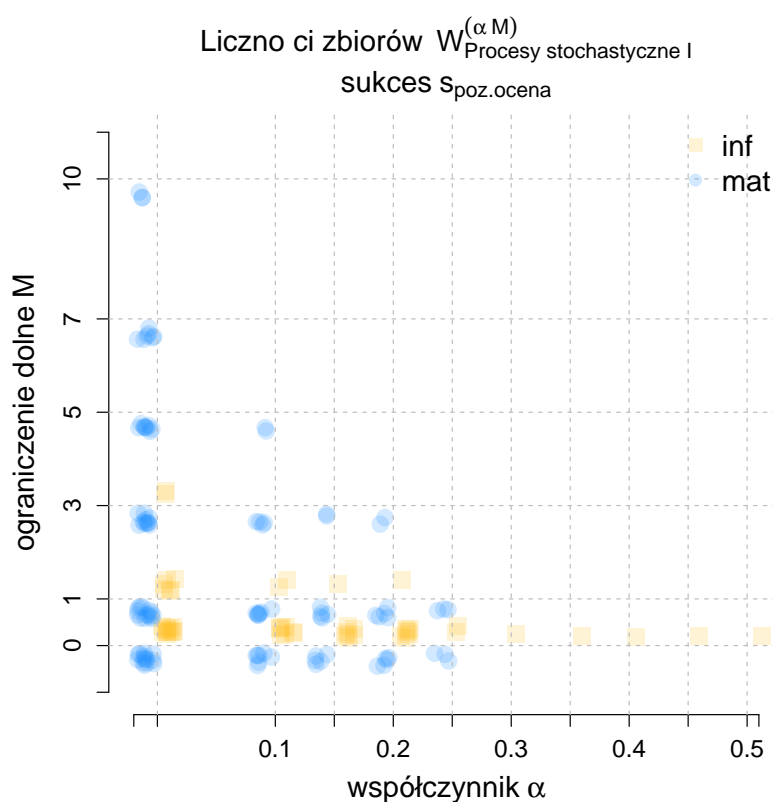
Tabela 4.19: Tablice dwudzielne dla problemu samowsparcia i sukcesu $s_{p*przecietna}$ dla kursu pr.stoch1.

	st. matematyki	st. informatyki
pk_1	0.19	NA
pk_0	0.35	0.48
wsp	-0.16	NA
$prop$	0.55	NA

Tabela 4.20: Wartości współczynników dla problemu samowsparcia i sukcesu $s_{p*przecietna}$ dla kursu pr.stoch1.

Podsumowanie

Postępując zgodnie z powyższym schematem wyznacza się pełne zbiory (α, M) -wspierające. Zaprezentowane powyżej zapisy kłamrowe zastępuje się wówczas odpowiednimi wykresami ilustrującymi moc zbiorów wspierających $W_K^{(\alpha, M)}$ z Rysunku 4.7. Wykresy te ułatwiają porównywanie wyników pomiędzy studentami matematyki i informatyki. Uzyskany efekt powstał w wyniku dodania zaburzenia losowego w obrębie danej pary parametrów. Dlatego jego integralną częścią jest tabela rzeczywistych wartości (tab. 4.21). Poszerzony zestaw wykresów i tabel można znaleźć w załączniku A. Dodatkowo można analizować wykres obrazujący zależności pomiędzy współczynnikami pk_0 i pk_1 , już w obrębie wybranej specjalizacji studentów. Przykład dla sukcesu pozytywnej oceny i studentów matematyki przedstawiono na Rysunku 4.8.



Rysunek 4.7: Wykres ilości kursów wspierających dla sukcesu $s_{poz.ocena}$.

4.2. Graficzne przedstawienie sieci (samo)wspierania

Po wykonaniu szczegółowej analizy i doboru odpowiednich parametrów należało wybrać końcowy sposób przedstawienia uzyskanych wyników. Posiłkując się teorią sieci przyczynowych i Bayesowskich, skonstruowano w tym celu grafy oddzielnie dla rozważanych definicji sukcesu, z uwzględnieniem podziału na kierunki studiów. Poniżej można znaleźć ich opis, a w dalszej części rozdziału - postaci czterech sporządzonych grafów.

	M-0	M-1	M-3	M-5	M-7	M-10
α 0	14-12	14-6	11-2	9-0	7-0	3-0
α 0.1	7-8	7-2	4-0	2-0	0-0	0-0
α 0.15	5-7	5-1	2-0	0-0	0-0	0-0
α 0.2	5-7	5-1	2-0	0-0	0-0	0-0
α 0.25	3-2	3-0	0-0	0-0	0-0	0-0
α 0.3	0-1	0-0	0-0	0-0	0-0	0-0
α 0.35	0-1	0-0	0-0	0-0	0-0	0-0
α 0.4	0-1	0-0	0-0	0-0	0-0	0-0
α 0.45	0-1	0-0	0-0	0-0	0-0	0-0
α 0.5	0-1	0-0	0-0	0-0	0-0	0-0

Tabela 4.21: Ilość kursów wspierających dla sukcesu $s_{poz.ocena}$: Studenci matematyki-studenci informatyki. Na przykładzie kursu Procesy Stochastyczne I.

4.2.1. Konstrukcja

Graf prezentujący sieć (samo)wspierania jest grafem o wierzchołkach odpowiadających kursom z puli kursów. Występujące w nim krawędzie można podzielić na trzy kategorie:

1. Krawędzie ciągłe - odpowiadają tym parom testowym, dla których wykryto wspieranie o dostatecznie dużej sile, tzn. co najmniej na poziomie α . Ich skierowanie jest zgodne z tzw. kierunkiem wsparcia, czyli prowadzi od kursu testowego do kursu bazowego, zaś szerokość odpowiada zdyskretyzowanym wartościom współczynnika wsp . Zostało to przedstawione na Rysunku 4.9 i w Tabeli 4.22 odpowiednio.
2. Pętle - odpowiadają tym kursom, które stanowią klasę samowspierającą.
3. Krawędzie dodatkowe - mają na celu redukcję izolowanych wierzchołków grafu, które powstały po skonstruowaniu grafu w oparciu o krawędzie ciągłe. Połączenie tej postaci było dołączane do grafu tylko w sytuacji istotnego wsparcia, tzn. gdy $wsp > 0$.

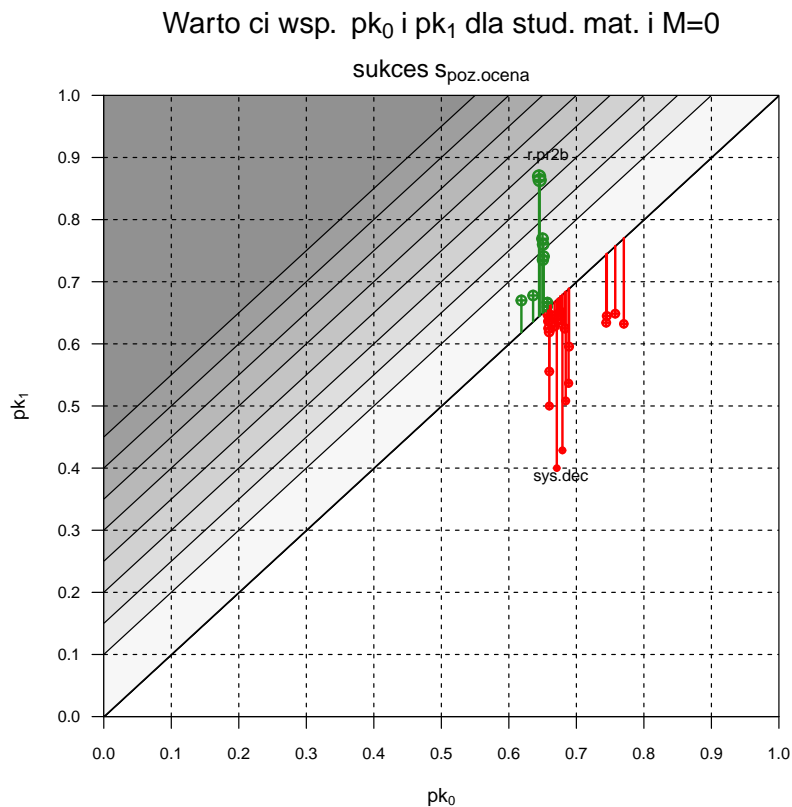
Dodatkowo grafy zawierają informację o wartości współczynnika proporcji w postaci etykiet dołączonych do każdych krawędzi.

4.2.2. Prezentacja

Końcowe postaci sieci (samo)wspierania znajdują się na Rysunkach 4.10, 4.11, 4.12, 4.13.

Analizując wyniki uzyskane dla studentów matematyki w obrębie sukcesu pozytywnej oceny widać, że relatywnie duże wsparcie jest związane z kursem *Ekonometria* - posiada on wiele przedmiotów, które ułatwiają jego zaliczenie w „pierwszym terminie”. Jednocześnie kurs ten nie stanowi wsparcia dla żadnego kursu. Dlatego racjonalnie jest realizować go w ostatnim etapie studiów. Podobne wnioski dotyczą *Teorii ryzyka w ubezpieczeniach*, czy *Modeli matematycznych w biologii i medycynie*. Z drugiej strony łatwo zauważyć kursy, które same nie posiadają przedmiotów wspierających, ale stanowią wsparcie dla innych. Są to *Algebra I**, *Statystyka II*, *Rachunek prawdopodobieństwa II**.

W przypadku studentów informatyki widocznych jest zdecydowanie mniej zależności. Na pierwszy plan wysuwają się kursy: *Inżynieria finansowa*, *Matematyka dyskretna* i *Równania różniczkowe cząstkowe* - przed ich realizacją warto rozważyć uczestnictwo w innych kursach, które je uzupełniają. W szczególności *Matematyka obliczeniowa I* wpływa korzystnie na wszystkie wymienione wyżej kursy. Innym kursem, który w myśl sieci (samo)wspierających może przynieść wymierne korzyści dla kilku kursów jednocześnie są *Funkcje analityczne*.



Rysunek 4.8: Wartości współczynników pk_1 i pk_0 dla *pr.stoch1*



Rysunek 4.9: Połączenie pary testowej w sieci (samo)wspierania.

Studenci zainteresowani nie tyle pozytywnym zaliczeniem przedmiotów, co wysokim, powinni skupić się na analizie kolejnych grafów. Grafy te są dużo bogatsze. Wśród studentów matematyki widać, że realizacja kursów opatrzonych * stawia studentów w uprzywilejowanej pozycji. Dla każdego z rozważanych tego typu kursów można wskazać inny kurs, dla którego stanowią one wsparcie. Same zaś nie posiadają kursów zwiększających szansę uzyskania oceny ponad*przeciętnej o 25% - wyjątkiem jest *Rachunek prawdopodobieństwa I**, dla którego wsparciem są *Funkcje analityczne*. W przypadku tego sukcesu pojawiły się również pary kursów, które wzajemnie na siebie wpływają, jak np. *Teoria sterowania* i *Procesy stochastyczne I*. Studentów zainteresowanych obydwooma takimi kursami zachęca się do ich jednoczesnej realizacji, co w myśl sieci (samo)wspierania pozwoli zoptymalizować szansę ponad*przeciętnego wyniku w każdym z nich.

Wśród studentów informatyki uwypukliła się z kolei rola *Funkcji analitycznych*. Warto odnotować, że *Statystyka I*, która wspierała *Inżynierię finansową* w uzyskaniu pozytywnego wyniku, wspiera ją również w uzyskaniu wyniku ponad*przeciętnego. Sam kurs *Inżynieria*

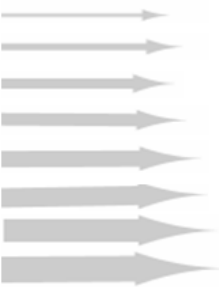
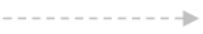

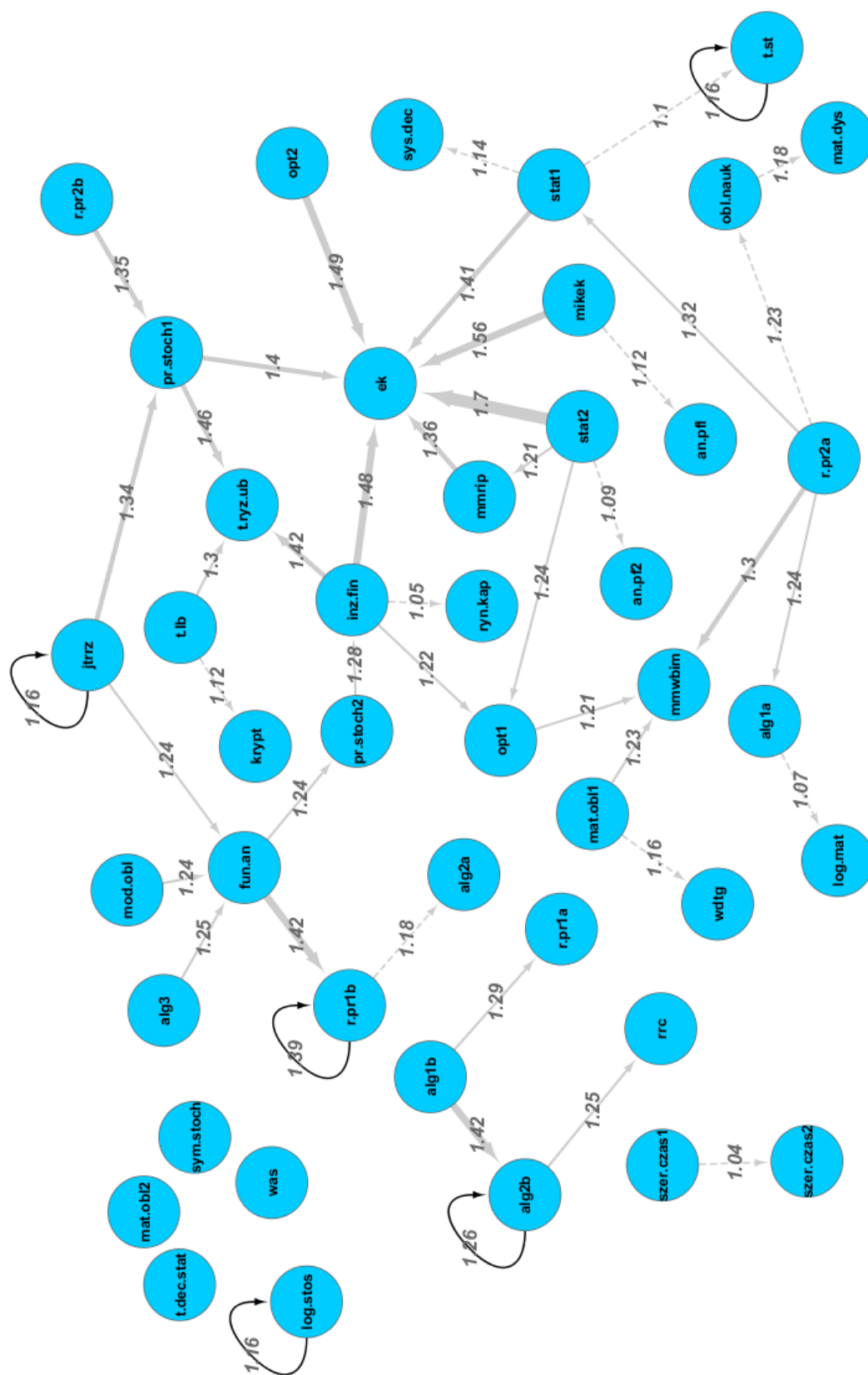
Rodzaj krawędzi	Współczynnik wsparcia	Postać
Krawędź ciągła	$[0.15 ; 0.20)$ $[0.20 ; 0.25)$ $[0.25 ; 0.30)$ $[0.30 ; 0.35)$ $[0.35 ; 0.40)$ $[0.40 ; 0.45)$ $[0.45 ; 0.50)$ $[0.50 ; 1.00)$	
Krawędź dodatkowa	-	
Pętla	-	

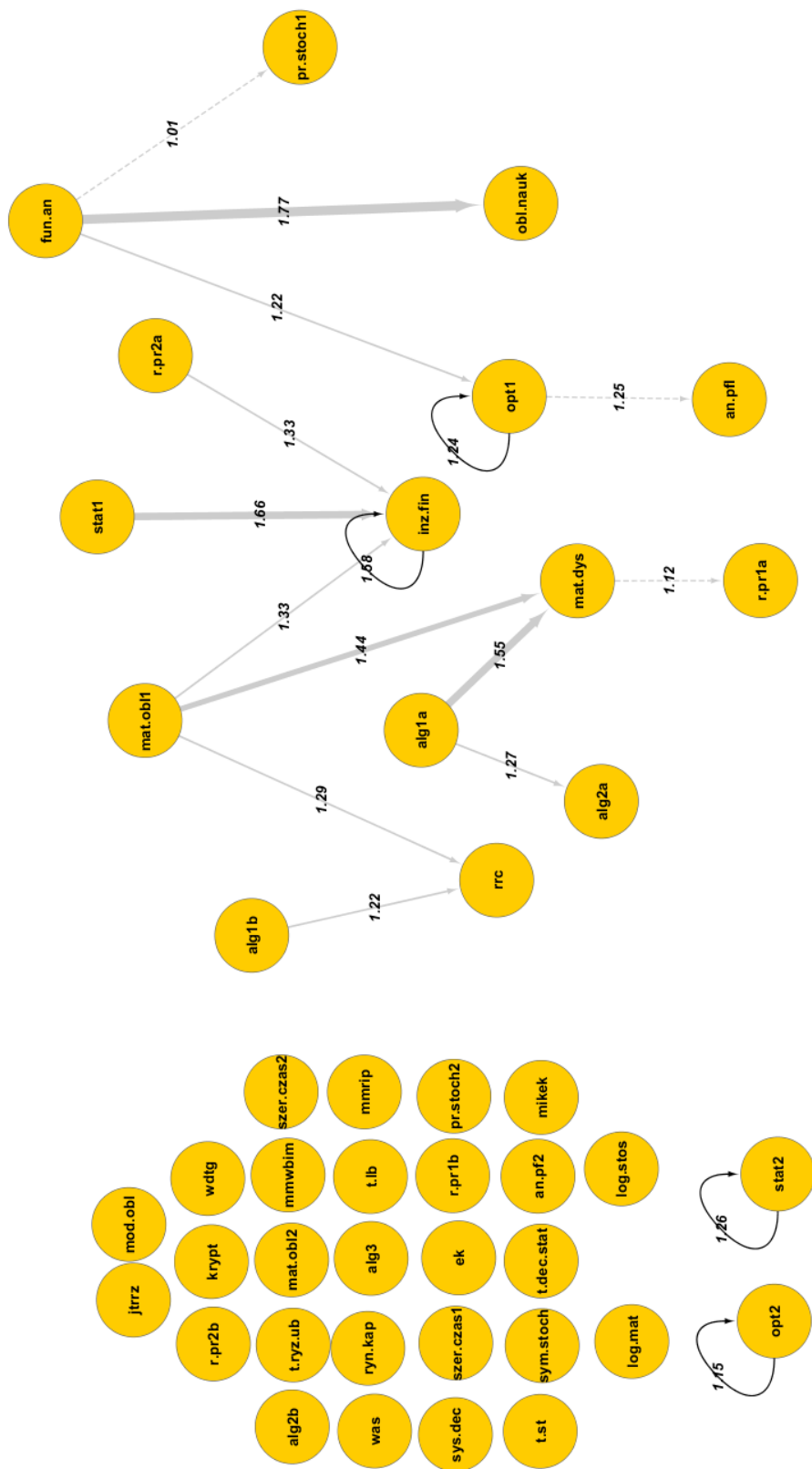
Tabela 4.22: Rodzaje krawędzi występujące w sieciach (samo)wspierania.

finansowa jest interesujący pod tym względem, iż jego rola odwróciła się w stosunku do studentów matematyki - z kursu wspierającego stał się on kursem wspieranym. Kursy opatrzone * nie odgrywają już tak istotnej roli jak wśród studentów matematyki. *Rachunek prawdopodobieństwa II** nie był kursem wspierającym, ani też sam nie stanowił wsparcia dla żadnej wartości $\alpha > 0$ przy ograniczeniu $M = 3$, o czym świadczy fakt, że nie został dołączony do zasadniczej części grafu nawet poprzez krawędź dodatkową (tab. 4.22). Z kolei *Algebra I** została dołączona do grafu tylko przez krawędzie dodatkowe. Jej kontynuacja, *Algebra II** sama posiada kurs wspierający, nie stanowiąc dla żadnego kursu wsparcia na poziomie co najmniej 25%. Dopiero *Rachunek prawdopodobieństwa I** stanowi kurs wspierający.

Należy zaznaczyć, że sieci (samo)wspierania nie zidentyfikowały dużej liczby kursów samowspierających. Nie można więc twierdzić, że powtarzanie kursu ułatwia jego realizację.

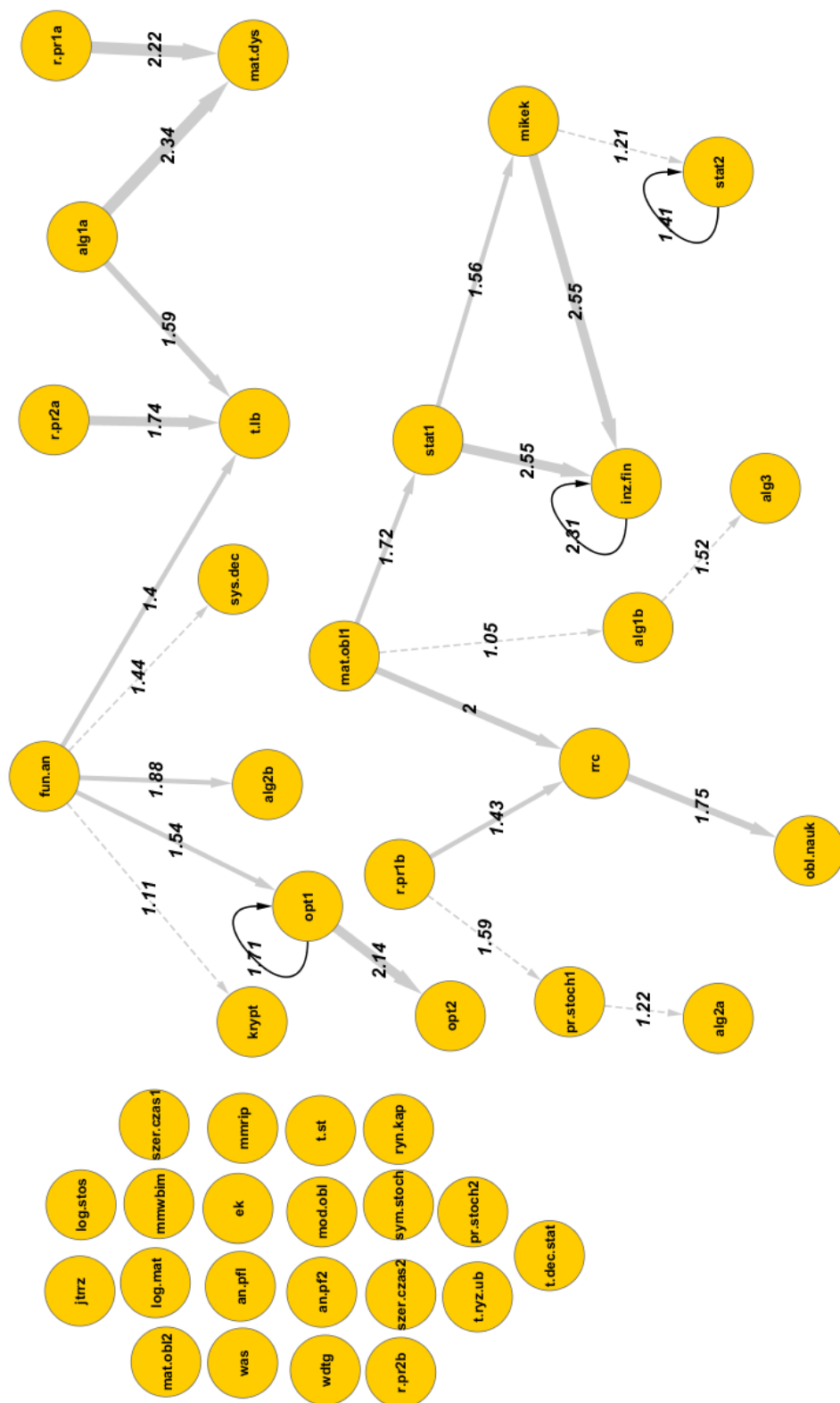


Rysunek 4.10: Sieć (samo)wspierania - sukces $s_{poz.ocena}$, studenci matematyki.



Rysunek 4.11: Sieć (samo)wspierania - sukces $s_{poz.ocena}$, studenci informatyki.





Rysunek 4.13: Sieć (samo)wspierania - sukces s_p^* przecietna, studenci informatyki.

Podsumowanie

Celem niniejszej pracy była analiza danych USOS związanych z Wydziałem Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego pod kątem znalezienia nowych zależności pomiędzy wybieranymi przez studentów kursami. Problem podzielono na dwie niezależne składowe.

Pierwsza skupia się na związkach pomiędzy wybieraniem poszczególnych grup kursów i bada ich wpływ na pozostałe kursy. Do tej analizy wykorzystano technikę sieci Bayesowskich, której idea została również opisana w prezentowanej pracy. Dodatkowo opisano główne funkcjonalności pakietów `gRain` i `bnlearn` programu `R`, które okazały się bardzo pomocne przy realizacji zadania. Ich wykorzystanie przyczyniło się do wykonania *mapy ciepła*, z której odczytuje się główne zależności w doborze kursów przez studentów oraz do zaproponowania struktury sieci Bayesowskiej - *USOS-sieci*. *USOS-sieć* umożliwia obserwację kodowanych w sieci zmian w czasie rzeczywistym. W pracy pokazano to zastosowanie w oparciu o arbitralnie wybrane ustalenia.

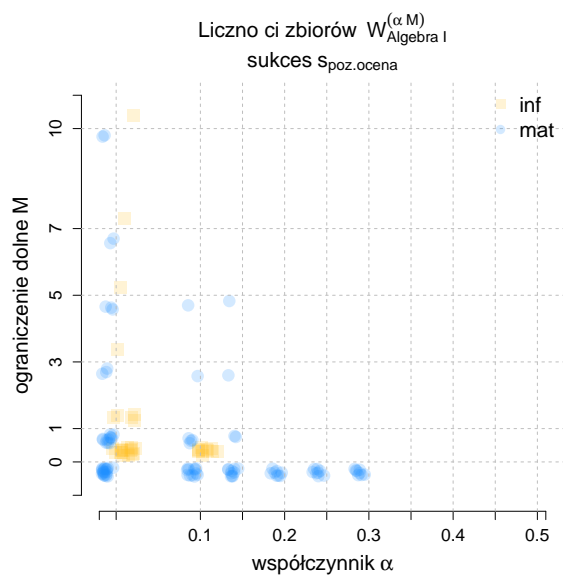
Druga składowa skupia się na znalezieniu kursów, które pozytywnie wpływają na pozostałe. Miarę wpływu sparametryzowano przez *sukces* związany z uzyskanymi przez studentów ocenami z egzaminów końcowych. W celu rozwiązania tego zadania opracowano metodę sieci (samo)wspierania, w całości zaimplementowaną w programie `R`. Dodatkowo zdecydowano się w niej uwzględnić zróżnicowanie studentów Wydziału MIMUW o ich zainteresowanie matematyczne bądź informatyczne. Z uwagi na brak tej informacji, została ona dodatkowo wyestymowana z danych. Uzyskane wyniki zaprezentowano w postaci grafów wykonanych już w programie `Cytoscape`.

Uzupełnieniem uzyskanych wyników związanych z sieciami (samo)wspierania jest Raport *Szczegóły wyznaczania sieci (samo)wspierających* umieszczony na płycie CD dołączonej do pracy. Zostały w nim zawarte pośrednie wyniki prowadzące do uzyskania końcowych grafów przyczynowych.

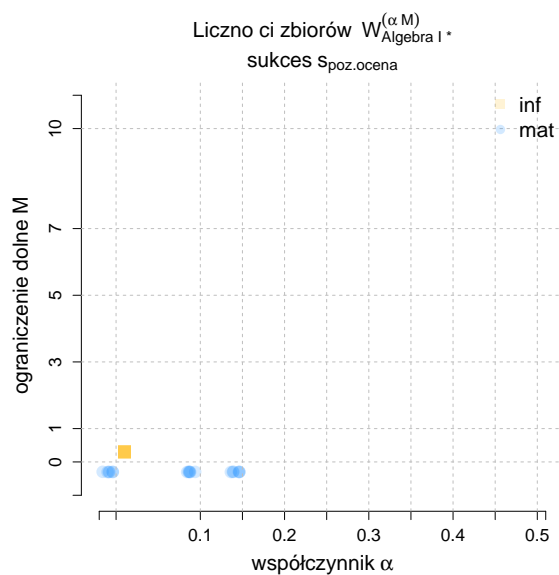
Dodatek A

Liczności zbiorów $W_K^{(\alpha, M)}$

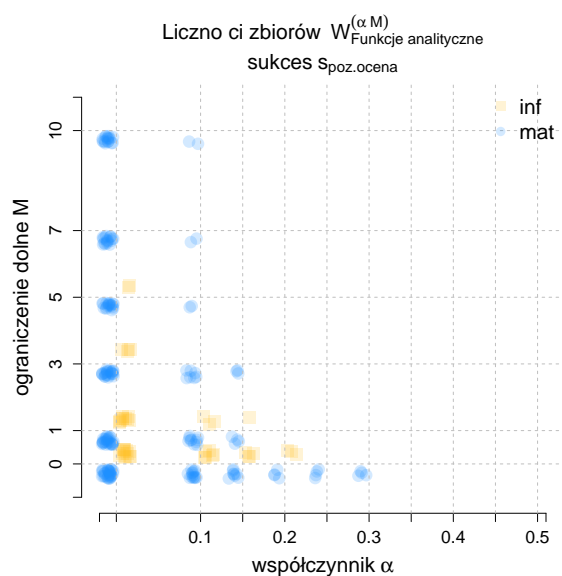
W tym dodatku zostały przedstawione wykresy i tabele liczności zbiorów $W_K^{(\alpha, M)}$, o których wspomniano w Rozdziale 4.1, dla przykładowych kursów. Dotyczą one tylko sukcesu $s_{poz.ocena}$, z uwzględnieniem podziału na studentów matematyki i informatyki, zgodnie z konwencją przedstawioną w w/w rozdziale. Analogiczne opracowanie dla sukcesu $s_{p*przecietna}$ można znaleźć w raporcie dołączonym do niniejszej pracy, już dla pełnej puli kursów.



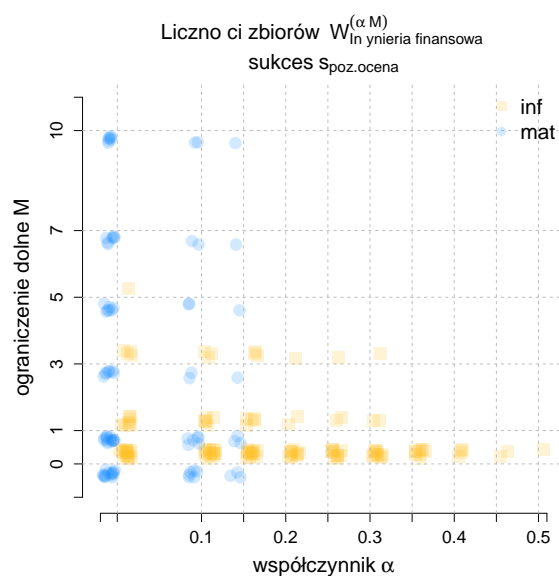
	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	20-16	12-5	3-1	3-1	2-1	2-1
alfa 0.1	12-11	4-0	1-0	1-0	0-0	0-0
alfa 0.15	10-0	2-0	1-0	1-0	0-0	0-0
alfa 0.2	8-0	0-0	0-0	0-0	0-0	0-0
alfa 0.25	8-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	8-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0



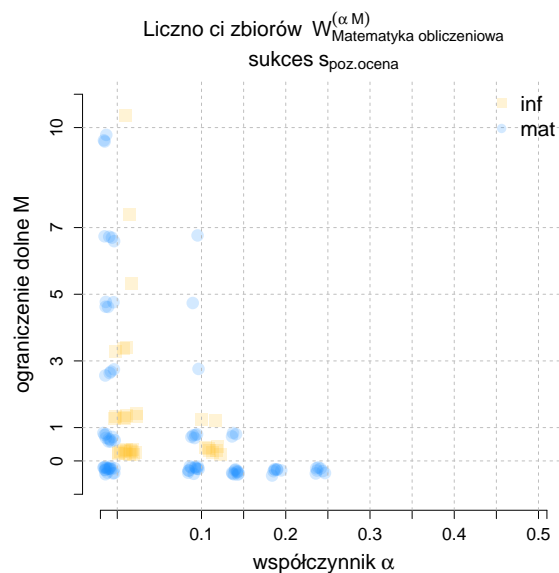
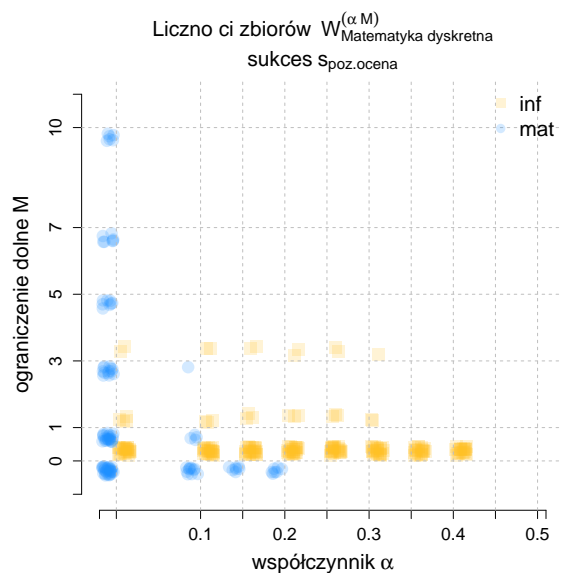
	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	6-8	0-0	0-0	0-0	0-0	0-0
alfa 0.1	6-0	0-0	0-0	0-0	0-0	0-0
alfa 0.15	6-0	0-0	0-0	0-0	0-0	0-0
alfa 0.2	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.25	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0



	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	31-13	27-10	23-4	18-2	17-0	17-0
alfa 0.1	14-6	10-3	7-0	2-0	2-0	2-0
alfa 0.15	8-4	4-1	3-0	0-0	0-0	0-0
alfa 0.2	4-3	0-0	0-0	0-0	0-0	0-0
alfa 0.25	4-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	4-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

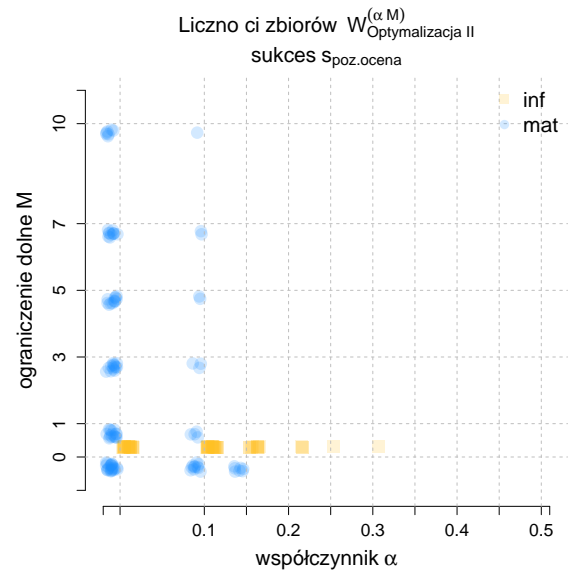
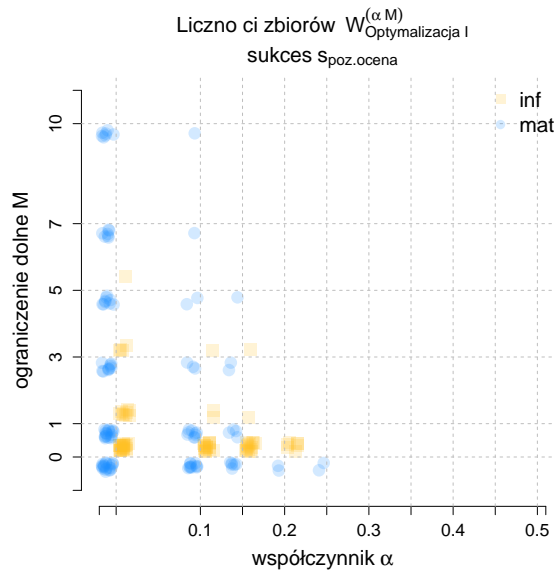


	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	14-17	14-8	7-4	7-1	7-0	7-0
alfa 0.1	6-15	6-6	2-3	2-0	2-0	2-0
alfa 0.15	3-14	3-5	1-3	1-0	1-0	1-0
alfa 0.2	0-11	0-2	0-1	0-0	0-0	0-0
alfa 0.25	0-11	0-2	0-1	0-0	0-0	0-0
alfa 0.3	0-11	0-2	0-1	0-0	0-0	0-0
alfa 0.35	0-9	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-4	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-2	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-1	0-0	0-0	0-0	0-0	0-0



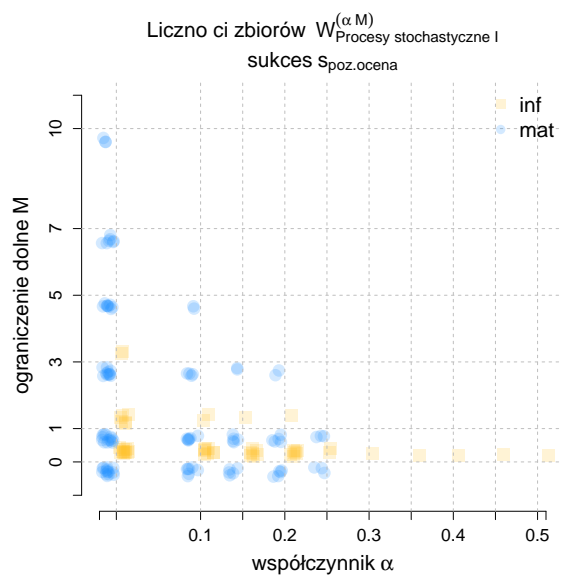
	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	30-25	23-4	13-2	8-0	7-0	6-0
alfa 0.1	11-24	4-3	1-2	0-0	0-0	0-0
alfa 0.15	7-24	0-3	0-2	0-0	0-0	0-0
alfa 0.2	7-24	0-3	0-2	0-0	0-0	0-0
alfa 0.25	0-24	0-3	0-2	0-0	0-0	0-0
alfa 0.3	0-23	0-2	0-1	0-0	0-0	0-0
alfa 0.35	0-21	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-21	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	17-12	10-7	4-3	4-1	4-1	3-1
alfa 0.1	12-7	5-2	1-0	1-0	1-0	0-0
alfa 0.15	10-0	3-0	0-0	0-0	0-0	0-0
alfa 0.2	7-0	0-0	0-0	0-0	0-0	0-0
alfa 0.25	7-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

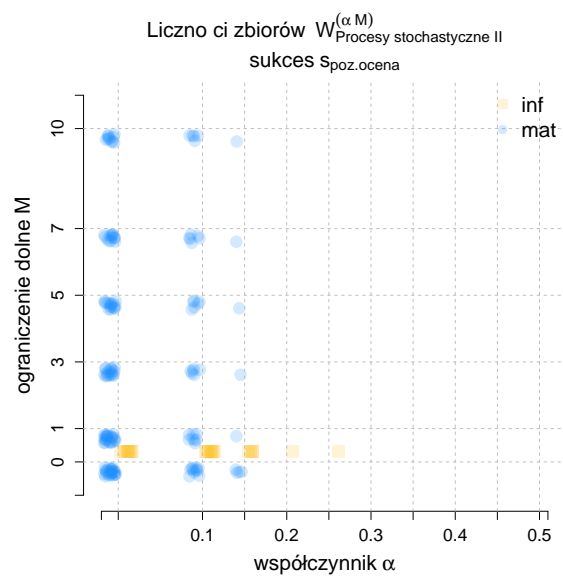


	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	20-23	18-9	10-4	9-1	7-0	7-0
alfa 0.1	11-16	9-2	3-1	2-0	1-0	1-0
alfa 0.15	6-15	4-1	2-1	1-0	0-0	0-0
alfa 0.2	2-7	0-0	0-0	0-0	0-0	0-0
alfa 0.25	2-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

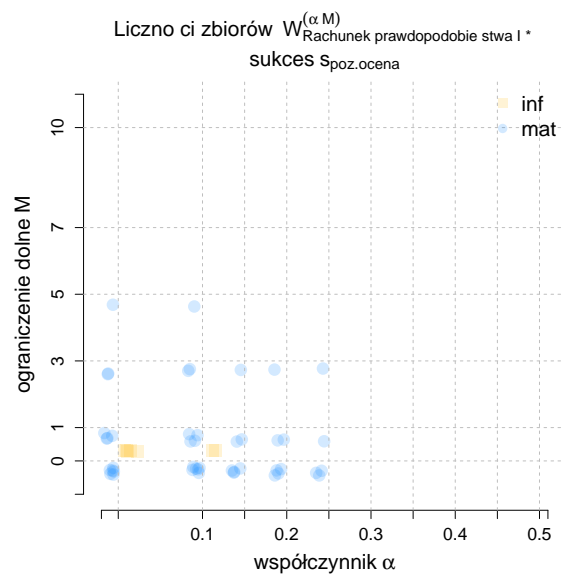
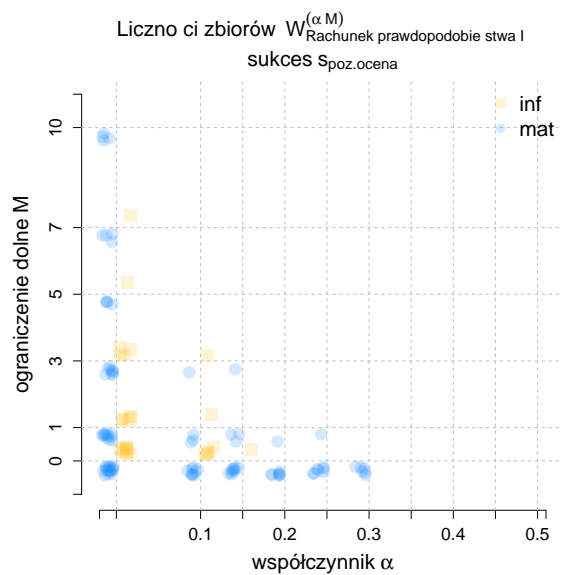
	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	23-22	16-0	14-0	12-0	11-0	7-0
alfa 0.1	11-22	4-0	3-0	2-0	2-0	1-0
alfa 0.15	7-10	0-0	0-0	0-0	0-0	0-0
alfa 0.2	0-3	0-0	0-0	0-0	0-0	0-0
alfa 0.25	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0



	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	14-12	14-6	11-2	9-0	7-0	3-0
alfa 0.1	7-8	7-2	4-0	2-0	0-0	0-0
alfa 0.15	5-7	5-1	2-0	0-0	0-0	0-0
alfa 0.2	5-7	5-1	2-0	0-0	0-0	0-0
alfa 0.25	3-2	3-0	0-0	0-0	0-0	0-0
alfa 0.3	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-1	0-0	0-0	0-0	0-0	0-0

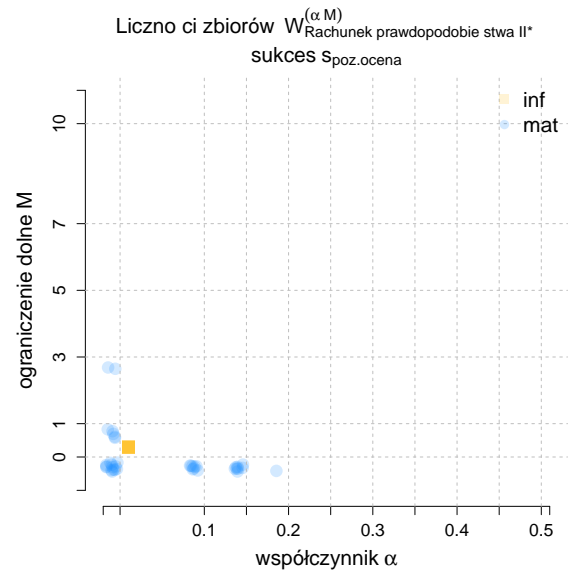
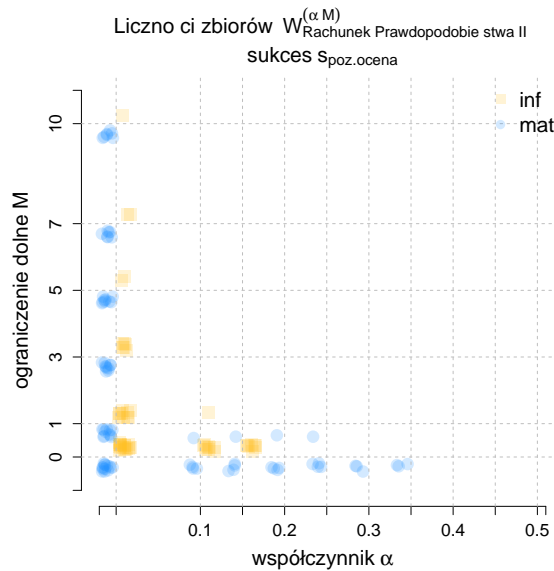


	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	27-10	24-0	18-0	16-0	14-0	11-0
alfa 0.1	11-10	8-0	6-0	6-0	6-0	5-0
alfa 0.15	4-5	1-0	1-0	1-0	1-0	1-0
alfa 0.2	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.25	0-1	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0



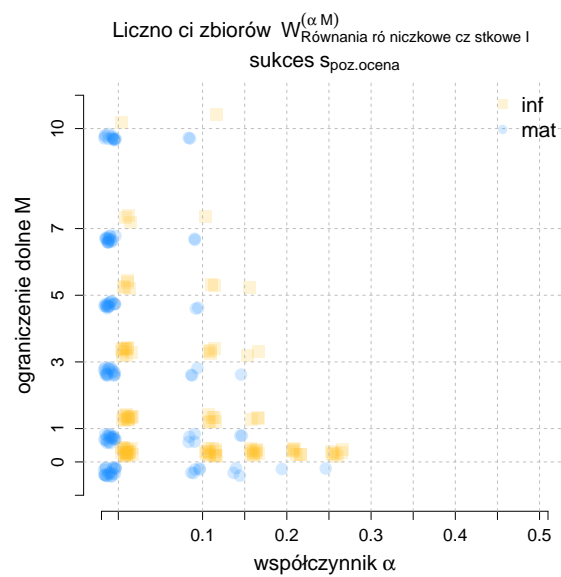
	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	17-9	11-5	8-4	4-1	4-1	4-0
alfa 0.1	9-5	3-1	1-1	0-0	0-0	0-0
alfa 0.15	9-1	3-0	1-0	0-0	0-0	0-0
alfa 0.2	7-0	1-0	0-0	0-0	0-0	0-0
alfa 0.25	7-0	1-0	0-0	0-0	0-0	0-0
alfa 0.3	6-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	6-5	4-0	2-0	1-0	0-0	0-0
alfa 0.1	6-2	4-0	2-0	1-0	0-0	0-0
alfa 0.15	4-0	2-0	1-0	0-0	0-0	0-0
alfa 0.2	4-0	2-0	1-0	0-0	0-0	0-0
alfa 0.25	3-0	1-0	1-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

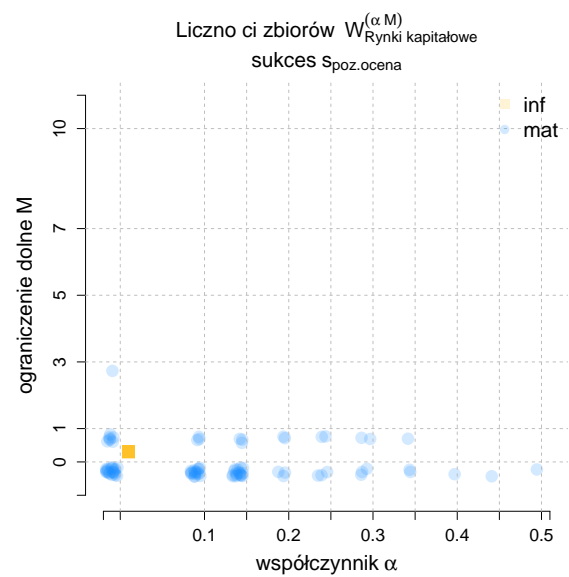


	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	14-18	11-10	10-6	9-2	7-2	7-1
alfa 0.1	4-9	1-1	0-0	0-0	0-0	0-0
alfa 0.15	4-8	1-0	0-0	0-0	0-0	0-0
alfa 0.2	4-0	1-0	0-0	0-0	0-0	0-0
alfa 0.25	4-0	1-0	0-0	0-0	0-0	0-0
alfa 0.3	3-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	3-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

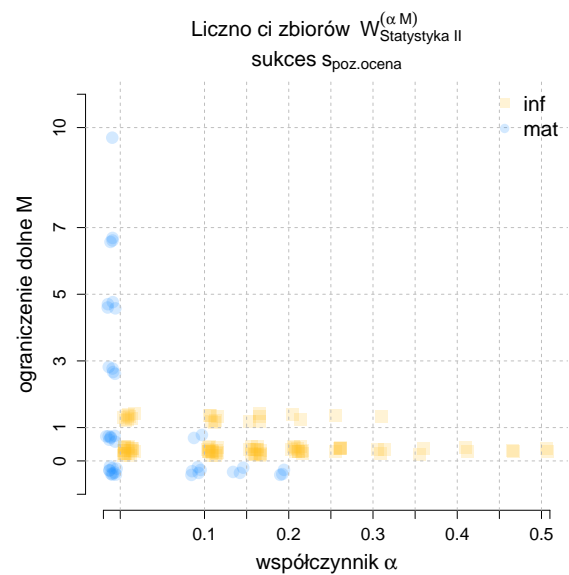
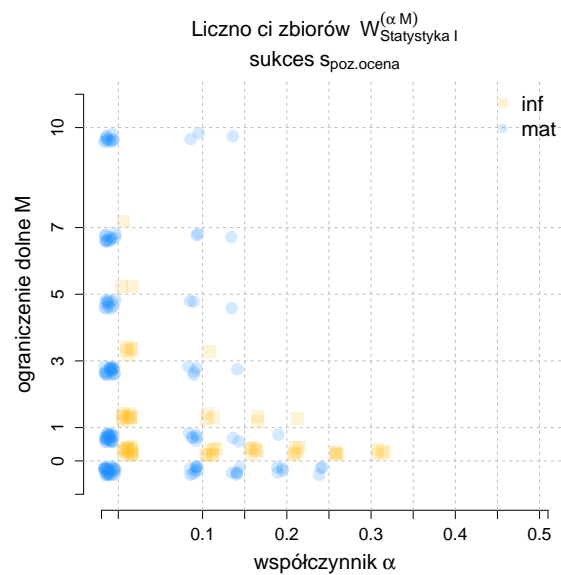
	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	13-12	5-0	2-0	0-0	0-0	0-0
alfa 0.1	8-0	0-0	0-0	0-0	0-0	0-0
alfa 0.15	8-0	0-0	0-0	0-0	0-0	0-0
alfa 0.2	1-0	0-0	0-0	0-0	0-0	0-0
alfa 0.25	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0



	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	20-19	19-12	17-10	16-5	16-3	13-1
alfa 0.1	5-13	4-6	3-4	2-2	2-1	2-1
alfa 0.15	3-10	2-3	1-2	0-1	0-0	0-0
alfa 0.2	1-7	0-0	0-0	0-0	0-0	0-0
alfa 0.25	1-7	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

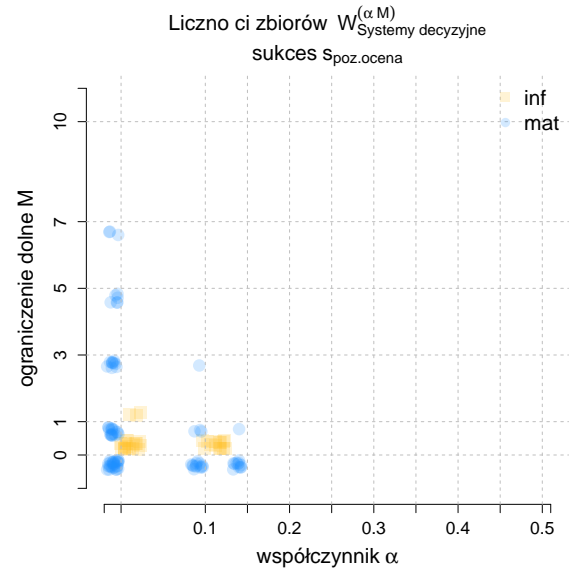
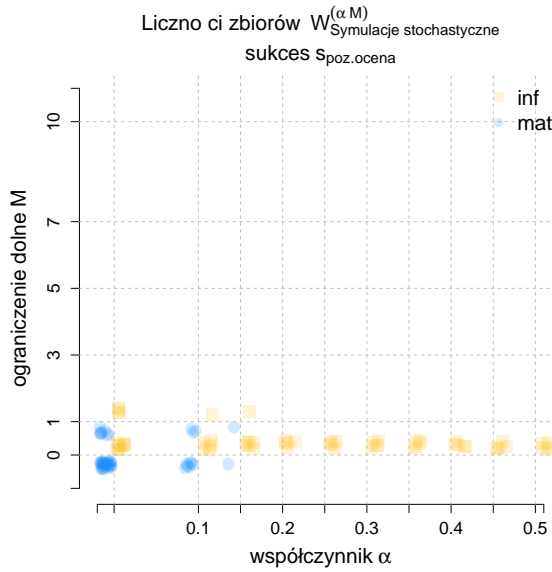


	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	21-16	7-0	1-0	0-0	0-0	0-0
alfa 0.1	17-0	3-0	0-0	0-0	0-0	0-0
alfa 0.15	17-0	3-0	0-0	0-0	0-0	0-0
alfa 0.2	3-0	2-0	0-0	0-0	0-0	0-0
alfa 0.25	3-0	2-0	0-0	0-0	0-0	0-0
alfa 0.3	3-0	2-0	0-0	0-0	0-0	0-0
alfa 0.35	2-0	1-0	0-0	0-0	0-0	0-0
alfa 0.4	1-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	1-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	1-0	0-0	0-0	0-0	0-0	0-0



	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	26-14	23-10	19-5	12-2	12-1	11-0
alfa 0.1	9-7	6-3	4-1	2-0	2-0	2-0
alfa 0.15	5-6	2-2	1-0	1-0	1-0	1-0
alfa 0.2	4-5	1-1	0-0	0-0	0-0	0-0
alfa 0.25	3-4	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-4	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	10-17	7-8	4-0	4-0	3-0	1-0
alfa 0.1	5-15	2-6	0-0	0-0	0-0	0-0
alfa 0.15	3-13	0-4	0-0	0-0	0-0	0-0
alfa 0.2	3-11	0-2	0-0	0-0	0-0	0-0
alfa 0.25	0-5	0-1	0-0	0-0	0-0	0-0
alfa 0.3	0-3	0-1	0-0	0-0	0-0	0-0
alfa 0.35	0-2	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-2	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-2	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-2	0-0	0-0	0-0	0-0	0-0



	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	30-9	8-4	0-0	0-0	0-0	0-0
alfa 0.1	7-6	3-1	0-0	0-0	0-0	0-0
alfa 0.15	1-6	1-1	0-0	0-0	0-0	0-0
alfa 0.2	0-5	0-0	0-0	0-0	0-0	0-0
alfa 0.25	0-5	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-5	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-5	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-5	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-5	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-5	0-0	0-0	0-0	0-0	0-0

	M-0	M-1	M-3	M-5	M-7	M-10
alfa 0	27-17	17-3	10-0	6-0	3-0	0-0
alfa 0.1	13-14	3-0	1-0	0-0	0-0	0-0
alfa 0.15	11-0	1-0	0-0	0-0	0-0	0-0
alfa 0.2	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.25	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.3	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.35	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.4	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.45	0-0	0-0	0-0	0-0	0-0	0-0
alfa 0.5	0-0	0-0	0-0	0-0	0-0	0-0

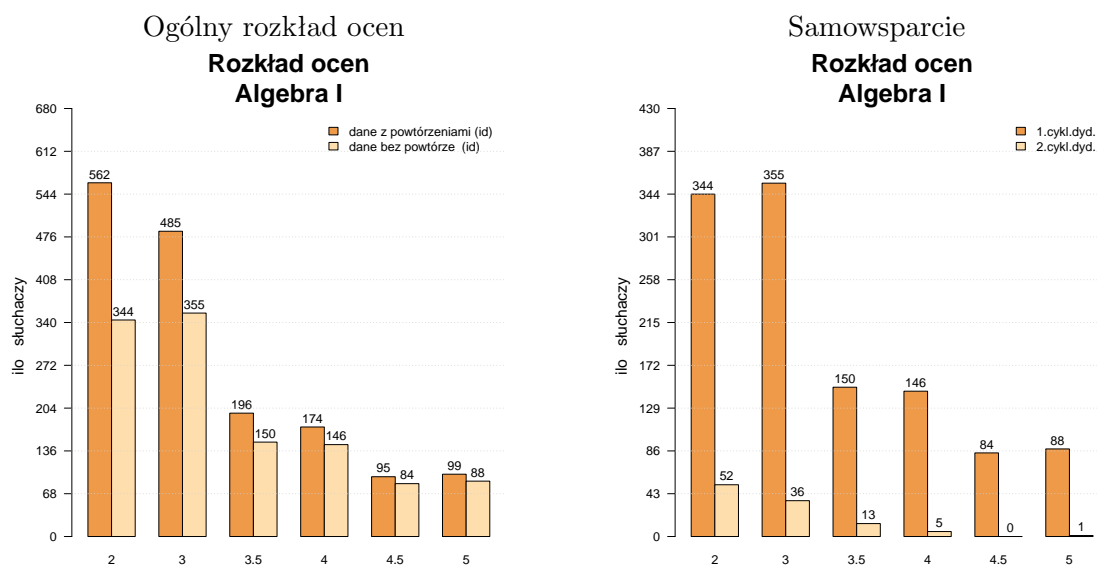
Tabela A.1: Wykresy i tabele liczności zbiorów $W_K^{(\alpha, M)}$ dla sukcesu $s_{\text{poz.ocena}}$.

Dodatek B

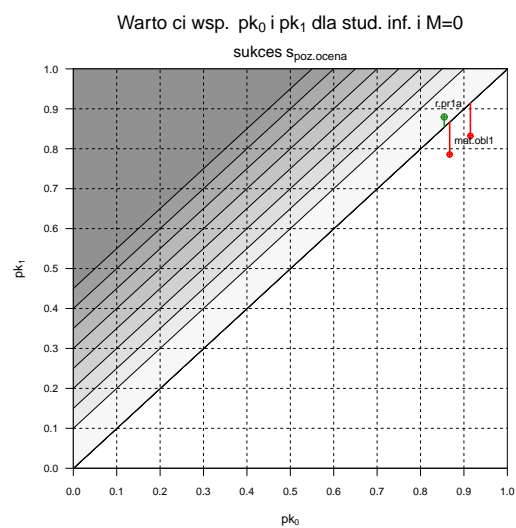
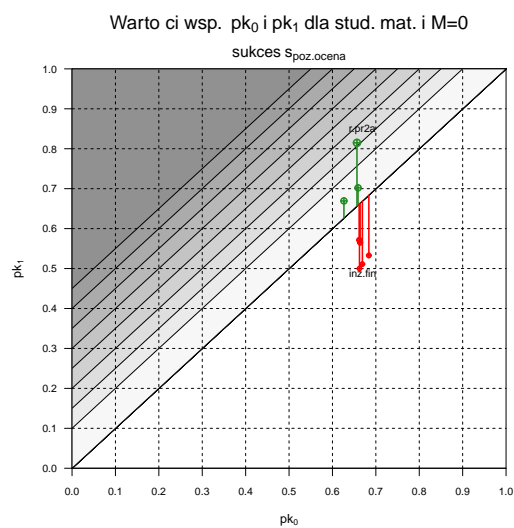
Raport szczegółowy - wybrane elementy

Uzupełnieniem niniejszej pracy magisterskiej jest Raport *Szczegóły wyznaczania sieci (samo)wspierania*, znajdujący się na dołączonej do pracy płycie CD. Raport przedstawia poszczególne komponenty analizy szczegółowej wykonanej dla wszystkich rozważanych kursów bazowych i par testowych. Jego opis, na przykładzie kursu bazowego *Procesy Stochastyczne I* oraz kilku wybranych kursów testowych, stanowi treść Rozdziału 4.1. Zamieszczony poniżej fragment Raportu nawiązuje do części Poznanie kursu bazowego, Samowsparcie, Podsumowanie, które można znaleźć w wyżej wymienionym rozdziale, w oparciu o kursy *Algebra I*, *Funkcje analityczne*, *Inżynieria finansowa*, *Matematyka dyskretna*, *Matematyka obliczeniowa*, *Optymalizacja I*, *Optymalizacja II*, *Rachunek Prawdopodobieństwa I*, *Rachunek Prawdopodobieństwa II*, *Równania różniczkowe cząstkowe I*, *Statystyka I*, *Statystyka II*, *Systemy decyzyjne*.

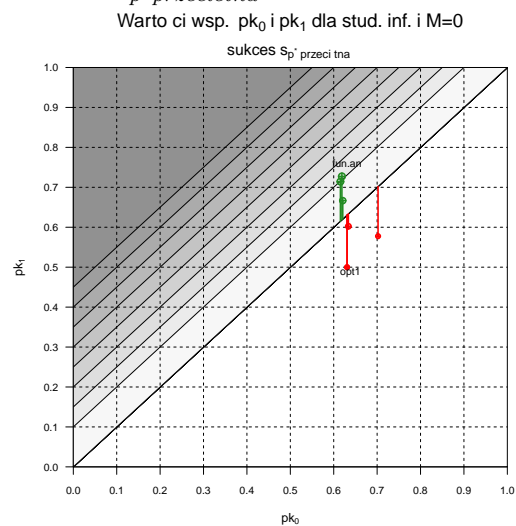
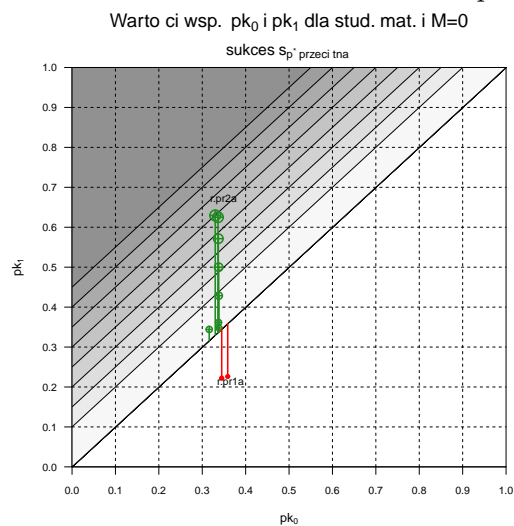
Kurs: Algebra I - alg1a



Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:



Podsumowanie wsparcia dla sukcesu $s_{p*przecitna}$:



Tablica dwudzielna dla problemu samowsparcia

	alg1a	
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	53/2	47/5
cykl.dyd.1	603/220	309/35

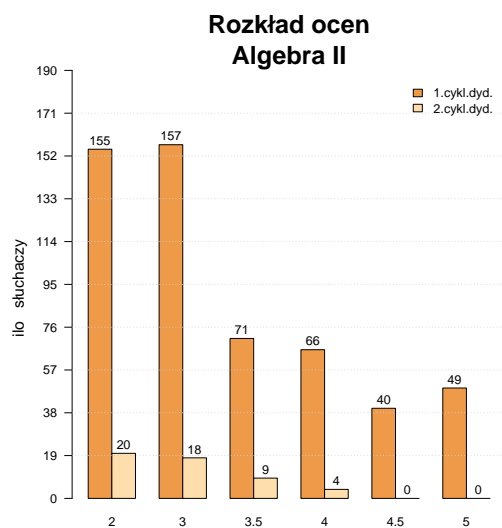
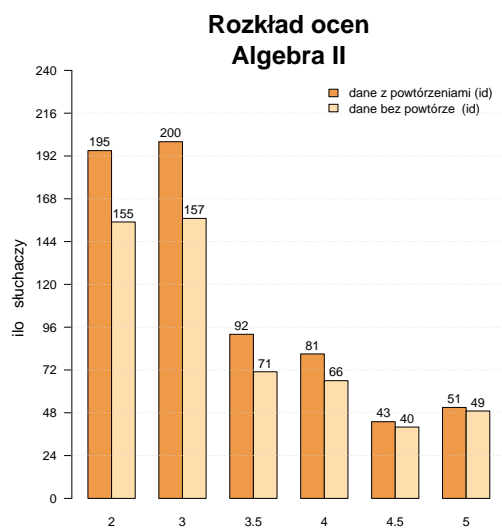
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.53	0.29
pk_0	0.66	0.86
wsp	-0.13	-0.58
$prop$	0.8	0.33

Kurs: Algebra II - alg2a

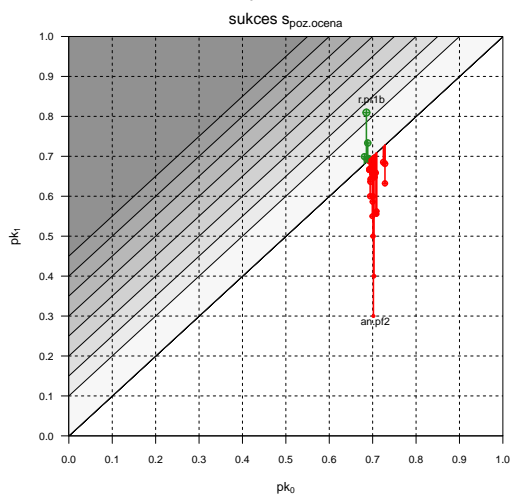
Ogólny rozkład ocen

Samowsparcie

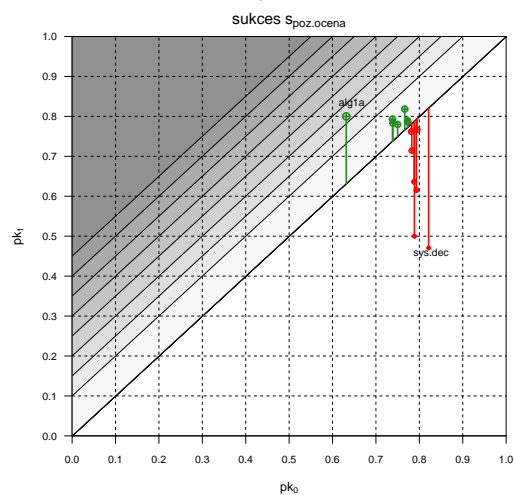


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$

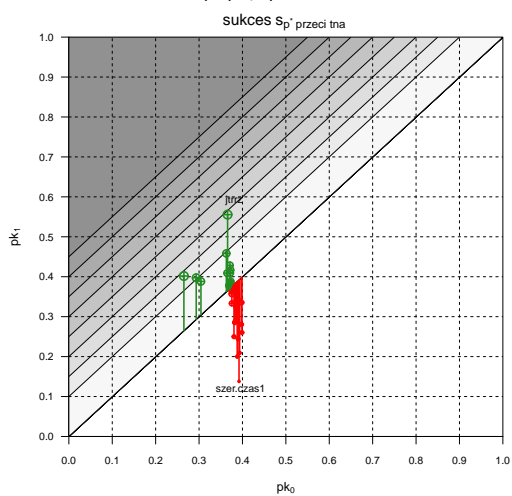


Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$

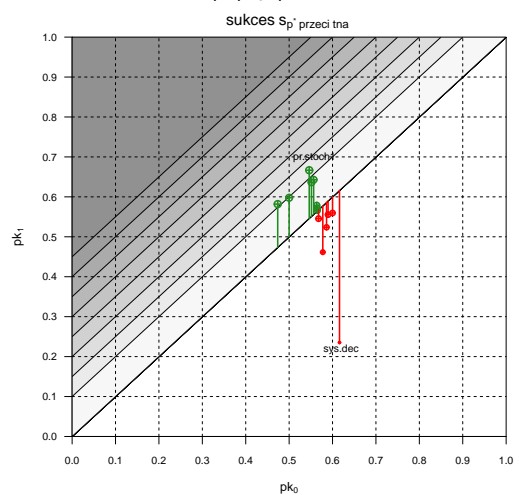


Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Tablica dwudzielna dla problemu samowsparcia

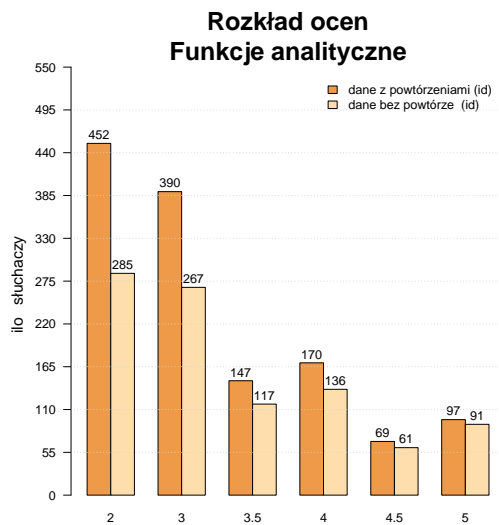
alg2a		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	28/3	17/3
cykl.dyd.1	283/100	126/29

uzyskane wyniki:

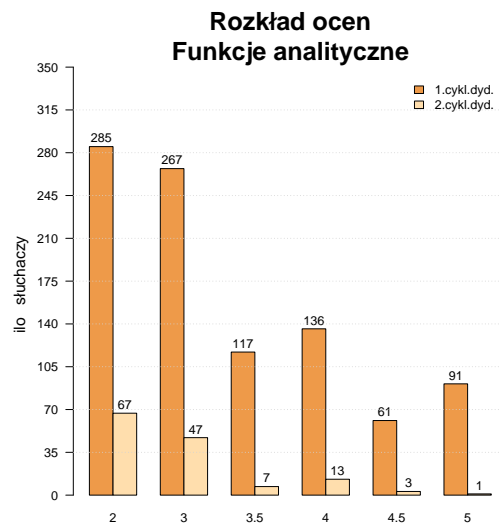
	st. matematyki	st. informatyki
pk_1	0.62	0.5
pk_0	0.69	0.78
wsp	-0.07	-0.28
$prop$	0.9	0.64

Kurs: Funkcje analityczne - fun.an

Ogólny rozkład ocen

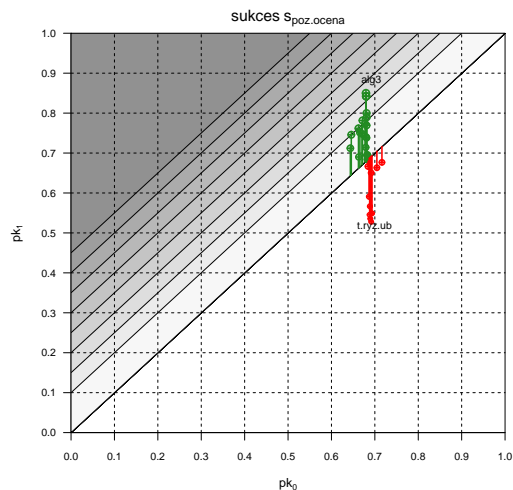


Samowsparcie

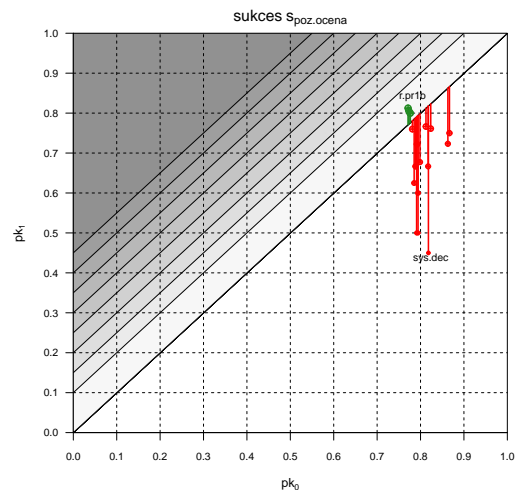


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

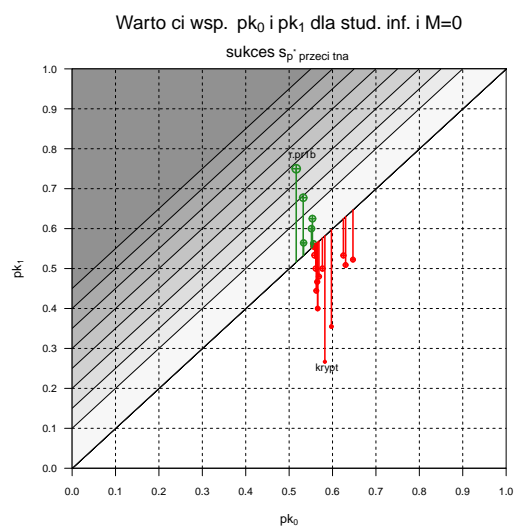
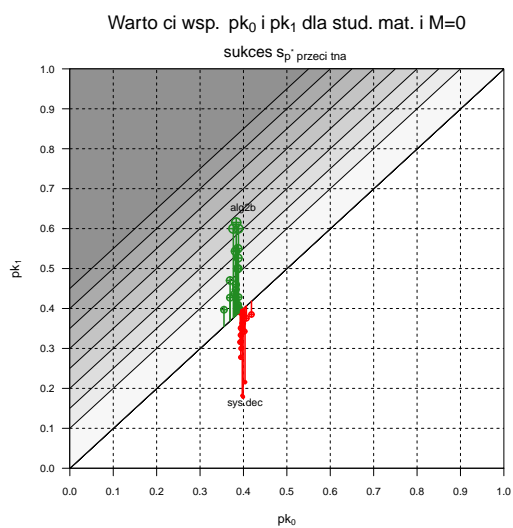
Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



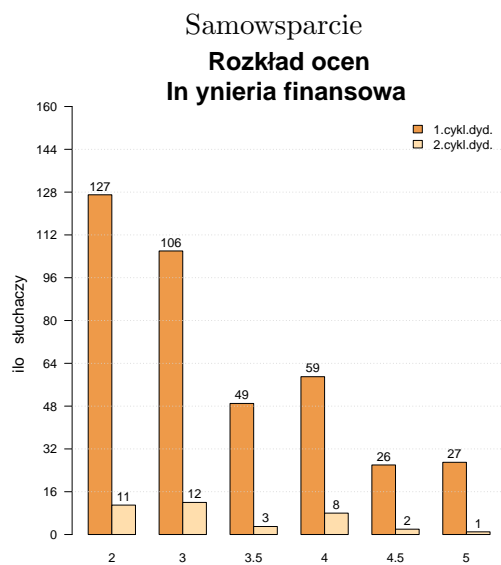
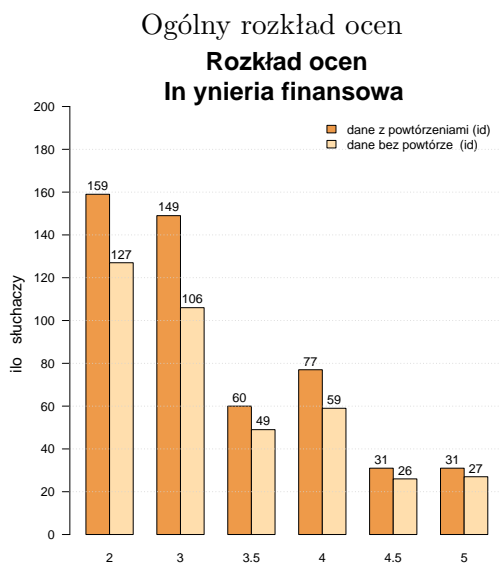
Tablica dwudzielna dla problemu samowsparcia

	fun.an	
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	65/6	61/6
cykl.dyd.1	528/144	244/41

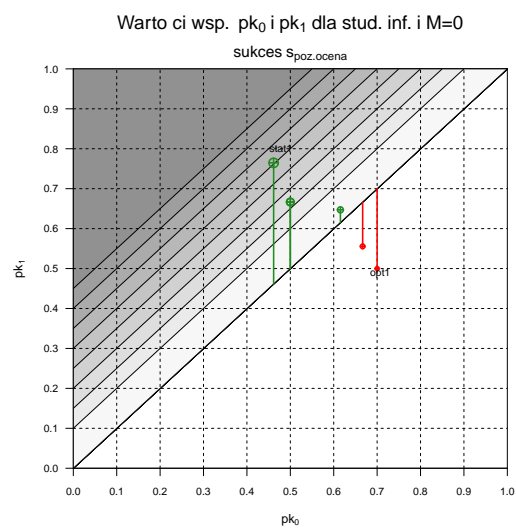
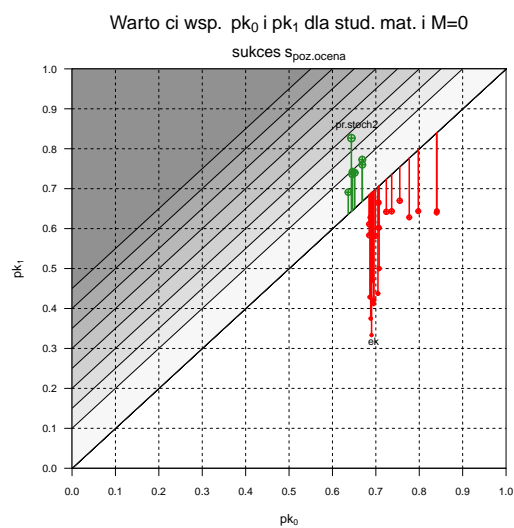
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.52	0.5
pk_0	0.68	0.78
wsp	-0.17	-0.28
$prop$	0.75	0.64

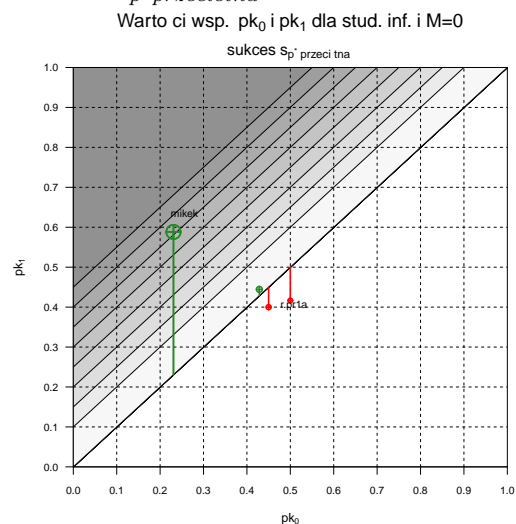
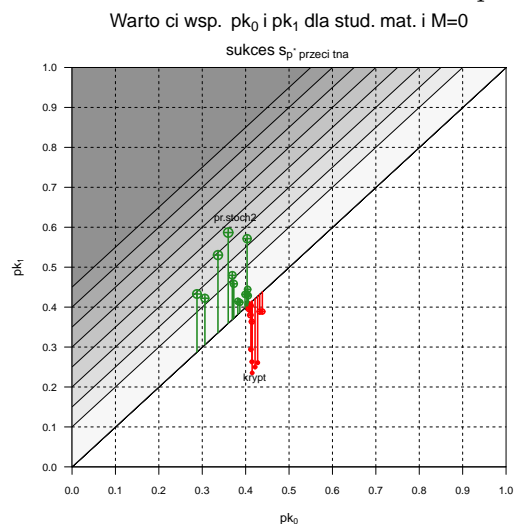
Kurs: Inżynieria finansowa - inz.fin



Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



Tablica dwudzielna dla problemu samowsparcia

	inz.fin	
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	24/2	11/0
cykl.dyd.1	248/19	116/11

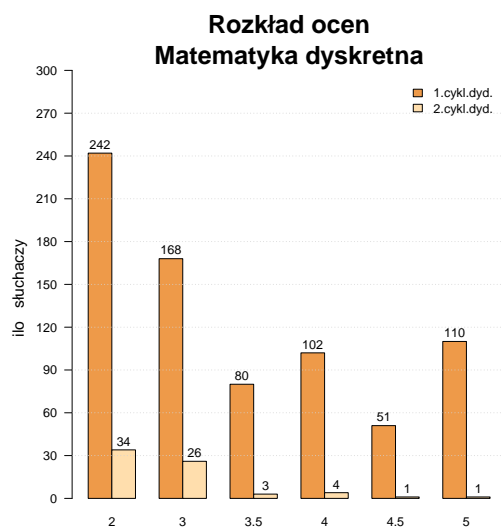
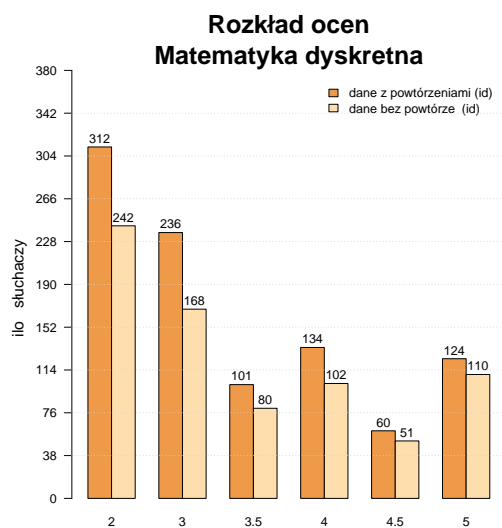
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.69	1
pk_0	0.68	0.63
wsp	0	0.37
$prop$	1.01	1.58

Kurs: Matematyka dyskretna - mat.dys

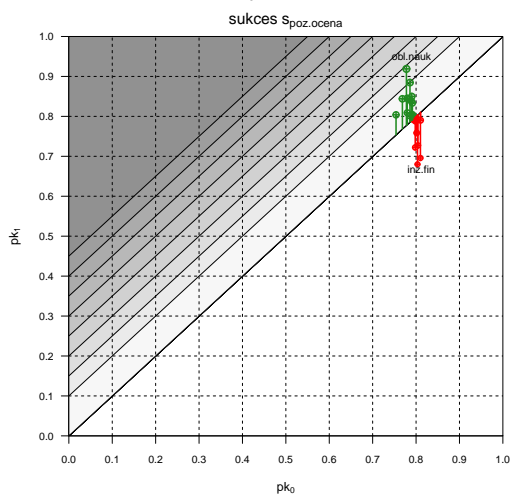
Ogólny rozkład ocen

Samowsparcie

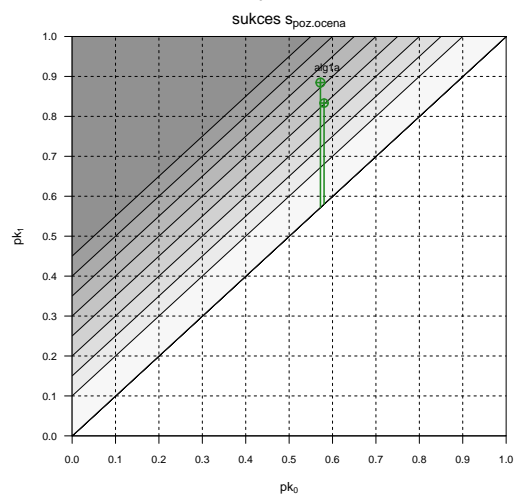


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$

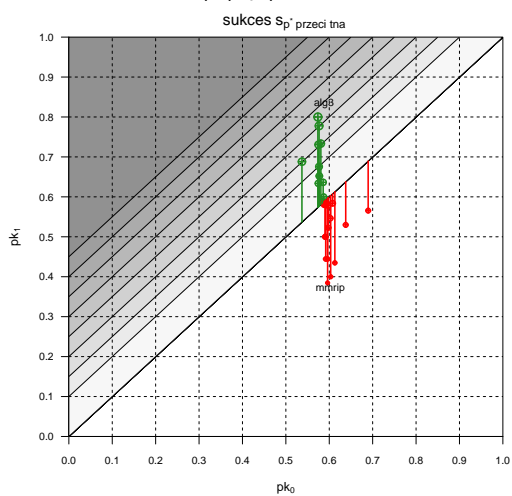


Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$

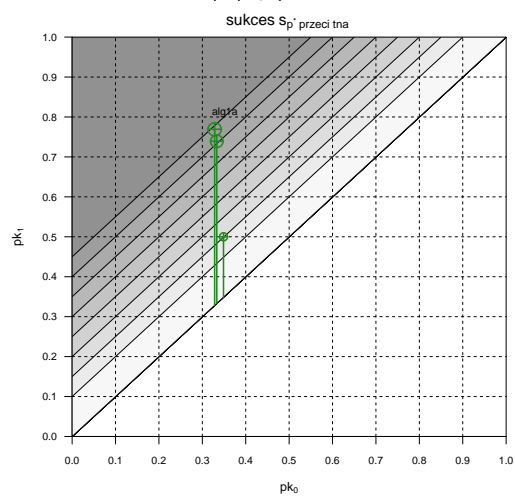


Podsumowanie wsparcia dla sukcesu $s_{p*przeciwna}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Tablica dwudzielna dla problemu samowsparcia

mat.dys		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	7/28	7/27
cykl.dyd.1	258/253	67/175

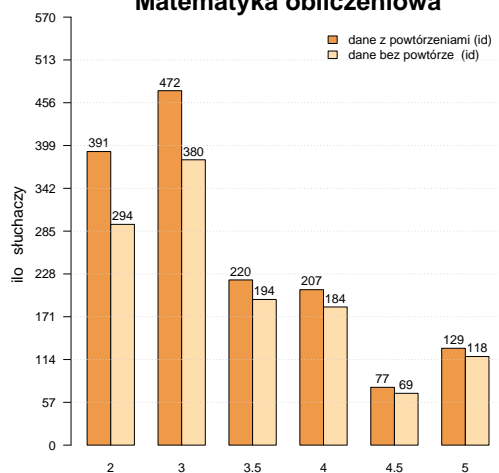
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.5	0.51
pk_0	0.79	0.59
wsp	-0.29	-0.08
$prop$	0.63	0.86

Kurs: Matematyka obliczeniowa - mat.obl1

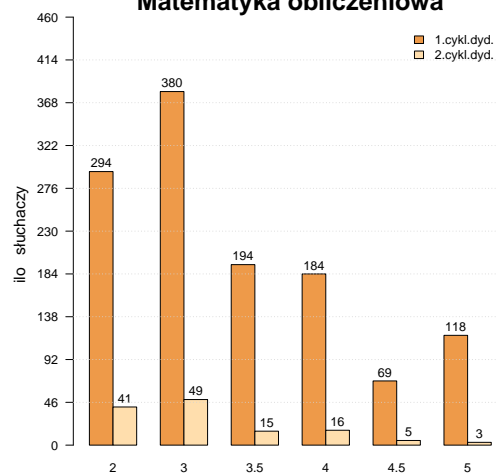
Ogólny rozkład ocen

Rozkład ocen
Matematyka obliczeniowa



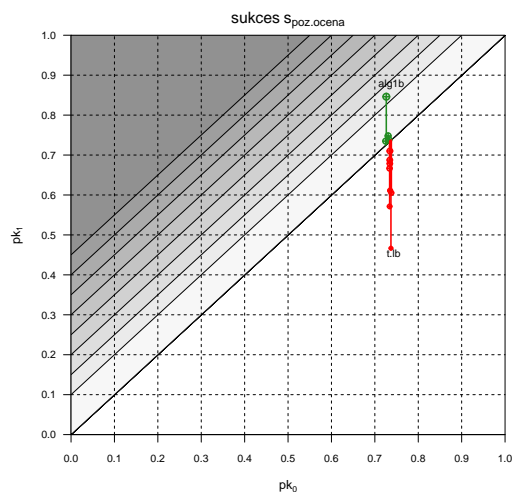
Samowsparcie

Rozkład ocen
Matematyka obliczeniowa

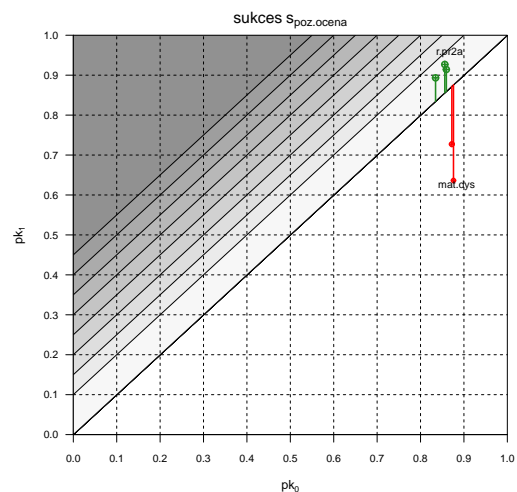


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

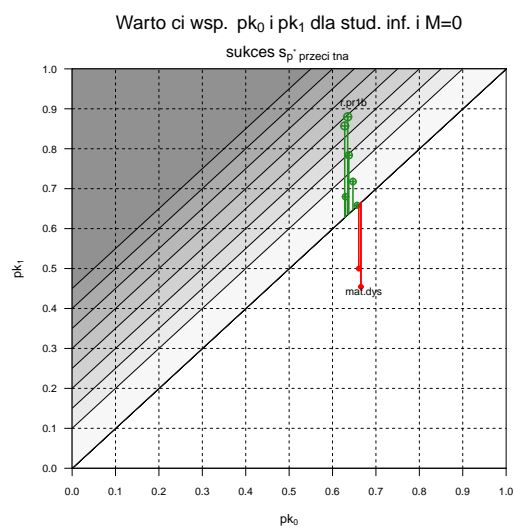
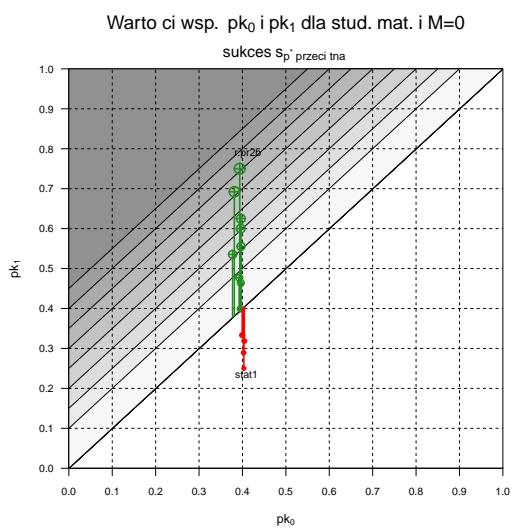
Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



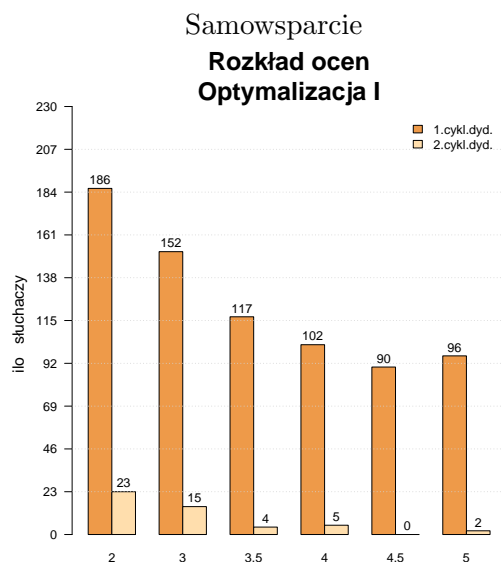
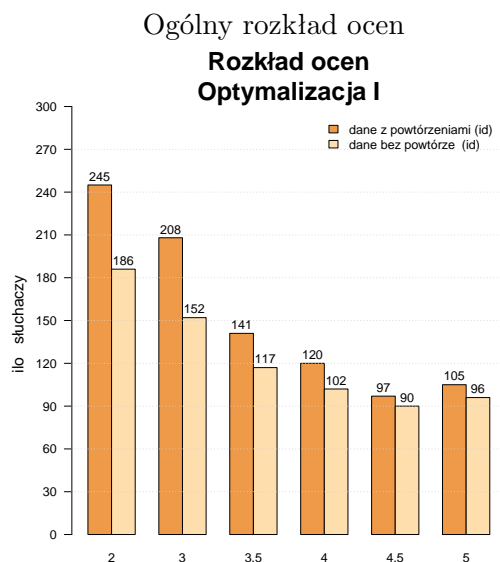
Tablica dwudzielna dla problemu samowsparcia

mat.obl1		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	83/5	39/2
cykl.dyd.1	705/240	257/37

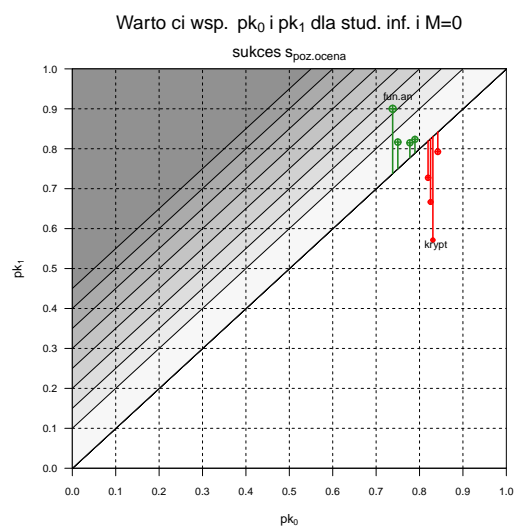
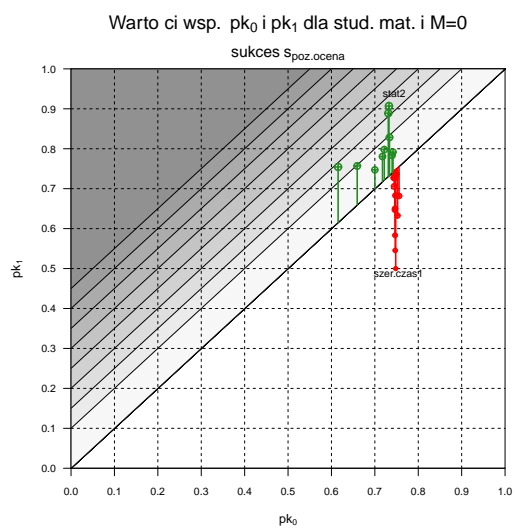
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.68	0.71
pk_0	0.73	0.87
wsp	-0.05	-0.15
$prop$	0.93	0.82

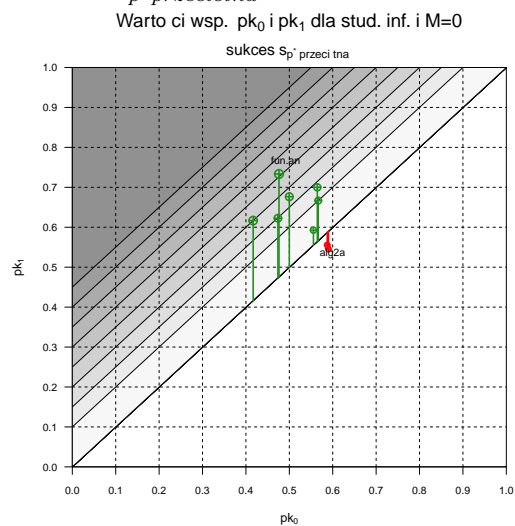
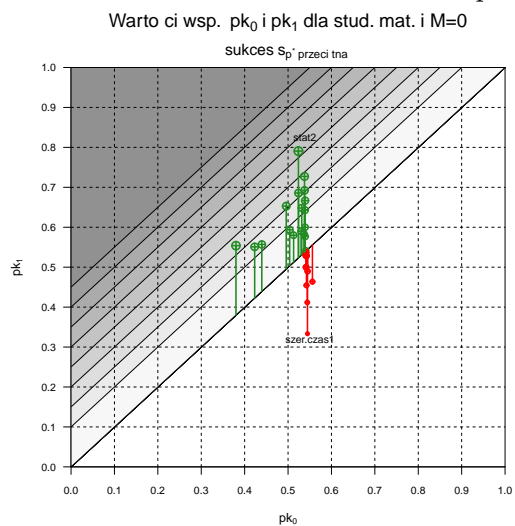
Kurs: Optymalizacja I - opt1



Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



Tablica dwudzielna dla problemu samowsparcia

opt1		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	25/1	23/0
cykl.dyd.1	499/58	172/14

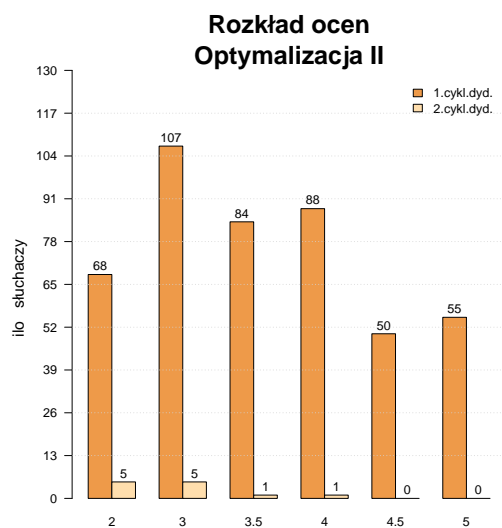
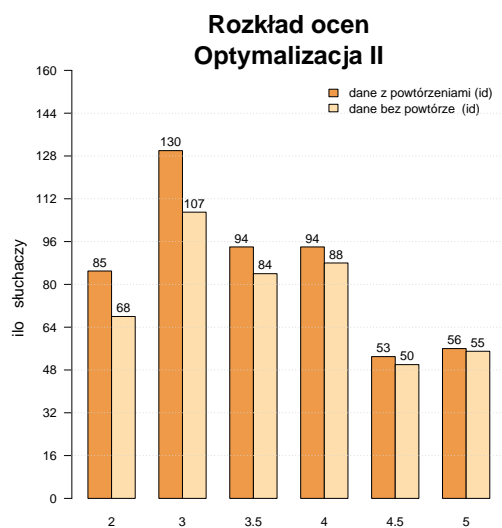
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.52	1
pk_0	0.74	0.81
wsp	-0.22	0.19
$prop$	0.7	1.24

Kurs: Optymalizacja II - opt2

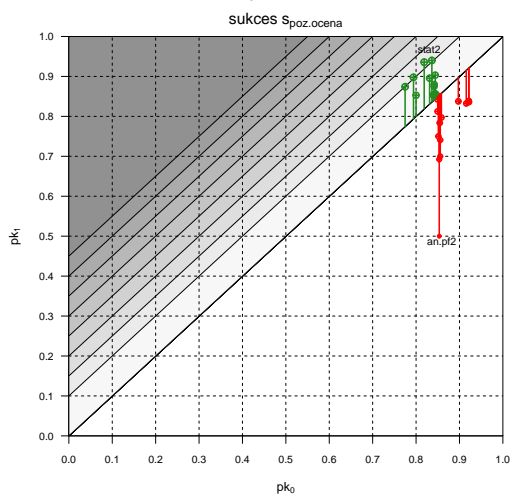
Ogólny rozkład ocen

Samowsparcie

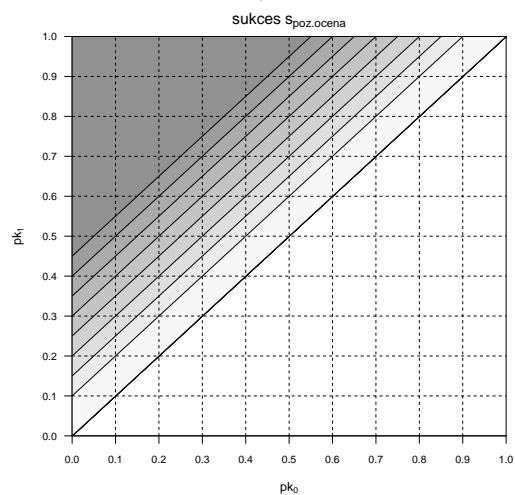


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$

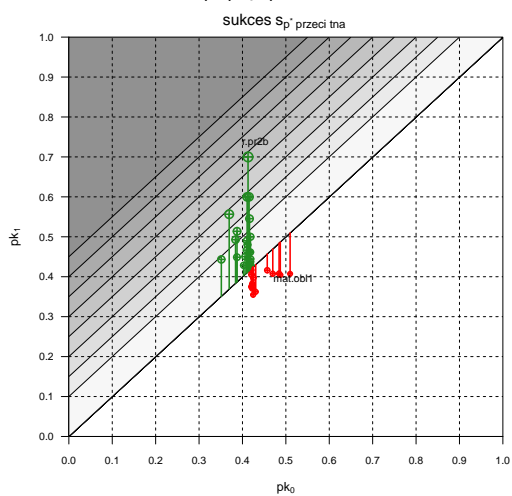


Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$

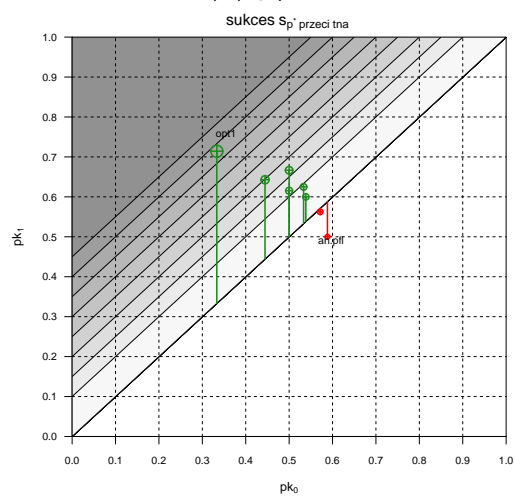


Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Tablica dwudzielna dla problemu samowsparcia

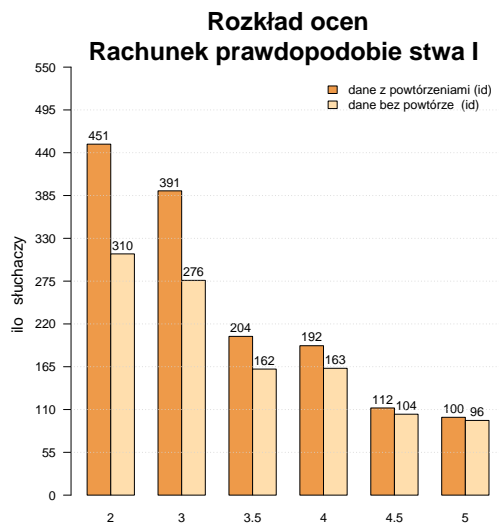
	opt2	
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	5/2	5/0
cykl.dyd.1	364/20	65/3

uzyskane wyniki:

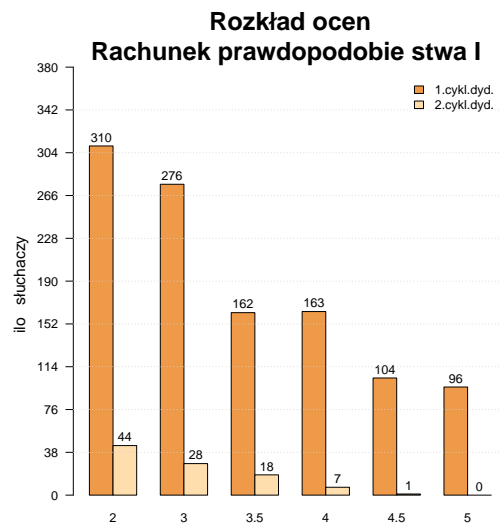
	st. matematyki	st. informatyki
pk_1	0.5	1
pk_0	0.85	0.87
wsp	-0.35	0.13
$prop$	0.59	1.15

Kurs: Rachunek prawdopodobieństwa I - r.pr1a

Ogólny rozkład ocen

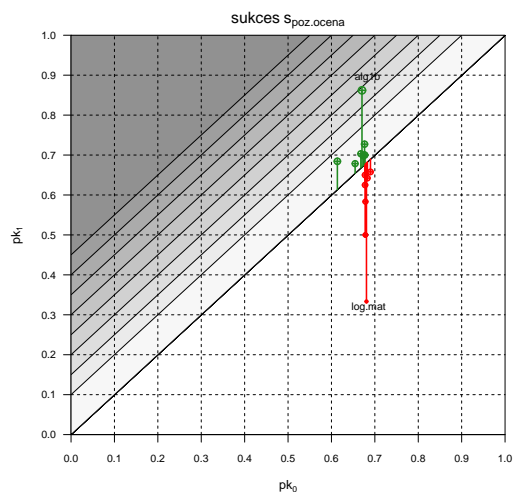


Samowsparcie

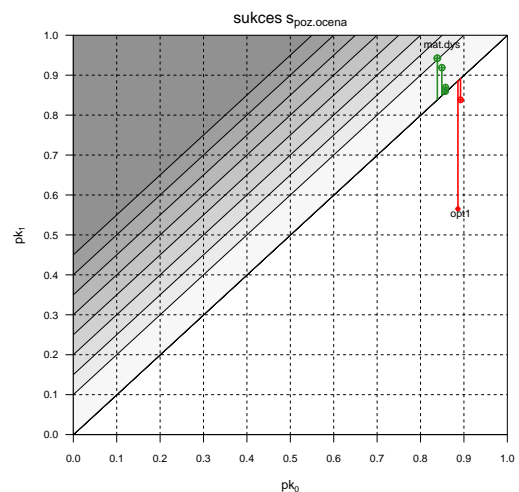


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

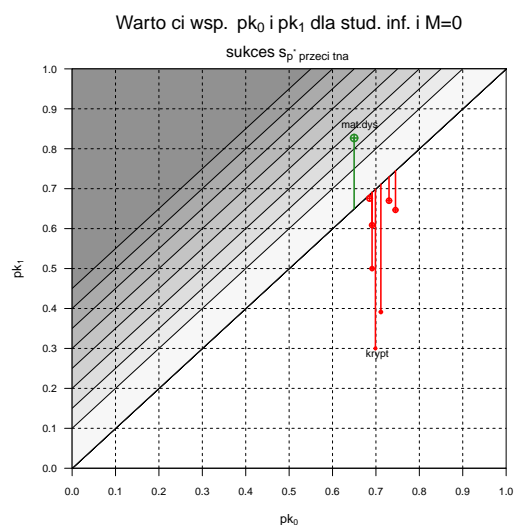
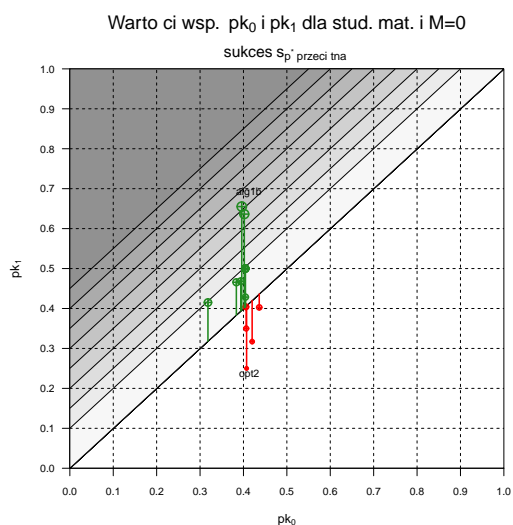
Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



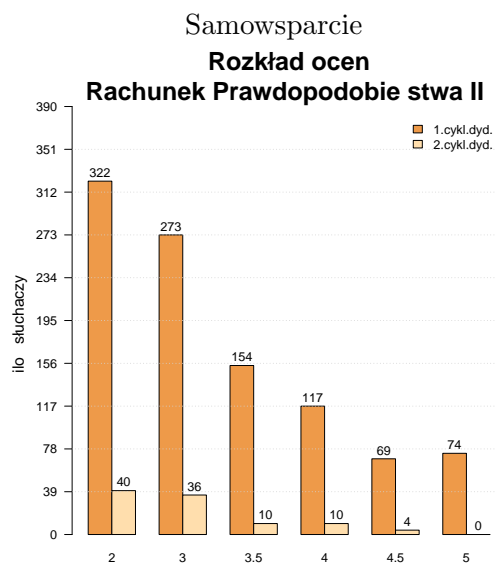
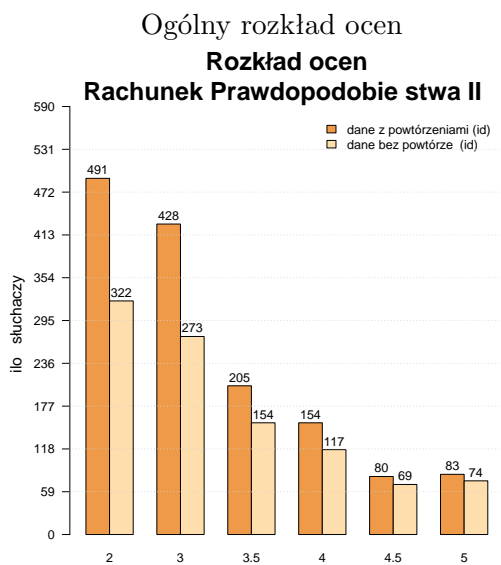
Tablica dwudzielna dla problemu samowsparcia

r.pr1a		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	48/6	37/7
cykl.dyd.1	570/231	272/38

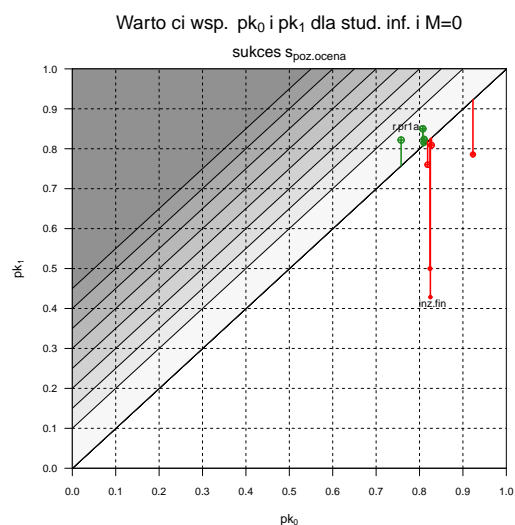
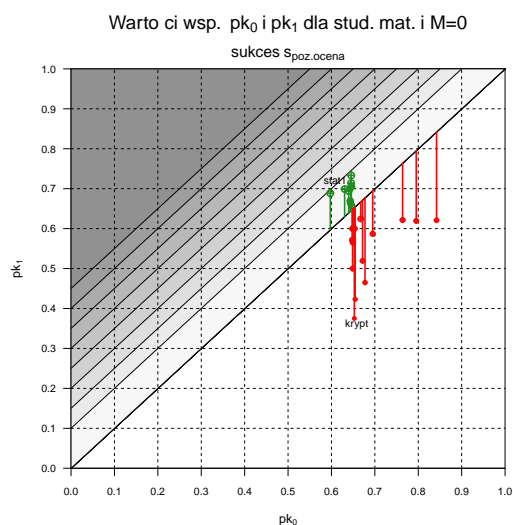
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.56	0.46
pk_0	0.68	0.86
wsp	-0.11	-0.4
prop	0.83	0.54

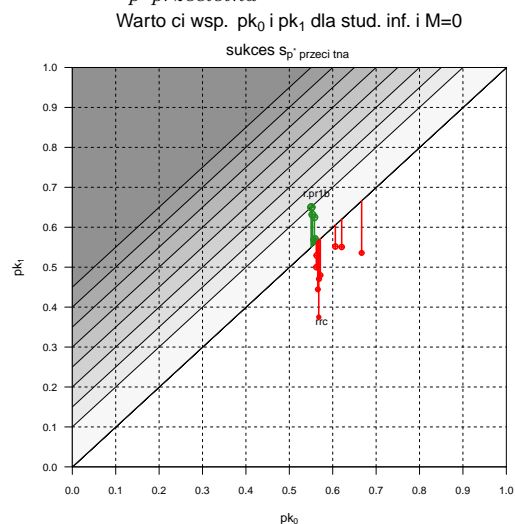
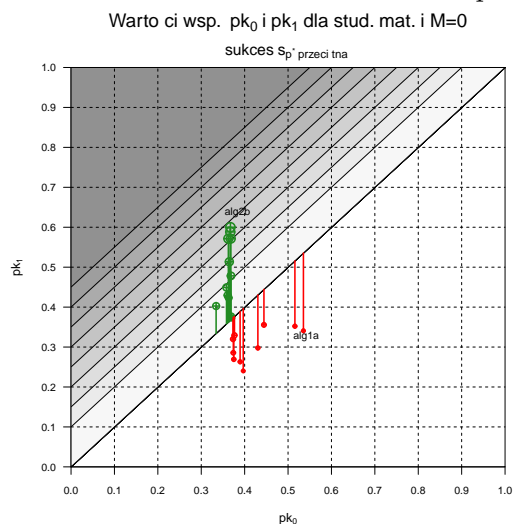
Kurs: Rachunek Prawdopodobieństwa II - r.pr2a



Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



Tablica dwudzielna dla problemu samowsparcia

r.pr2a		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	55/5	35/5
cykl.dyd.1	519/168	283/39

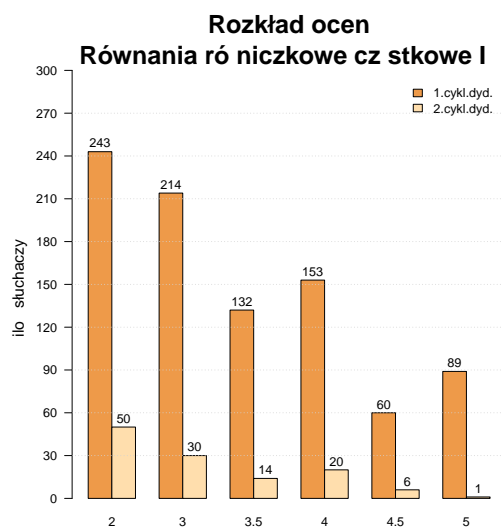
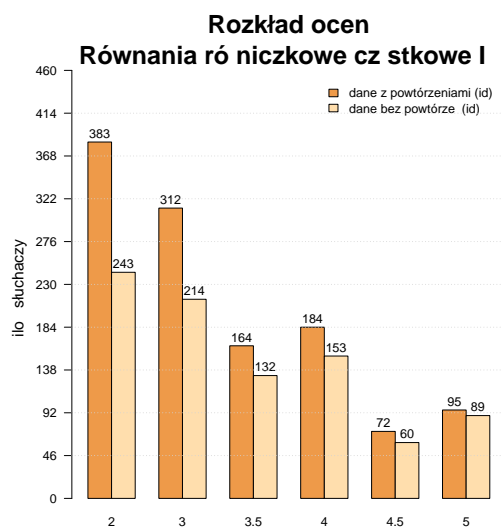
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.61	0.5
pk_0	0.65	0.81
wsp	-0.04	-0.31
$prop$	0.94	0.62

Kurs: Równania różniczkowe cząstkowe I - rrc

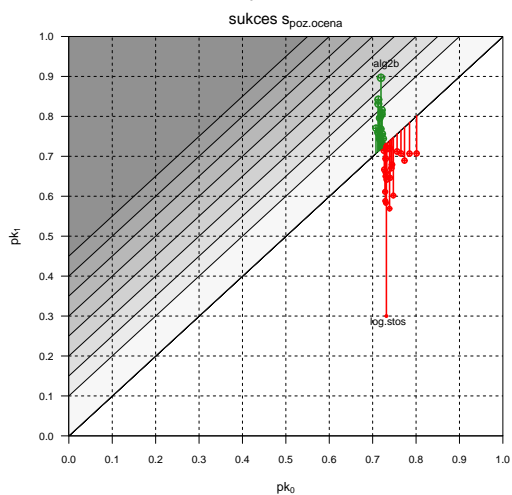
Ogólny rozkład ocen

Samowsparcie

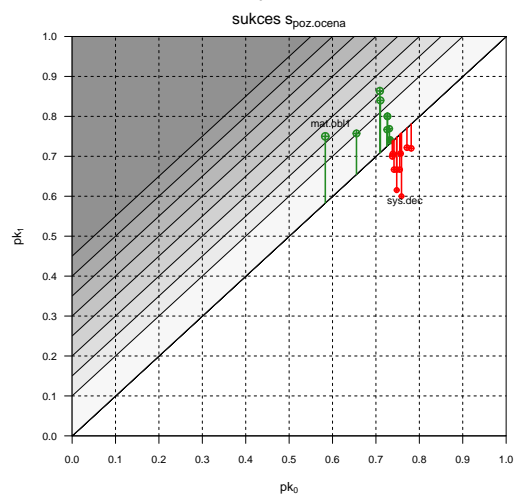


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

Wartości wsp. pk_0 i pk_1 dla stud. mat. i $M=0$

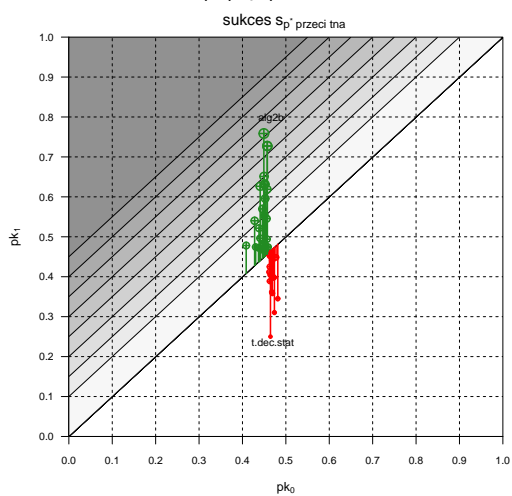


Wartości wsp. pk_0 i pk_1 dla stud. inf. i $M=0$

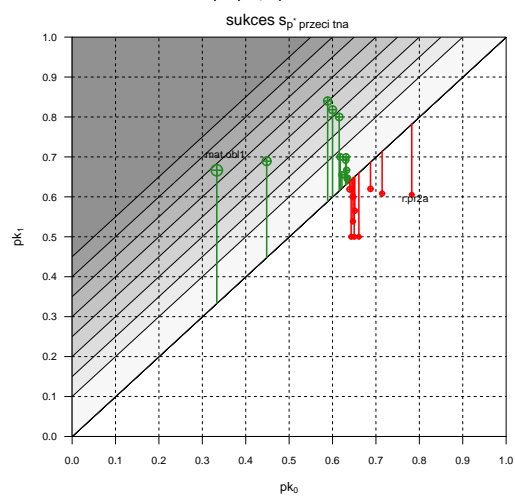


Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:

Wartości wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Wartości wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Tablica dwudzielna dla problemu samowsparcia

	rrc	
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	65/6	46/4
cykl.dyd.1	551/97	208/35

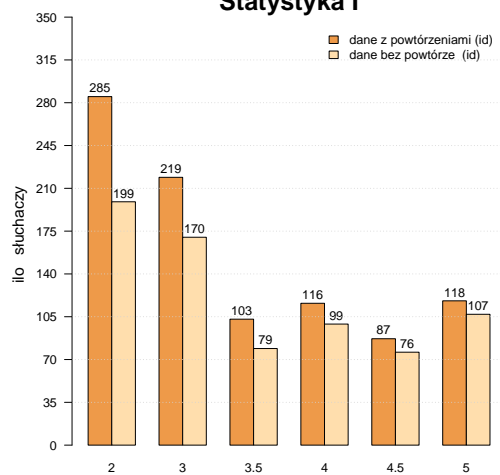
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.59	0.6
pk_0	0.73	0.73
wsp	-0.14	-0.13
$prop$	0.81	0.82

Kurs: Statystyka I - stat1

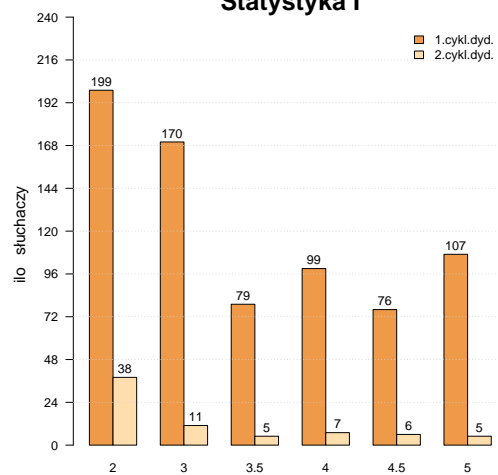
Ogólny rozkład ocen

**Rozkład ocen
Statystyka I**



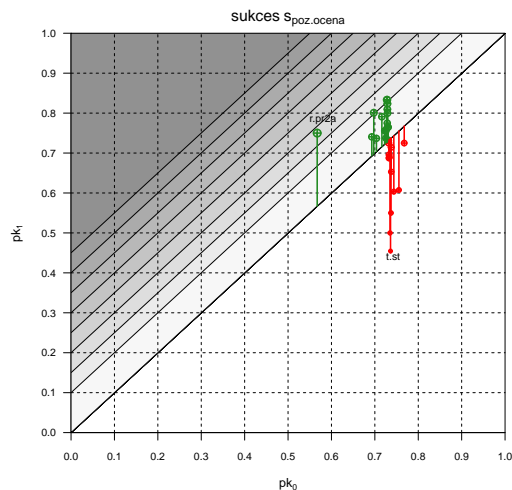
Samowsparcie

**Rozkład ocen
Statystyka I**

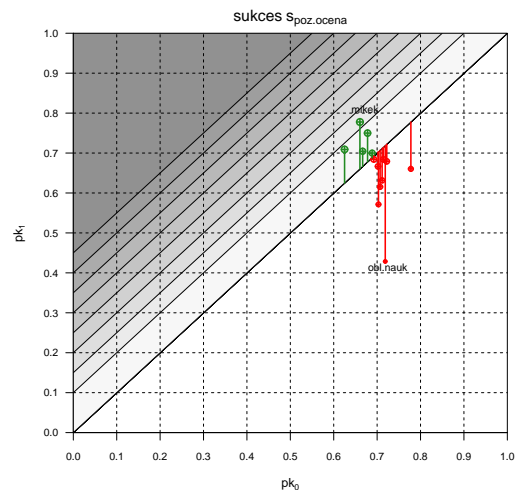


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

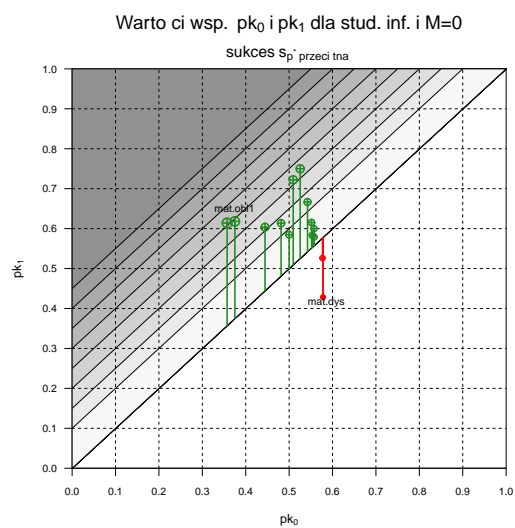
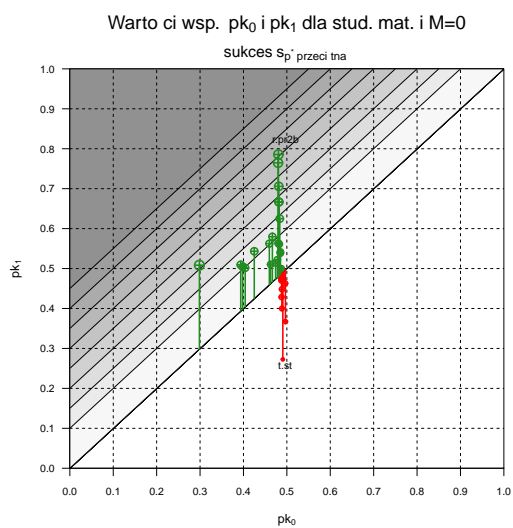
Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



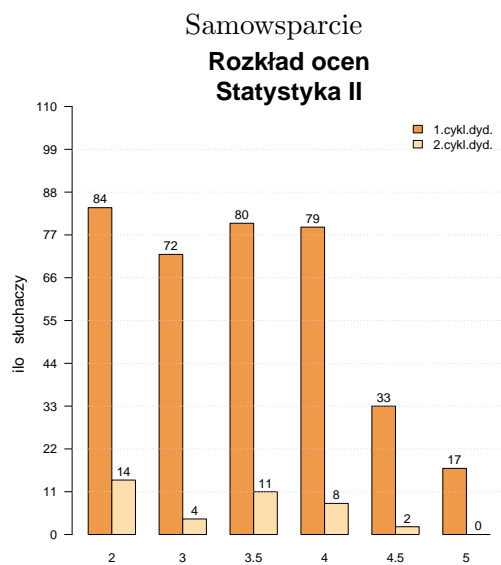
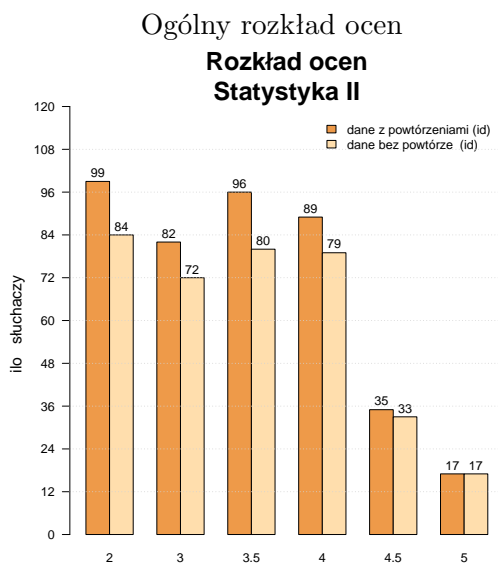
Tablica dwudzielna dla problemu samowsparcia

	stat1	
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	31/3	36/2
cykl.dyd.1	482/49	177/22

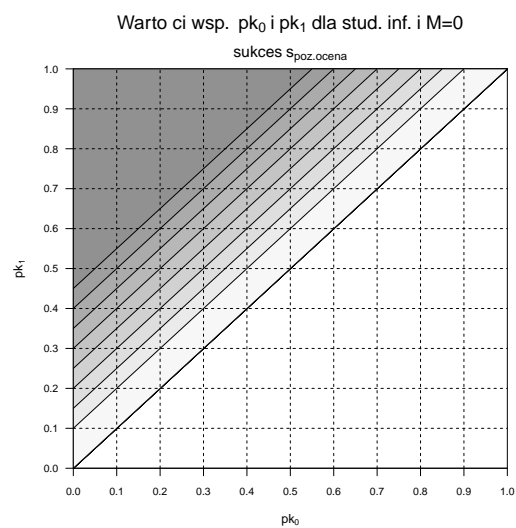
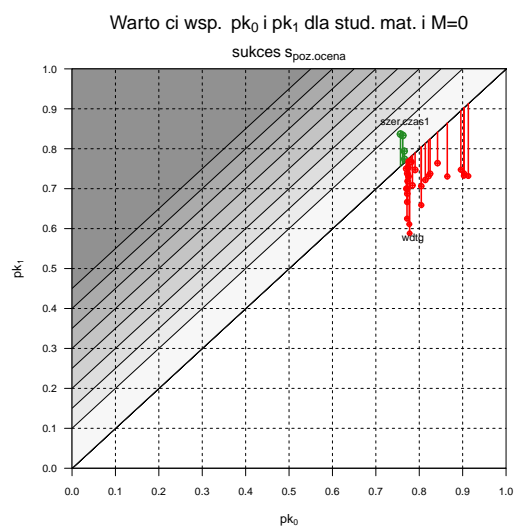
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.46	0.6
pk_0	0.73	0.69
wsp	-0.27	-0.09
$prop$	0.63	0.87

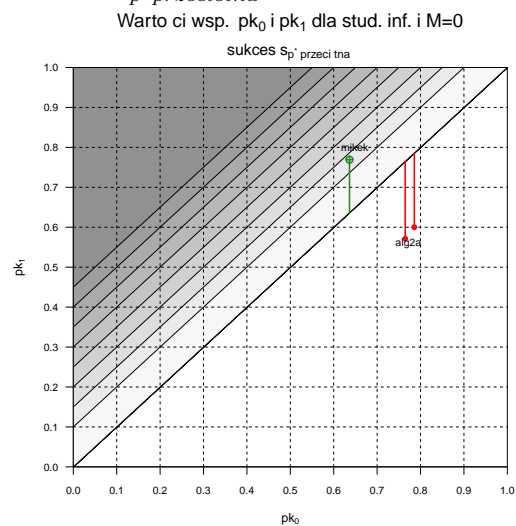
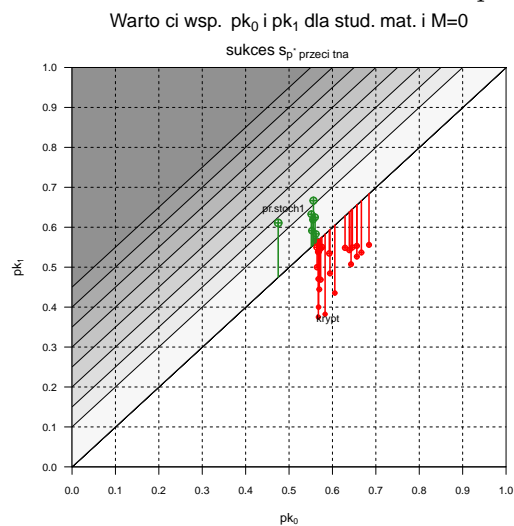
Kurs: Statystyka II - stat2



Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:



Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:



Tablica dwudzielna dla problemu samowsparcia

stat2		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	23/2	14/0
cykl.dyd.1	262/19	79/5

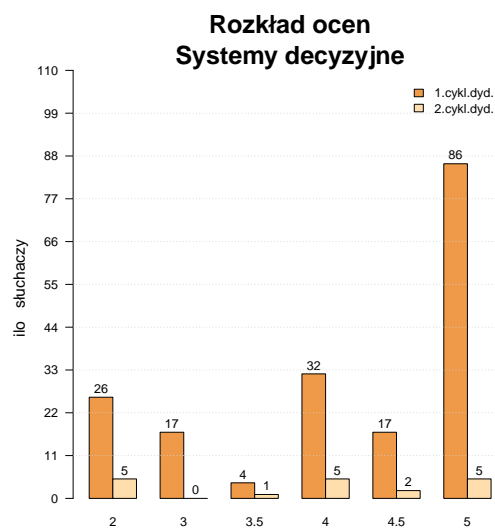
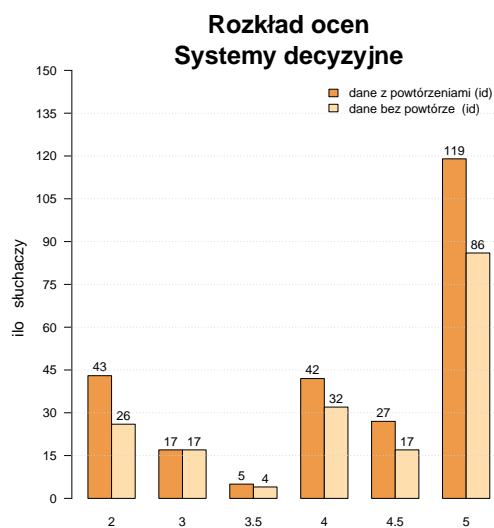
uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.62	1
pk_0	0.77	0.79
wsp	-0.15	0.21
$prop$	0.81	1.26

Kurs: Systemy decyzyjne - sys.dec

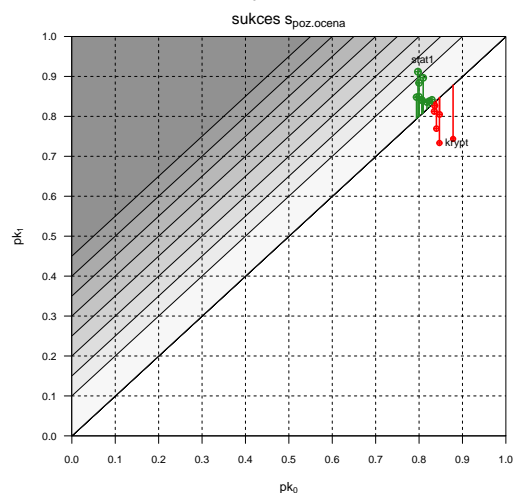
Ogólny rozkład ocen

Samowsparcie

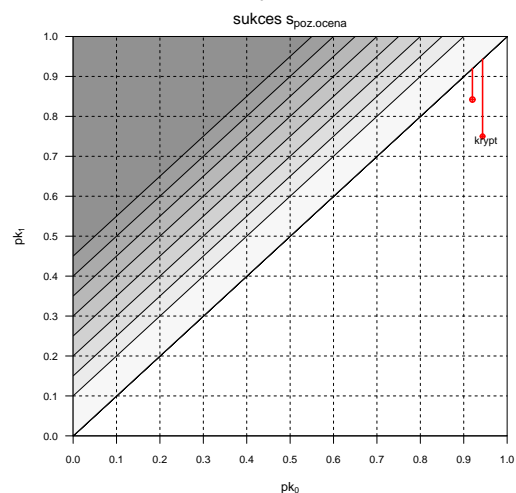


Podsumowanie wsparcia dla sukcesu $s_{poz.ocena}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$

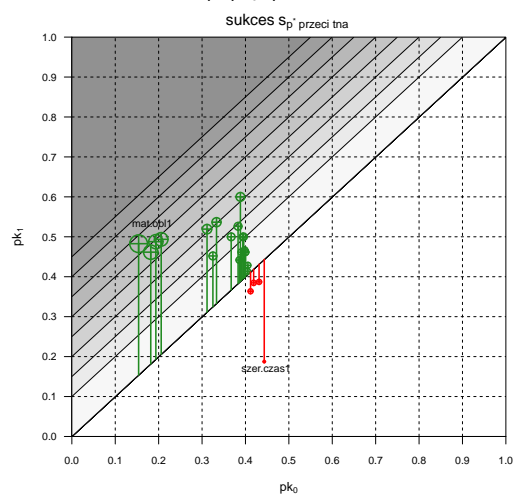


Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$

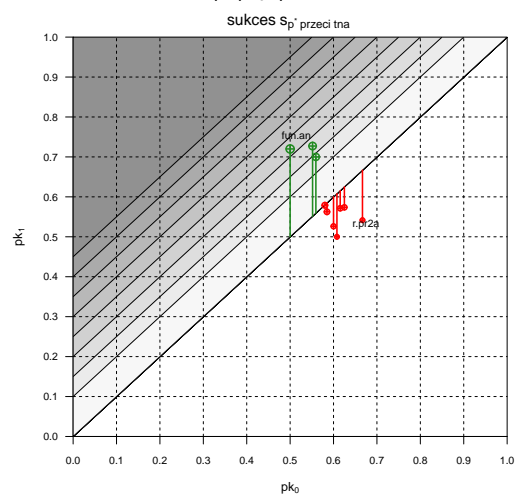


Podsumowanie wsparcia dla sukcesu $s_{p*przecietna}$:

Warto ci wsp. pk_0 i pk_1 dla stud. mat. i $M=0$



Warto ci wsp. pk_0 i pk_1 dla stud. inf. i $M=0$



Tablica dwudzielna dla problemu samowsparcia

sys.dec		
	$s_{poz.ocena} = 1$	$s_{poz.ocena} = 0$
cykl.dyd.2	10/3	4/1
cykl.dyd.1	94/62	19/7

uzyskane wyniki:

	st. matematyki	st. informatyki
pk_1	0.71	0.75
pk_0	0.83	0.9
wsp	-0.12	-0.15
$prop$	0.86	0.83

Dodatek C

Płyta CD

Do prezentowanej pracy magisterskiej dołączono płytę CD, która zawiera:

1. Niniejszą pracę w formacie pdf;
2. Skrypty programu R umożliwiające wykonanie przedstawionych analiz;
3. Pełną wersję Raportu *Szczegóły wyznaczania sieci (samo)wspierania* przedstawionego pokrótce w poprzednim załączniku.

Bibliografia

- [1] Biecek P., *Przewodnik po pakiecie R*, Oficyna Wydawnicza GiS, Wrocław, 2008, ISBN: 978-83-89020-79-6.
- [2] Diestel R., *Graph Theory*, Springer, New York, 1997, ISBN: 0-387-98210-8.
- [3] Grotkowski F., *USOS: System raportowania i analiz statystycznych*, Praca magisterska, wrzesień 2011.
- [4] Højsgaard S *Graphical Independence Networks with the gRain package for R*, Journal of Statistical Software, marzec 2012,
<http://www.jstatsoft.org/v46/i10/paper>.
- [5] Jakubowski J., Sztencel R., *Wstęp do teorii prawdopodobieństwa*, SCRIPT, Warszawa, 2004, ISBN: 83-89716-01-1.
- [6] Jensen F.V., *Bayesian Networks and Decision Graphs*, Springer, New York, 2007, ISBN: 978-0-387-68281-5.
- [7] Kański K., Praca magisterska - w przygotowaniu.
- [8] Kjærulff U.B., Madsen A.L, *Probabilistic Networks - An Introduction to Bayesian Networks and Influence Diagrams*,
<http://people.cs.aau.dk/~uk/papers/pgm-book-I-05.pdf>.
- [9] Koronacki J., Mileniczuk J., *Statystyka dla studentów kierunków technicznych i przyrodniczych*, WNT, Warszawa, 2009, ISBN: 978-83-204-3633-4.
- [10] Koski T., Noble J.M., *Bayesian Netorks. An Introduction*, 2009, ISBN: 978-0-470-74304-1.
- [11] Noble J.M., *Bayesian Netowrks Lecutre Notes*, notatki do kursu prowadzonego na Wydziale MIMUW dostępne pod adresem:
<http://www.mai.liu.se/~jonob/warsaw/lecturenotes.pdf>.
- [12] Podręcznik użytkownika programu Cytoscape:
<http://cytoscape.org/manual/Cytoscape2.6Manual.html>.
- [13] Podręcznik użytkownika programu GeNIe:
http://genie.sis.pitt.edu/wiki/GeNIe_Documentation.
- [14] Repozytorium CRAN:
<http://cran.r-project.org/>.

- [15] Scutari M. *Learning Bayesian Networks with the **bnlearn** R Package*, Jurnal of Statistical Software, lipiec 2010,
<http://www.jstatsoft.org/v35/i03/paper>.
- [16] Strona główna programu Cytoscape:
<http://www.cytoscape.org/>.
- [17] Strona główna programu GeNIe:
<http://genie.sis.pitt.edu/>.
- [18] Strona główna projektu R:
<http://www.r-project.org/>.
- [19] Strona główna projektu USOS:
<http://usos.edu.pl/>.
- [20] Strona główna Wydziału MIMUW:
<http://www.mimuw.edu.pl/>.