

**Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie**

Wydział Informatyki, Elektroniki i Telekomunikacji

KATEDRA INFORMATYKI



DOKUMENTACJA TECHNICZNA

TOMASZ KASPRZYK, DANIEL OGIELA, JAKUB STĘPAK

**SYSTEM OBLICZAJĄCY WYNIKI WYBORÓW DLA
UOGÓLNIENIA SYSTEMU K-BORDA**

PROMOTOR:

dr hab. inż. Piotr Faliszewski

Kraków 2016

Spis treści

1. Dziedzina problemu	5
1.1. Metoda obliczania wyników wyborów	5
1.1.1. Metoda Bordy	5
1.1.2. Metoda k-Borda	5
1.1.3. Uogólnienie - system ℓ_p -Borda	5
1.2. Format danych wejściowych	6
2. Architektura Django	7
3. Opis modułów	8
3.1. Moduł administracji kont	8
3.1.1. Opis ogólny	8
3.1.2. Komponenty programowe	8
3.2. Moduł zarządzania wyborami	8
3.2.1. Opis ogólny	8
3.2.2. Komponenty programowe	9
3.3. Moduł wizualizacji wyborów i wyników wyborów	10
3.3.1. Opis ogólny	10
3.3.2. Komponenty programowe	10
3.4. Moduł generacji i wczytywania wyborów	11
3.4.1. Opis ogólny	11
3.4.2. Komponenty programowe	11
3.5. Moduł obliczania wyników wyborów	11
3.5.1. Opis ogólny	11
3.5.2. Komponenty programowe	11
3.6. Moduł URL Resolver	12
3.6.1. Opis ogólny	12
3.6.2. Komponenty programowe	12
4. Współpraca modułów i innych komponentów	13
4.1. Diagram komunikacji	13
4.2. Ścieżki przejść	13

4.2.1.	Logowanie.....	13
4.2.2.	Wylogowanie.....	14
4.2.3.	Rejestracja.....	14
4.2.4.	Wyświetlenie listy wszystkich wyborów	15
4.2.5.	Stworzenie wyborów.....	15
4.2.6.	Usunięcie wyborów.....	16
4.2.7.	Wyświetlenie informacji szczegółowych dla danych wyborów.....	16
4.2.8.	Stworzenie nowego wyniku dla danych wyborów	17
4.2.9.	Wyświetlenie szczegółowych informacji o danym wyniku danych wyborów.....	17
4.2.10.	Wczytanie wyborów z pliku .soc	18
4.2.11.	Generacja wyborów z rozkładu normalnego.....	18
5.	Struktura obiektowa przechowywanych danych	20
5.1.	Opis modeli wykorzystywanych w ORM.....	20
5.1.1.	Election	20
5.1.2.	Candidate	21
5.1.3.	Voter	21
5.1.4.	Preference.....	22
5.1.5.	Point	22
5.1.6.	Result	23
5.1.7.	GeneticAlgorithmSettings.....	23
5.2.	Diagram ERD	24
6.	Algorytmy	25
6.1.	Algorytm zachłanny	25
6.1.1.	Idea algorytmu	25
6.1.2.	Funkcja zadowolenia.....	25
6.1.3.	Szczegóły implementacji	26
6.1.4.	Pseudokod	27
6.2.	Algorytm zachłanny według zasady Chamberlina-Couranta	27
6.2.1.	Idea algorytmu	28
6.2.2.	Funkcja zadowolenia.....	28
6.2.3.	Szczegóły implementacyjne.....	28
6.3.	Algorytm genetyczny	29
6.3.1.	Osobnik	29
6.3.2.	Krzyżowanie	29
6.3.3.	Mutacja.....	29
6.3.4.	Parametry algorytmu.....	29

6.3.5. Cykl algorytmu	30
6.3.6. Wpływ parametrów działania na wyniki osiągnięte przez algorytm	30
7. Porównanie algorytmów	34
7.1. Zrównanie jakości działania algorytmów zachłannych.....	34
7.1.1. Działanie algorytmów zachłannych	34
7.1.2. Wpływ liczby wyborców na wartość graniczną parametru p	34
7.1.3. Wpływ odchylenia standardowego na wartość graniczną parametru p	36
7.1.4. Wpływ liczby kandydatów na wartość graniczną parametru p	36
7.1.5. Wpływ rozmiaru komitetu na wartość graniczną parametru p	37
7.1.6. Stopień pokrywania się zwycięzców dla różnych wartości parametru p	38

1. Dziedzina problemu

1.1. Metoda obliczania wyników wyborów

1.1.1. Metoda Bordy

Niech v będzie głosem nad zbiorem kandydatów C . Wynik według Bordy kandydata c w v jest równy

$$\beta(i) = ||C|| - i, \quad (1.1)$$

gdzie i - pozycja kandydata w ciągu v .

Wynik c w wyborach jest sumą wyników c u każdego z wyborców

1.1.2. Metoda k-Borda

Rozszerzenie metody Bordy. Wynik, zamiast dla jednego kandydata, obliczany jest dla ciągu kandydatów. f_{kB} - funkcja zadowolenia z komitetu, (i_1, \dots, i_k) - ciąg pozycji kandydatów

$$f_{kB} = \sum_{j=1}^k \beta(i_j) \quad (1.2)$$

Przykład $C = \{c_1, c_2, c_3, c_4\}$ - zbiór kandydatów, $v = (c_2, c_1, c_4, c_3)$ - głos Niech $k = 2$ (wybory 2 spośród 4) $w = (c_4, c_3)$. Najpierw określamy pozycje kandydatów z komitetu w w v : $pos_v(w) = (3, 4)$, zatem wynik komitetu w dla głosu v wynosi $f_{kB}(3, 4) = (3) + (4) = ||C|| - 3 + (||C|| - 4) = 1 + 0 = 1$

1.1.3. Uogólnienie - system ℓ_p -Borda

Zanim wprowadzone zostanie pojęcie uogólnionego systemu $k - Borda$ warto przypomnieć wzór na normę ℓ_p

Norma ℓ_p

$$\ell_p(x_1, x_2, \dots, x_n) = \sqrt[p]{x_1^p + x_2^p + \dots + x_n^p} \quad (1.3)$$

Wówczas, w uogólnionej wersji metody $k - Borda$, funkcja zadowolenia f_{kB} (1.2) zostaje uzależniona również od parametru p ze wzoru (1.3). Norma liczona jest z wyników według Bordy, $\beta(i)$ (1.1).

Wzór uogólniony funkcji zadowolenia przyjmuje zatem postać:

$$f_{\ell_p B}(p, (i_1, \dots, i_k)) = \sqrt[p]{[(i_1)]^p + [(i_2)]^p + \dots + [(i_k)]^p} \quad (1.4)$$

Systemy $k - Borda$ i Chamberlina-Couranta są szczególnymi przypadkami zdefiniowanego powyżej systemu $\ell_p - Borda$:

Dla $p = 1$, l_1 :

$$f_{\ell_p B}(1, (i_1, \dots, i_k)) = \beta(i_1) + \beta(i_2) + \dots + \beta(i_k) = f_{kB}(i_1, \dots, i_k) \quad (1.5)$$

Dla $p = \infty$, $l_\infty = \max$:

$$\begin{aligned} f_{\ell_p B}(\infty, (i_1, \dots, i_k)) \\ &= \lim_{p \rightarrow \infty} \sqrt[p]{\beta[(i_1)]^p + \beta[(i_2)]^p + \dots + [\beta(i_k)]^p} \quad (1.6) \\ &= \max \beta(i_1), \beta(i_2), \dots, \beta(i_k) = \beta(i_1) = f_{CC} \end{aligned}$$

1.2. Format danych wejściowych

Struktura danych wejściowych została zaczerpnięta z zestawów danych zamieszczonych na stronie PrefLib¹. Wykorzystywanym typem zestawów danych są zestawy SOC^{2 3}.

Zgodnie ze specyfikacją, każdy prawidłowy plik z rozszerzeniem *.soc* posiada profil będący całkowitą, przejściową i asymetryczną relacją na grupie obiektów.

Preferencje definiowane są jako ciągi liter oddzielone przecinkami, przykładowo:

A, B, C;

Powyższy zapis oznacza, że, spośród trzech opcji A, B, C, A jest bardziej preferowana niż B i C oraz B jest bardziej preferowana niż C. Relacja silnego porządku jest wówczas określana przez przecinek.

Format pliku *.soc*:

```
<liczba kandydatów>
1, <nazwa kandydata>
2, <nazwa kandydata>
...
<liczba kandydatów>, <nazwa kandydata>
<liczba głosujących>, <liczba głosów policzonych>, <liczba unikalnych
głosów>
<liczba powtórzeń głosu>, <głos>
...
<liczba powtórzeń głosu>, <głos>
```

Przy czym <głos> to pojedyncza *preferencja* opisana powyżej.

¹<http://www.preflib.org/data/format.php#soc>

²SOC - Strict Orders - Complete List

³Szczegółowe opisy wyborów odnoszących się do poszczególnych plików można znaleźć pod adresem <http://www.preflib.org/data/index.php#ed>

2. Architektura Django

Django to framework webowy napisany w *Pythonie*. Dostarcza wysokopoziomowych abstrakcji pozwalających na szybkie i wygodne pisanie przejrzystych aplikacji.

Jego architektura koncepcyjnie przypomina wzorzec architektoniczny *Model-View-Controller*, jednak, jak przyznają sami twórcy ¹, *Django* nie do końca wpasowuje się w klasyczne ujęcie *MVC*.

Django dostarcza mapowania obiektowo-relacyjnego, dzięki którym całość modeli można ująć w *Pythonie*. Wygodny *ORM* zazwyczaj wystarcza do obsługi bazy danych, jednak zawsze istnieje możliwość użycia bezpośrednio *SQL*.

Widoki w *Django* spełniają dwójaką funkcję - służą zarówno przekazaniu danych do wyświetlenia, jak i ich modyfikacji. W wyświetleniu danych użytkownikowi pośredniczą szablony (*templates*), które opakowują przekazane dane do postaci *HTML*'a, który może wyświetlić przeglądarka internetowa. Dzięki temu wybór danych, jakie mają zostać pokazane użytkownikowi, jest oddzielony od samego sposobu ich prezentacji.

Za kontroler z klasycznego *MVC* można uznać sam framework dostarczający wspomnianej obsługi bazy danych czy mapowania adresów *URL* do poszczególnych widoków.

¹<https://docs.djangoproject.com/en/1.10/faq/general/>

Django appears to be a MVC framework, but you call the Controller the “view”, and the View the “template”. How come you don’t use the standard names?

3. Opis modułów

3.1. Moduł administracji kont

3.1.1. Opis ogólny

Moduł odpowiedzialny za zarządzanie kontami użytkowników. Umożliwia czynności logowania, wylogowania oraz rejestracji. Nie współpracuje z innymi modułami. Korzysta jedynie z bazy danych w celu weryfikacji użytkowników.

3.1.2. Komponenty programowe

Komponenty programowe dla tego modułu znajdują się w pakiecie *ecs.accounts*

Typ komponentu	Komponenty	Wykorzystanie
widoki	klasa <code>LoginView</code>	logowanie
	klasa <code>RegisterView</code>	rejestracja
szablony	<code>login.html</code>	logowanie
	<code>register.html</code>	rejestracja
formularze	klasa <code>LoginForm</code>	logowanie
	klasa <code>RegistrationForm</code>	rejestracja

3.2. Moduł zarządzania wyborami

3.2.1. Opis ogólny

Moduł odpowiedzialny za administrację wyborami i ich wynikami. Umożliwia podstawowe operacje na wyborach i ich wynikach oraz nawigację między nimi. Pozwala na wyświetlenie listy wszystkich wyborów i ich wyników, stworzenie nowych wyborów lub wyniku, czy usunięcie wyborów. W celu wykonania niektórych zadań współpracuje z modułem obliczania wyników wyborów oraz modułem wizualizacji wyborów i ich wyników. Jest to najbardziej rozbudowany moduł.

3.2.2. Komponenty programowe

Typ komponentu	Komponenty	Wykorzystanie
widoki	klasa ElectionListView	wyświetlenie listy wszystkich wyborów
	klasa ElectionCreateView	stworzenie wyborów
	klasa ElectionDeleteView	usunięcie wyborów
	klasa ElectionDetailView	wyświetlenie informacji szczegółowych o danych wyborach
	klasa ResultCreateView	stworzenie wyniku dla danych wyborów - wyświetlenie formularza określającego parametry wyniku
	klasa ResultDetailsView	wyświetlenie danego wyniku danych wyborów
	klasa ResultDeleteView	usuwanie pojedynczego wyniku wyborów
szablony	election_list.html	wyświetlenie listy wszystkich wyborów, usunięcie wyborów - strona wyświetlana po pomyślnym usunięciu wyborów
	election_create.html	stworzenie wyborów
	election_details.html	stworzenie wyborów - strona wyświetlana po pomyślnym stworzeniu wyborów, wyświetlenie informacji szczegółowych o danych wyborach
	election_delete.html	usunięcie wyborów
	result_create.html	stworzenie nowego wyniku wyborów - strona z formularzem
	result_details.html	wyświetlenie danego wyniku danych wyborów
	result_delete.html	potwierdzenie usunięcia rezultatu
formularze	klasa ElectionForm	stworzenie wyborów
	klasa ResultForm	stworzenie wyniku dla danych wyborów - formularz określający parametry wyniku

3.3. Moduł wizualizacji wyborów i wyników wyborów

3.3.1. Opis ogólny

Moduł odpowiedzialny za stworzenie wykresu 2D wizualizującego wybory i jego wyniki. Wizualizacja dotyczy tylko wyborów wygenerowanych z rozkładu normalnego.

Wyborcy i kandydaci są reprezentowani jako punkty na płaszczyźnie. Punkty reprezentujące wyborców i kandydatów mają na wykresie odmienne kolory. Na wykresie wyników wyborów punkty reprezentujące zwycięzców wyborów są powiększone.

Moduł współpracuje z modułem zarządzania wyborów, który zleca mu zadanie wizualizacji wyborów lub jego wyników. W celu wykonania zadania moduł wizualizacji wyborów i wyników wyborów pobiera dane z bazy danych.

3.3.2. Komponenty programowe

Typ komponentu	Komponenty	Wykorzystanie
widoki	klasa <code>ScatterChartMixin</code>	pobranie danych potrzebnych do wygenerowania wykresów (punkty, tytuł wykresu, serii danych), dziedziczy po klasie <code>View</code>
	klasa <code>ElectionChartView</code>	wizualizacja wyborów, klasa dziedziczy po klasie <code>ScatterChartMixin</code> , pobranie współrzędnych kandydatów i wyborców
	klasa <code>ResultChartView</code>	wizualizacja wyników wyborów, klasa dziedziczy po klasie <code>ScatterChartMixin</code> , pobranie współrzędnych kandydatów, wyborców oraz zwycięzców
	klasa <code>AlgorithmsComparisonChartView</code>	zestawia dla danych wyborów czasy wykonania obliczeń przez różne algorytmy dla różnych wartości parametru p

3.4. Moduł generacji i wczytywania wyborów

3.4.1. Opis ogólny

Moduł odpowiedzialny za generację wyborów z rozkładu normalnego oraz wczytywanie wyborów z pliku formatu *.soc*. Moduł współpracuje z modułem zarządzania wyborami, któremu zleca po wykonaniu swoich zadań, stworzenie i wysłanie użytkownikowi odpowiedniej strony internetowej. Moduł zapewnia wygenerowanie wyborów z rozkładu normalnego według wskazanych parametrów oraz weryfikację danych przy wczytywaniu wyborów z pliku. Po stworzeniu wyborów moduł komunikuje się z bazą danych w celu utrwalenia wyborów.

3.4.2. Komponenty programowe

Typ komponentu	Komponenty	Wykorzystanie
Widoki	klasa <code>ElectionLoadDataFormView</code>	wczytanie wyborów z pliku
	klasa <code>ElectionGenerateDataFormView</code>	generacja wyborów z rozkładu normalnego
Szablony	<code>election_load_data.html</code>	strona z formularzem do wskazania pliku
	<code>election_generate_data.html</code>	strona z formularzem do wskazania parametrów wyborów i rozkładu normalnego
	<code>election_details.html</code>	strona wyświetlana po poprawnym wczytaniu danych z pliku
Formularze	klasa <code>ElectionLoadDataForm</code>	wczytanie z pliku

3.5. Moduł obliczania wyników wyborów

3.5.1. Opis ogólny

Moduł odpowiedzialny za obliczanie wyników wyborów. Zapewnia różne algorytmy do wykonania zadania. Użytkownik ma wybór między algorytmem genetycznym, dwoma algorytmami zachłannymi oraz algorytmem typu brute-force. Moduł współpracuje z modułem zarządzania wyborami, który zleca mu wykonanie zadania.

3.5.2. Komponenty programowe

Wszystkie komponenty programowe dotyczące modułu obliczania wyników wyborów zawierają się w pakiecie `ecs.elections.algorithms`.

Klasy odpowiedzialne za poszczególne algorytmy:

- `BruteForce` - odpowiedzialna za algorytm typu brute-force

- GreedyAlgorithm - odpowiedzialna za algorytm zachłanny zależny od parametru p
- GreedyCC - odpowiedzialna za algorytm zachłanny niezależny od parametru p
- GeneticAlgorithm - odpowiedzialna za algorytm genetyczny

3.6. Moduł URL Resolver

3.6.1. Opis ogólny

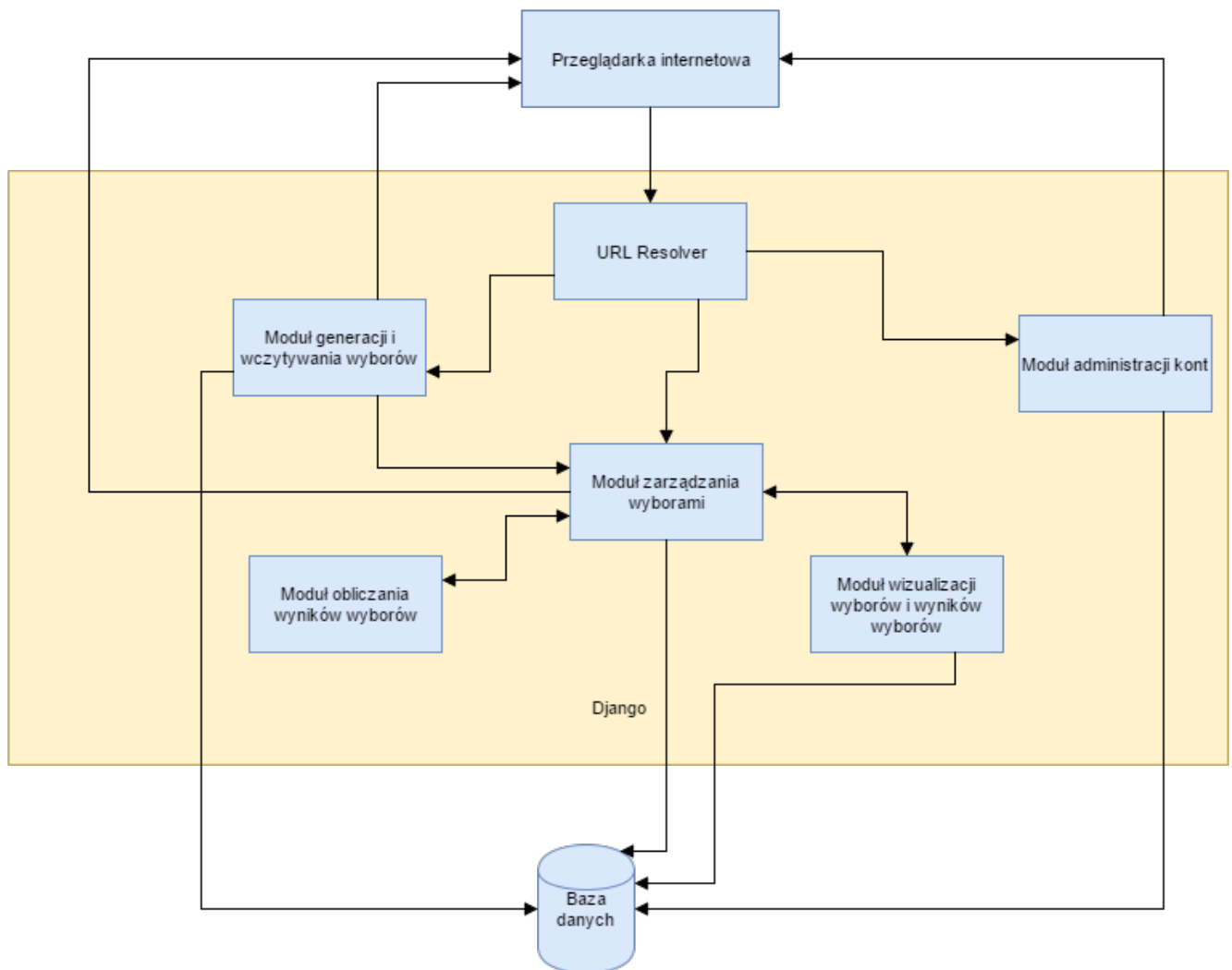
Moduł odpowiedzialny za przekazywanie żądań otrzymywanych przez klienta do odpowiednich modułów.

3.6.2. Komponenty programowe

Przyporządkowania żądań użytkownika do odpowiednich modułów (adresów URL do widoków) znajdują się w pliku `ecs.elections.urls.py`.

4. Współpraca modułów i innych komponentów

4.1. Diagram komunikacji



Rysunek 4.1: Diagram komunikacji

4.2. Ścieżki przejść

4.2.1. Logowanie

1. Użytkownik za pomocą przeglądarki internetowej wysyła żądanie logowania.
2. Moduł URL Resolver kieruje żądanie do modułu administracji kont.

3. Moduł administracji kont przesyła do przeglądarki stronę z formularzem do zalogowania (pole *Username* i *Password*).
4. Użytkownik wypełnia formularz i za pomocą przeglądarki internetowej wysyła uzupełniony formularz.
5. Moduł URL Resolver przekazuje żądanie do modułu administracji kont.
6. Moduł administracji kont przeprowadza uwierzytelnienie użytkownika.
7. W zależności od rezultatu uwierzytelnienia podjęte są trzy możliwe akcje:
 - jeżeli uwierzytelnienie powiodło się, Moduł administracji kont loguje użytkownika i wysyła do przeglądarki stronę z wiadomością o sukcesie logowania
 - jeżeli podana nazwa użytkownika nie istnieje w bazie, Moduł administracji kont wysyła do przeglądarki internetowej stronę z formularzem do zalogowania (pole *Username* i *Password*) i z informacją o nieistnieniu podanej nazwy użytkownika. Następuje przejście do kroku nr 4
 - jeżeli podana nazwa użytkownika istnieje w bazie, ale podane hasło jest nieprawidłowe, Moduł administracji kont wysyła do przeglądarki internetowej stronę z formularzem do zalogowania (pole *Username* i *Password*) i z informacją o błędnym hasle. Następuje przejście do kroku nr 4

4.2.2. Wylogowanie

1. Użytkownik za pomocą przeglądarki internetowej wysyła żądanie wylogowania.
2. Moduł URL Resolver kieruje żądanie do modułu administracji kont.
3. Moduł administracji kont wylogowuje użytkownika i przesyła do przeglądarki stronę domową aplikacji.

4.2.3. Rejestracja

1. Użytkownik za pomocą przeglądarki internetowej wysyła żądanie zarejestrowania użytkownika.
2. Moduł URL Resolver kieruje żądanie do modułu administracji kont.
3. Moduł administracji kont przesyła do przeglądarki stronę z formularzem do rejestracji (pola *E-mail*, *Username*, *Password*, *First name* i *Last name*).
4. Użytkownik wypełnia formularz i za pomocą przeglądarki internetowej wysyła uzupełniony formularz.
5. Moduł URL Resolver przekazuje żądanie do modułu administracji kont.
6. Moduł administracji kont przeprowadza walidację danych.

7. W zależności od rezultatu walidacji:

- jeżeli się powiodła, Moduł administracji kont tworzy konto nowego użytkownika, zapisuje je do bazy danych, loguje użytkownika do nowo utworzonego konta i przesyła do przeglądarki stronę z wiadomością o sukcesie logowania
- w przeciwnym wypadku, zostaje wysłana do przeglądarki strona z formularzem do rejestracji wraz z informacją o znalezionym problemie. Następuje przejście do kroku nr 4

4.2.4. Wyświetlenie listy wszystkich wyborów

1. Użytkownik klikając przycisk *Your elections* na stronie, wysyła za pomocą przeglądarki internetowej żądanie wyświetlenia listy swoich wyborów.
2. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ElectionListView*).
3. Moduł zarządzania wyborami kieruje zapytanie do bazy danych o obiekty reprezentujące wybory danego użytkownika.
4. Po pobraniu z bazy potrzebnych danych moduł zarządzania wyborami za pomocą szablonu `election_list.html` tworzy stronę i wysyła do przeglądarki internetowej użytkownika.

4.2.5. Stworzenie wyborów

1. Użytkownik klikając przycisk *New election* na stronie, wysyła za pomocą przeglądarki internetowej żądanie stworzenie nowych wyborów.
2. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ElectionCreateView*).
3. Widok posiada zdefiniowany szablon `election_create.html` i formularz *ElectionForm*. Wykorzystując te komponenty tworzy i wysyła do użytkownika stronę z formularzem.
4. Użytkownik wypełnia formularz (pola *Name* i *Committee size*) i wysyła uzupełniony formularz.
5. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ElectionCreateView*).
6. Widok wykorzystując klasę formularza pobiera dane z formularza, tworzy nowy obiekt wyborów i zapisuje go do bazy.
7. Widok wykorzystując szablon `election_details.html` tworzy stronę i wysyła ją do przeglądarki użytkownika.

4.2.6. Usunięcie wyborów

1. Użytkownik klikając ikonę kosza na stronie, wysyła za pomocą przeglądarki internetowej żądanie usunięcia wyborów.
2. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ElectionDeleteView*).
3. Widok posiada zdefiniowany szablon `election_delete.html`, który wykorzystuje do stworzenia i wysłania strony potwierdzającej akcję użytkownika.
4. Użytkownik potwierdza chęć usunięcia wyborów przez naciśnięcie przycisku *Yes, delete this*. (W przypadku wyboru przycisku *No, cancel* akcja nie zostaje przeprowadzona a użytkownik zostaje przeniesiony na stronę z listą swoich wyborów).
5. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ElectionDeleteView*).
6. Widok usuwa wybory z bazy danych i wykorzystując szablon `election_list.html` tworzy i wysyła stronę do przeglądarki internetowej użytkownika.

4.2.7. Wyświetlenie informacji szczegółowych dla danych wyborów

1. Użytkownik klikając na link wyborów na stronie z listą swoich wyborów, wysyła za pomocą przeglądarki internetowej żądanie wyświetlenia informacji szczegółowych dla danych wyborów.
2. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ElectionDetailView*).
3. Widok pobiera z bazy danych listę głosujących oraz listę wyników wyborów.
4. Jeżeli wybory zostały wygenerowane z rozkładu normalnego, widok przekierowuje działanie programu do modułu wizualizacji wyborów i wyników wyborów (widok *ElectionChartView*). Jeżeli wybory nie zostały wygenerowane z rozkładu normalnego, następuje przejście do kroku nr 6
5. Widok pobiera z bazy danych obiekty będące punktami 2D reprezentującymi wyborców i kandydatów. Ustawia parametry wykresu a następnie wysyła wszystkie dane z powrotem do widoku *ElectionDetailView* umożliwiając stworzenie wykresu.
6. Widok wykorzystując szablon `election_details.html` tworzy i wysyła stronę do przeglądarki użytkownika.

4.2.8. Stworzenie nowego wyniku dla danych wyborów

1. Użytkownik klikając na link *Add new result* na stronie ze szczegółowymi informacjami o danych wyborach, wysyła za pomocą przeglądarki internetowej żądanie stworzenia nowego wyniku dla danych wyborów.
2. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ResultCreateView*).
3. Widok wykorzystując formularz *ResultForm* i szablon *result_create.html* tworzy oraz wysyła do użytkownika stronę z formularzem do określenia parametrów wyniku (pola: parametr *p* do obliczania normy oraz typ algorytmu).
4. Użytkownik wypełnia formularz i wysyła go do systemu.
5. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ResultCreateView*).
6. Widok wczytuje dane z formularza, pobiera z bazy danych obiekt reprezentujący wybory a następnie przekierowuje działanie programu do modułu obliczania wyników wyborów przekazując konieczne parametry.
7. Moduł obliczania wyników wyborów liczy zwycięzców wyborów za pomocą odpowiedniego algorytmu. Następnie przekazuje sterowanie z powrotem do widoku *ResultCreateView* przekazując wyniki.
8. Widok zapisuje wynik do bazy danych i przechodzi do ścieżki *Wyświetlenie informacji szczegółowych dla danych wyborów*.

4.2.9. Wyświetlenie szczegółowych informacji o danym wyniku danych wyborów

1. Użytkownik klikając na link danego wyniku wyborów na stronie ze szczegółowymi informacjami o danych wyborach, wysyła za pomocą przeglądarki internetowej żądanie wyświetlenia szczegółowych informacji o danym wyniku dla danych wyborów.
2. Moduł URL Resolver kieruje żądanie do modułu zarządzania wyborami (widok *ResultDetailView*).
3. Widok pobiera z bazy danych informacje o zwycięzcach danych wyborów.
4. Jeżeli wybory zostały wygenerowane z rozkładu normalnego, widok przekierowuje działanie programu do modułu wizualizacji wyborów i wyników wyborów (widok *ResultChartView*). Jeżeli wybory nie zostały wygenerowane z rozkładu normalnego, następuje przejście do kroku nr 6

5. Widok pobiera z bazy danych obiekty będące punktami 2D reprezentującymi wyborców, kandydatów i zwycięzców. Ustawia parametry wykresu a następnie wysyła wszystkie dane z powrotem do widoku `ResultDetailsView`, umożliwiając stworzenie wykresu.
6. Widok wykorzystując szablon `result_details.html` tworzy i wysyła stronę do przeglądarki użytkownika.

4.2.10. Wczytanie wyborów z pliku .soc

1. Użytkownik klikając na przycisk *Load data from file* na stronie ze szczegółowymi informacjami o danych wyborach, wysyła za pomocą przeglądarki internetowej żądanie wczytania wyborów z pliku `.soc`.
2. Moduł URL Resolver kieruje żądanie do modułu generacji i wczytywania wyborów (widok `ElectionLoadDataFormView`).
3. Widok wykorzystując formularz `ElectionLoadDataForm` i szablon `election_load_data.html` tworzy oraz wysyła do użytkownika stronę z formularzem do wybrania pliku, z którego będą wczytywane wybory.
4. Użytkownik wskazuje plik i wysyła uzupełniony formularz do systemu.
5. Moduł URL Resolver kieruje żądanie do modułu generacji i wczytywania wyborów (widok `ElectionLoadDataFormView`).
6. Widok parsuje plik wejściowy i waliduje dane. W przypadku poprawnego wczytania pliku wejściowego następuje zapis wczytanych obiektów do bazy danych. Za pomocą szablonu `election_details.html` zostaje stworzona i wysłana strona do przeglądarki użytkownika.

4.2.11. Generacja wyborów z rozkładu normalnego

1. Użytkownik klikając na przycisk *Generate election from normal distribution* na stronie ze szczegółowymi informacjami o danych wyborach, wysyła za pomocą przeglądarki internetowej żądanie wygenerowania wyborów za pomocą rozkładu normalnego.
2. Moduł URL Resolver kieruje żądanie do modułu generacji i wczytywania wyborów (widok `ElectionGenerateDataFormView`).
3. Widok wykorzystując formularz `ElectionGenerateDataForm` i szablon `election_generate_data.html` tworzy oraz wysyła do użytkownika stronę z formularzem do określenia parametrów wyborów oraz i rozkładu normalnego.
4. Użytkownik uzupełnia formularz i wysyła go do systemu.
5. Moduł URL Resolver kieruje żądanie do modułu generacji i wczytywania wyborów (widok `ElectionGenerateDataFormView`).

6. Widok wczytuje dane z formularza i wykorzystując generator, generuje według wczytanych parametrów kandydatów, wyborców oraz liczy ich preferencje. Dane zostają zapisane do bazy. Następnie program przechodzi do 2-go punktu ścieżki *Wyświetlanie informacji szczegółowych dla danych wyborów*.

5. Struktura obiektowa przechowywanych danych

5.1. Opis modeli wykorzystywanych w ORM

Poniżej zamieszczono opis klas wykorzystywanych przez *Django* do odwzorowania obiektowej architektury systemu informatycznego na bazę danych.

5.1.1. Election

Election
+ user: ForeignKey(User) + name: CharField(150) + committee_size: Integer()
+ get_absolute_url(): Union[Str,Unicode] + is_set_up(): Boolean + is_generated(): Boolean

Rysunek 5.1: Obiekt mapowany na tabelę wyborów

Pola:

- `user` - użytkownik aplikacji
- `name` - nazwa wyborów
- `committee_size` - rozmiar zwycięskiego komitetu

Metody:

- `get_absolute_url()` - zwraca adres URL, który daje dostęp do danych wyborów z poziomu dokumentu *HTML*
- `is_set_up()` - zwraca prawdę jeśli dla danych wyborów zostali określani kandydaci i głosujący lub fałsz w przeciwnym przypadku

5.1.2. Candidate

Candidate
+ name: CharField(50, Nullable) + position: ForeignKey(Point, Nullable) + preferences: ManyToManyField(Voter, Preference) + election: ForeignKey(Election, related_name='candidates') + soc_id: Integer(Nullable)

Rysunek 5.2: Obiekt mapowany na tabelę z danymi kandydata

Pola:

- name - nazwa kandydata
- position - położenie kandydata w układzie współrzędnych, pole opcjonalne dla kandydatów generowanych z rozkładu normalnego
- preferences - pole mapowane na relację wiele do wielu łączącej kandydata i głosującego w tabeli *Preference*
- election - wybory, do których należy kandydat
- soc_id - identyfikator kandydata w odpowiadającym wyborom pliku w formacie *.soc*, pole opcjonalne dla wyborów parsowanych z pliku

5.1.3. Voter

Voter
+ repeats: Integer(default=1) + position: ForeignKey(Point, Nullable) + election: ForeignKey(Election, related_name='voters')
+ set_preferences_by_ids(ids: List(Int)): void + calculate_committee_score(committee: List(Candidate), p: Int, candidates_number: Int)

Rysunek 5.3: Obiekt reprezentujący głos w wyborach

Pola:

- repeats - liczba powtórzeń głosów w wyborach
- position - współrzędne wyborcy na wykresie, pole opcjonalne, stosowane dla wyborów generowanych z rozkładu normalnego
- election - wybory, do których należy głosujący

Metody:

- `set_preferences_by_ids()` - metoda wykorzystywana przy parsowaniu plików w formacie *.soc*. Na podstawie ciągu identyfikatorów wyborców pobranego z pliku tworzy ciąg kandydatów ułożony według preferencji wyborcy. Pobrane dane zapisuje w bazie danych w postaci rekordu w tabeli *Preference*.
- `calculate_committee_score(committee: List(Candidate), p: Int, candidates_number: Int)` - oblicza wynik komitetu *committee* na podstawie parametru *p* i liczby kandydatów *candidates_number*. Wynik mnożony jest przez liczbę powtórzeń głosu.

5.1.4. Preference

Preference
+ candidate: ForeignKey(Candidate) + voter: models.ForeignKey(Voter, related_name='preferences') + preference: Integer(NonNull)

Rysunek 5.4: Obiekt reprezentujący pozycję kandydata w liście preferencji głosującego

Pola:

- *candidate* - kandydat związany z preferencją
- *voter* - głosujący związany z preferencją
- *preference* - pozycja preferencji na liście preferencji głosującego

5.1.5. Point

Point
+ x: Integer + y: Integer
+ distance(): Float

Rysunek 5.5: Obiekt przechowujący współrzędne wyborcy lub kandydata w układzie współrzędnych.

Pola - współrzędne w układzie kartezjańskim:

- *x*: Int

- `y: Int`

Metody:

- `distance(other: Point) : Float` - metoda zwracająca odległość punktu od punktu `other` w metryce euklidesowej w przestrzeni R^2

5.1.6. Result

Result
+ election: ForeignKey(Election, related_name='results') + p_parameter: Integer() + winners: ManyToManyField(Candidate) + algorithm: CharField(1) + time: Float(Nullable) + score: Float(Nullable)
+ calculate_score(): Float + get_absolute_url(): Union[Str, Unicode]

Rysunek 5.6: Obiekt mapowany na tabelę z danymi wyników wyborów: rodzaj algorytmu, parametry wywołania algorytmu, otrzymany wynik punktowy komitetu oraz czas działania algorytmu

Pola:

- `election` - wybory, dla których obliczono rezultat
- `p_parameter` - parametr p normy ℓ_p
- `winners` - zwycięski komitet
- `algorithm` - identyfikator algorytmu
- `time` - czas wykonania obliczeń
- `score` - wynik punktowy zwycięskiego komitetu

5.1.7. GeneticAlgorithmSettings

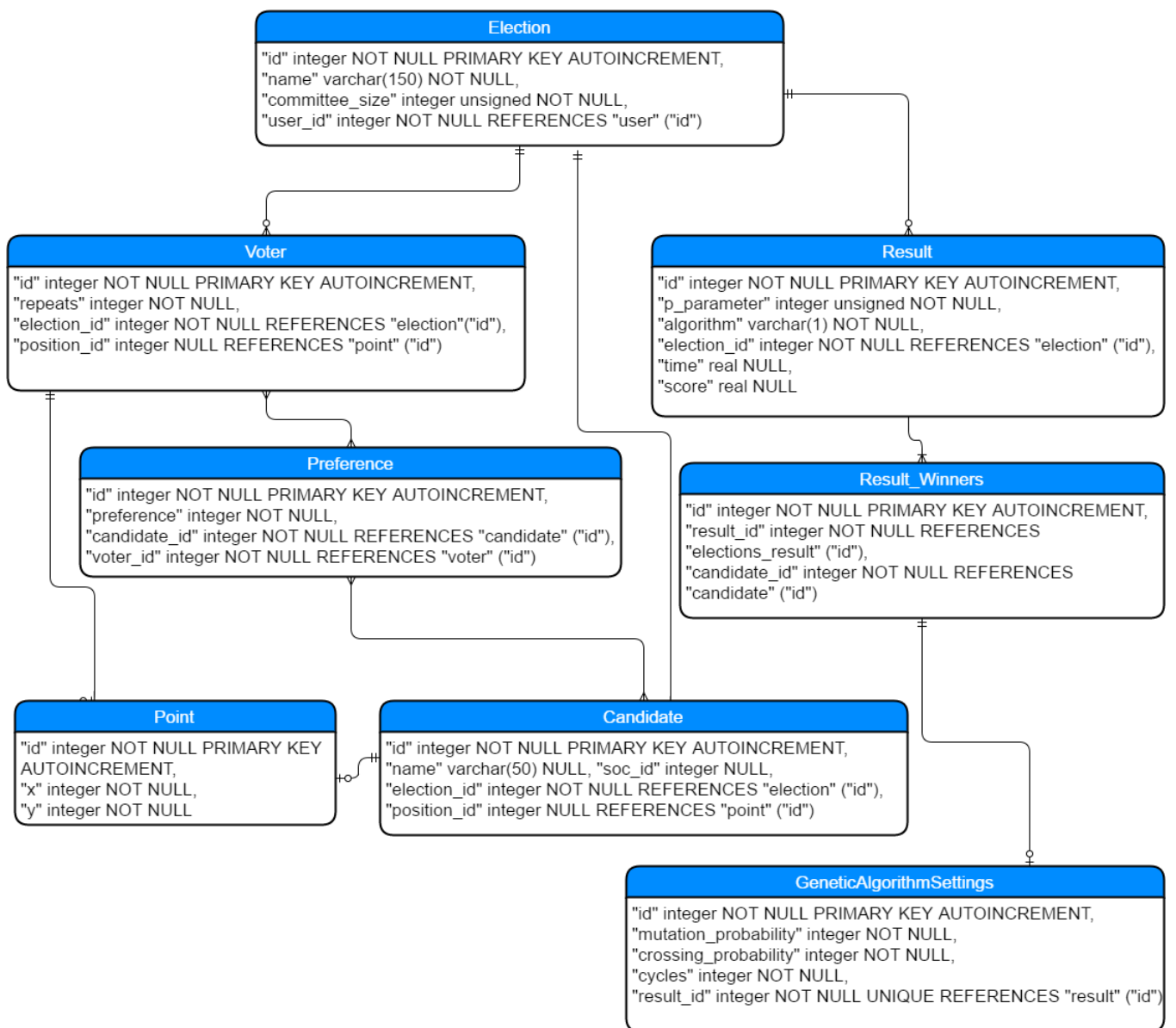
GeneticAlgorithmSettings
+ result: OneToOneField(Result) + mutation_probability: Integer(Min=0, Max=100, Default=10) + crossing_probability: Integer(Min=0, Max=100, Default=20) + cycles: Integer(Default=50)

Rysunek 5.7: Klasa mapowana na tabelę gromadząca dane specyficzne dla konfiguracji algorytmu genetycznego. Wykorzystywana do zapisania specyficznych dla algorytmu genetycznego parametrów wywołania.

Pola:

- `result` - rezultat algorytmu genetycznego
- `mutation_probability` - prawdopodobieństwo wystąpienia mutacji danego osobnika
- `crossing_probability` - prawdopodobieństwo wymiany puli genów pomiędzy parą osobników
- `cycles` - liczba iteracji algorytmu genetycznego

5.2. Diagram ERD



Rysunek 5.8: Diagram ERD

6. Algorytmy

6.1. Algorytm zachłanny

6.1.1. Idea algorytmu

Idea algorytmu polega na dołączaniu kolejnego kandydata do zwycięskiego komitetu w każdej iteracji. W związku z tym główna pętla algorytmu ma tyle iteracji, ile liczy rozmiar zwycięskiego komitetu. Jeżeli któryś z kandydatów został wybrany w danej iteracji jest pewne, iż znajdzie się on w zwycięskim Komitecie. Wybór kandydata w danej iteracji dokonywany jest w sposób zachłanny. Spośród kandydatów, którzy jeszcze nie zostali wybrani do zwycięskiego komitetu, dołącza do zwycięzców ten, który razem z wcześniej wybranymi kandydatami tworzy komitet o największym uznaniu wśród wyborców. Mając więc na danym etapie algorytmu komitet złożony z k zwycięzców, w kolejnej iteracji poszukiwany jest $k + 1$ -szy kandydat spośród pozostałych. Ma on stworzyć komitet liczący $k + 1$ członków, który jest najlepszy wśród wszystkich innych możliwości. W każdej iteracji algorytmu wybierany jest kandydat, który najwięcej dodaje do aktualnego poziomu zadowolenia wyborców. Stąd można uznać opisany algorytm jako zachłanny.

6.1.2. Funkcja zadowolenia

Aktualny poziom zadowolenia wyborców na danym etapie algorytmu określany jest za pomocą uprzednio zdefiniowanego uogólnienia systemu $k - Borda$. Satysfakcja pojedynczego wyborcy z komitetu liczącego na danym etapie algorytmu k zwycięzców, określana jest za pomocą funkcji:

$$f_{\ell_p - Borda}(i_1, i_2, \dots, i_k) = \sqrt[p]{[\beta(i_1)]^p + [\beta(i_2)]^p + \dots + [\beta(i_k)]^p} \quad (6.1)$$

Funkcja bazuje na normie ℓ_p oraz punktacji $k - Borda$. Dokładne informacje na temat funkcji umieszczone są w rozdziale *Dziedzina problemu*. Ogólne zadowolenie wyborców z komitetu jest obliczane jako suma pojedynczych zadowoleń wszystkich wyborców. Algorytm działa w ten sposób, że przed wyborem pierwszego zwycięzcy, poziom zadowolenia wyborców wynosi 0, gdyż zadowolenie każdego pojedynczego wyborcy wynosi 0 - zadowolenie z komitetu liczącego 0 osób zgodnie z powyższą funkcją wynosi 0. W pierwszej iteracji wybierany jest pierwszy zwycięzca w ten sposób, żeby suma pojedynczych zadowoleń wszystkich wyborców z jednoosobowego komitetu była największa. W kolejnej iteracji dobierany jest do wybranego wcześniej komitetu jednoosobowego drugi zwycięzca w ten sposób, by suma pojedynczych zadowoleń wszystkich wyborców z dwuosobowego komitetu była największa. W drugiej iteracji algorytmu brane więc pod uwagę są tylko wybrane komitety dwuosobowe - takie, które składają się z wcześniej wybranego zwycięzcy i dobranego wciąż jeszcze walczącego o zwycięstwo kandydata. W trzeciej iteracji do wybranego wcześniej komitetu dwuosobowego dobierany jest trzeci zwycięzca w ten sposób, by suma pojedynczych zadowoleń wszystkich wyborców z trzyosobowego komitetu była

największa. Podobnie jak w drugiej iteracji, pod uwagę brane są tylko wybrane komitety trzyosobowe - takie, które składają się z wcześniej wybranego komitetu dwuosobowego i dobranego wciąż jeszcze walczącego o zwycięstwo kandydata. W kolejnych iteracjach algorytmu, kandydaci dobierani są w sposób analogiczny. Kandydaci dobierani są aż do osiągnięcia określonego rozmiaru komitetu. Poziom zadowolenia wyborców po każdej iteracji algorytmu rośnie.

6.1.3. Szczegóły implementacji

Algorytm zaimplementowany jest w klasie `GreedyAlgorithm` znajdującej się w pakiecie `ecs.elections.algorithms.greedy_algorithm`. Główna pętla algorytmu znajduje się w metodzie `run()`. Przed główną pętlą programu do pamięci operacyjnej pobrane są obiekty z bazy danych reprezentujące kandydatów, wyborców oraz preferencje wyborców. Ponadto odpowiednio uzupełniane są struktury danych reprezentujące aktualne zadowolenie wyborców z wybranych kandydatów, aktualny skład zwycięskiego komitetu oraz kandydatów wciąż walczących o zwycięstwo, którzy stanowią wszystkich kandydatów bez tych, którzy już zostali wybrani do zwycięskiego komitetu. Początkowo poziom zadowolenia każdego wyborcy wynosi 0, skład zwycięskiego komitetu jest pusty, a kandydaci wciąż walczący o zwycięstwo stanowią zbiór wszystkich kandydatów.

W głównej pętli algorytmu znajdują się dwie wewnętrzne pętle. Bardziej zewnętrzna pętla iteruje po wciąż walczących o zwycięstwo kandydatach, bardziej wewnętrzna pętla iteruje po wszystkich wyborcach. Wskazane dwie wewnętrzne pętle mają na celu wyłonienie najlepszego kandydata spośród jeszcze niewybranych kandydatów do zwycięskiego komitetu. Dzieje się to poprzez obliczenie poziomu zadowolenia wyborców z komitetu złożonego z kandydatów już wcześniej wybranych (na początku komitet zwycięski jest pusty) wraz z dodatkowo dobranym kandydatem wcześniej niewybranym. Obliczany jest poziom zadowolenia dla każdego możliwego dobrania (iteracja po kandydatach jeszcze niewybranych) i ostatecznie w danej iteracji głównej pętli algorytmu do zwycięskiego komitetu, dołączany jest kandydat, który wraz z wcześniej wybranymi kandydatami stanowi największy poziom zadowolenia wyborców. W algorytmie najlepszy kandydat w danej iteracji dobierany jest w ten sposób, że wskazany jest aktualny lider spośród kandydatów jeszcze niewybranych wraz z poziomem satysfakcji wyborców z komitetu utworzonego z tego lidera i z wcześniej wybranych kandydatów. Jeżeli poziom satysfakcji wyborców z komitetu z innym niewybranym jeszcze kandydatem okaże się większy od tego z aktualnym liderem, aktualny lider zostaje zmieniony a poziom satysfakcji zostaje aktualizowany. Ostatecznie w danej iteracji głównej pętli algorytmu, do zwycięskiego komitetu dołączany zostaje lider po zakończeniu iteracji po wszystkich jeszcze niewybranych kandydatach. Ponieważ poziom zadowolenia z komitetu ustalany jest jako suma pojedynczych zadowoleń wszystkich wyborców, w celu obliczenia funkcji satysfakcji dla komitetu złożonego z wcześniej wybranych kandydatów wraz z dobranym kandydatem wcześniej niewybranym, następuje iteracja po wszystkich wyborcach. Dla każdego wyborcy oceniane jest pojedyncze zadowolenie z komitetu i dodawane do sumy zadowolenia wszystkich wyborców. Pojedyncze zadowolenie liczone jest w ten sposób, że wyciągana jest punktacja badanego kandydata według punktacji k – Borda, a następnie liczona jest norma ℓ_p z aktualnego

zadowolenia danego wyborcy (z wcześniej wybranych kandydatów) i punktów $k - Borda$ przyznanych dla badanego kandydata. Po wybraniu zwycięzcy, ostatnim krokiem w głównej pętli algorytmu jest aktualizacja struktur danych reprezentujących aktualny stan algorytmu. Do zwycięskiego komitetu zostaje dołączony wybrany w danej iteracji kandydat, przy jednoczesnym usunięciu go ze zbioru kandydatów jeszcze niewybranych. Aktualizowane zostają również pojedyncze satysfakcje wszystkich wyborców w oparciu o wybranego w danej iteracji kandydata, który w tym momencie staje się zwycięzcą.

6.1.4. Pseudokod

Data: k - rozmiar komitetu, $porzadek_preferencji$ - porządek preferencji m kandydatów dla n wyborców, p - parametr funkcji zadowolenia

Result: zwycięski zbiór k kandydatów - zbiór *REZULTAT*

$REZULTAT \leftarrow \emptyset$;

$C \leftarrow$ zbiór wszystkich m kandydatów;

$V \leftarrow$ zbiór wszystkich n wyborców;

for $i \leftarrow 1$ **to** k **do**

for $c \in C \setminus REZULTAT$ **do**

for $v \in V$ **do**

$punkty_Bordy \leftarrow$

$liczba_punktow_w_porzadku_preferencji(c, v, porzadek_preferencji)$;

$zadowolenie_v \leftarrow aktualne_zadowolenie(v)$;

$zadowolenie_v_wraz_z_kandydatem_c \leftarrow$

$\sqrt[p]{zadowolenie_v^p + punkty_Bordy^p}$;

$zadowolenie_z_kandydata_c += zadowolenie_v_wraz_z_kandydatem_c$;

end

if $zadowolenie_z_kandydata_c >$

$zadowolenie_z_najlepszego_dotychczas_kandydata$ **then**

$prowadzacy_kandydat_w_danej_iteracji \leftarrow c$;

$zadowolenie_z_najlepszego_dotychczas_kandydata \leftarrow$

$zadowolenie_z_kandydata_c$;

end

$uaktualnij_zadowolenia_wyborcow(prowadzacy_kandydat_w_danej_iteracji)$;

$REZULTAT \leftarrow REZULTAT \cup prowadzacy_kandydat_w_danej_iteracji$;

end

end

Algorithm 1: Pseudokod algorytmu zachłannego

6.2. Algorytm zachłanny według zasady Chamberlina-Couranta

Algorytm ten nie aproksymuje wyników wyborów na bazie funkcji satysfakcji opisującej system $\ell_p - Borda$, ale na szczególnym przypadku tej funkcji gdy $p \rightarrow \infty$. W związku z tym parametr p

nie wpływa na działanie algorytmu.

6.2.1. Idea algorytmu

Algorytm został zaproponowany przez Lu i Boutilier’a. Idea jest taka sama jak dla pierwszego algorytmu zachłannego. W każdej iteracji głównej pętli algorytmu do zwycięskiego komitetu dobrany zostaje jeden kandydat. Zwycięzca w danej iteracji zostaje dobrany według postępowania zachłannego. Wygrywa kandydat, który najwięcej dodaje do ogólnego poziomu zadowolenia wyborców. Różnica w stosunku do pierwszego algorytmu zachłannego polega na innej funkcji zadowolenia, na bazie której inaczej obliczane są pojedyncze zadowolenia wyborców i tym samym inaczej obliczane jest ogólne zadowolenie wyborców z danego komitetu, które jest sumą pojedynczych satysfakcji wszystkich wyborców. W regule *Chamberlina-Couranta* położony jest nacisk na regułę reprezentacji proporcjonalnej. Każdy wyborca reprezentowany jest przez jednego, wyraźnego kandydata i tylko punkty tego kandydata wpływają na wartość pojedynczego zadowolenia wyborcy z danego komitetu. Jest to szczególny przypadek uogólnienia systemu $k - Borda$, w którym $p \rightarrow \infty$. Jednak dla odpowiednio dużych wartości parametru p opisana heurystyka dobrze sprawdza się dla uogólnienia systemu $k - Borda$.

6.2.2. Funkcja zadowolenia

Aktualny poziom zadowolenia wyborców na danym etapie algorytmu określany jest za pomocą funkcji satysfakcji *Chamberlina-Couranta*:

$$f_{CC}(i_1, i_2, \dots, i_k) = \beta(i_1) \quad (6.2)$$

Jest ona skrajnym przypadkiem dla uogólnienia systemu $k - Borda$, kiedy $p \rightarrow \infty$. Zgodnie z tą funkcją, pojedyncze zadowolenie danego wyborcy z komitetu może wzrosnąć tylko w przypadku, gdy do komitetu dołączany zostaje kandydat, któremu dany wyborca przydzielił więcej punktów $k - Borda$ od punktów aktualnego reprezentanta. Jeżeli taki kandydat zostanie dołączony do zwycięzców, to on staje się reprezentantem dla danego wyborcy w aktualnym stanie algorytmu. Jest to odmienna sytuacja od pierwszego algorytmu zachłannego, w którym pojedyncze zadowolenie danego wyborcy z komitetu rośnie wtedy, gdy nowo dobrany kandydat ma jakiegokolwiek przydzielone punkty $k - Borda$. W tamtym przypadku nowo dołączony zwycięzca nie musi być najlepszy spośród wszystkich aktualnie wybranych kandydatów do zwycięskiego komitetu, aby polepszyć indywidualne zadowolenie danego wyborcy i tym samym zbiorcze zadowolenie wyborców.

6.2.3. Szczegóły implementacyjne

Algorytm zaimplementowany jest w klasie `GreedyCC` znajdującej się w pakiecie `ecs.elections.algorithms.greedy_cc`. Główna pętla algorytmu znajduje się w metodzie `run()`. Implementacja i kolejne kroki algorytmu są takie same lub bardzo podobne jak dla pierwszego algorytmu zachłannego zależnego od parametru p . Największa różnica jest w najbardziej wewnętrznej pętli iterującej po wszystkich wyborcach. Zaimplementowana jest tu funkcja zadowolenia pojedynczego wyborcy, różniąca się od analogicznej funkcji dla pierwszego algorytmu zachłannego.

Badany kandydat zwiększa wartość pojedynczej satysfakcji danego wyborcy tylko wtedy, gdy ma on przydzielone u wyborcy najwięcej punktów $k - Borda$ w porównaniu z dotychczasowymi zwycięzcami. W celu sprawdzenia czy potencjalny zwycięzca w danej iteracji, zwiększa satysfakcję danego wyborcy, porównane są dwie wartości: punkty $k - Borda$ przyznane przez wyborcę dla potencjalnego zwycięzcy i punkty $k - Borda$ przyznane przez wyborcę dla najlepszego spośród kandydatów aktualnego zwycięskiego komitetu. Jeśli punkty dla badanego kandydata są większe od punktów najlepszego dotychczasowego zwycięzcy, dodatkowa satysfakcja dla badanego kandydata od danego wyborcy jest uzupełniana o różnicę między punktami $k - Borda$ przydzielonymi dla badanego kandydata i punktami $k - Borda$ przydzielonymi dla najlepszego dotychczas kandydata. W danej iteracji wygrywa kandydat, który wniósł najwięcej dodatkowej satysfakcji do ogólnego zadowolenia wyborców. Drugą istotną różnicą wspomnianą przy omawianiu pierwszej, jest sposób porównania w danej iteracji niewybranych jeszcze kandydatów w celu ustalenia najlepszego z nich. Polega on na zbadaniu jaką dodatkową satysfakcję do ogólnego poziomu zadowolenia wnosi badany kandydat i porównywaniu tych dodatkowych satysfakcji, a nie wartości ogólnego poziomu zadowolenia jak w przypadku pierwszego algorytmu.

6.3. Algorytm genetyczny

Algorytmy genetyczne polegają na przedstawieniu różnych wyników problemu jako *osobników*, a następnie, w sposób losowy, ich krzyżowanie i mutowanie. Po każdym takim cyklu osobniki są poddawane ocenie i do dalszych cykli pozostawiane są tylko najlepsze okazy. Algorytm opiera się na znanej z biologii ewolucji.

6.3.1. Osobnik

W naszej implementacji osobnikiem jest wybierany komitet (podzbiór kandydatów). Po utworzeniu jest on oceniany według funkcji satysfakcji i przypisywany jest mu wynik.

6.3.2. Krzyżowanie

Krzyżowanie osobników - komitetów polega na losowym wybraniu kandydatów z dwóch osobników, tak by łącznie było ich k i żaden się nie powtarzał (krzyżowane osobniki mogą zawierać tych samych kandydatów).

6.3.3. Mutacja

Mutacja osobnika - komitetu polega na losowej zmianie jednego z kandydatów z komitetu na innego, wcześniej niewybranego.

6.3.4. Parametry algorytmu

System pozwala wybrać:

- liczbę cykli działania algorytmu

- część puli jaka jest poddawana krzyżowaniu (osobniki są z niej losowo wybierane do krzyżowania)
- prawdopodobieństwo wystąpienia mutacji w pojedynczym osobniku

6.3.5. Cykl algorytmu

Na cykl algorytmu składa się:

- wybór par osobników do krzyżowania się
- krzyżowanie osobników
- mutacja osobników zgodnie z zadaniem prawdopodobieństwem
- wybór N osobników do przejścia do kolejnego cyklu

Liczba N określa wielkość puli osobników i jest wybierana na początku działania algorytmu. W naszej implementacji przyjęliśmy stałą liczbę 50 osobników.

6.3.6. Wpływ parametrów działania na wyniki osiągane przez algorytm

Aby przekonać się jak dobór parametrów ma wpływ na wyniki osiągane przez algorytm, przeprowadziliśmy kilka testów na losowo wygenerowanych wyborach (korzystając z funkcji systemu losującej kandydatów i wyborców na podstawie rozkładu normalnego).

Dla przykładu - dla wyborów 20 spośród 100 kandydatów, przy 100 wyborcach, czasy działania algorytmu wahały się w zakresie od 4,3s do 150,3s. Przy czym wyniki różniły się również znacząco - komitety uzyskiwały od 10715,65 do 10867,96 punktów. W tej serii zostały przetestowane wszystkie kombinacje parametrów spośród następujących:

liczba cykli $\in \{20, 50, 100\}$

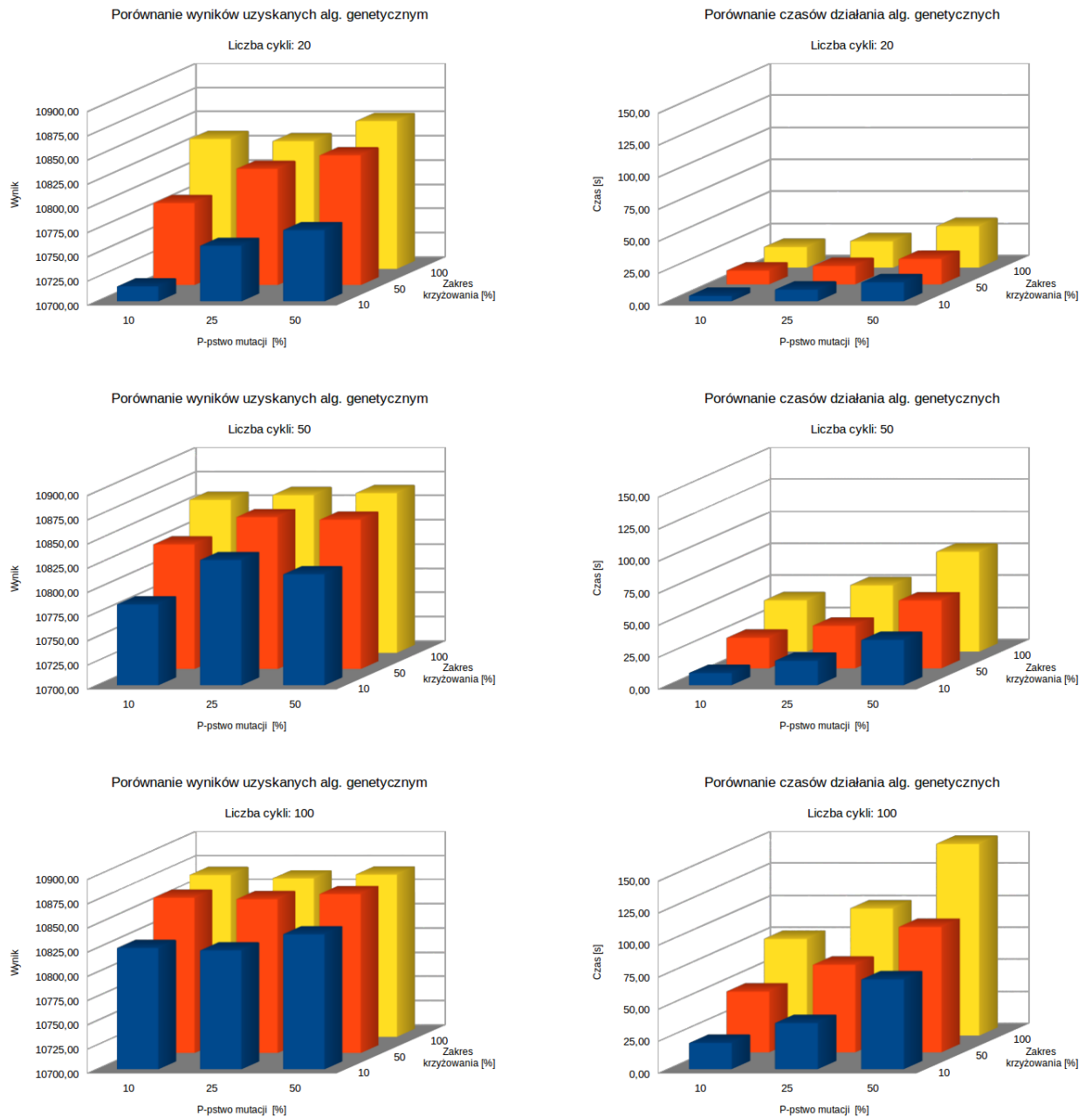
prawdopodobieństwo mutacji $\in \{10\%, 25\%, 50\%\}$

krzyżowanie $\in \{10\%, 50\%, 100\%\}$

Okazuje się jednak, że nawet te parametry nie wystarczają do wyznaczenia najlepszego wyniku - obliczenia w 200 cyklach (z prawdopodobieństwem mutacji 50% i krzyżowaniem całej puli) dały jeszcze wyższy wynik - 10868.15, przy czasie 364s.

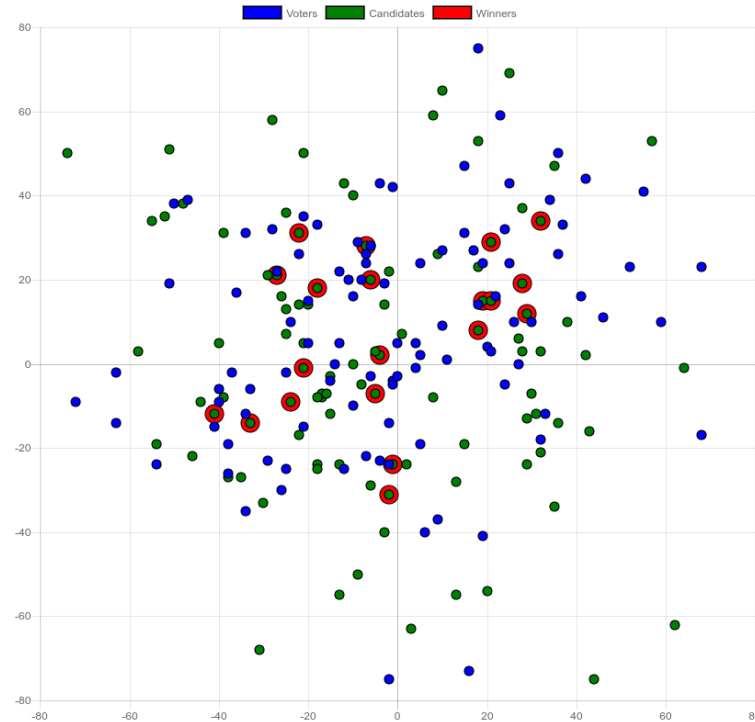
Należy mieć na uwadze, że w działaniu algorytmu genetycznego, z powodu dużego udziału czynnika losowego, wiele zależy od niego i wyniki są z tym obarczone.

Poniżej znajdują się wykresy obrazujące czas i wyniki osiągane przez algorytm genetyczny dla tego zestawu danych przy różnych parametrach:

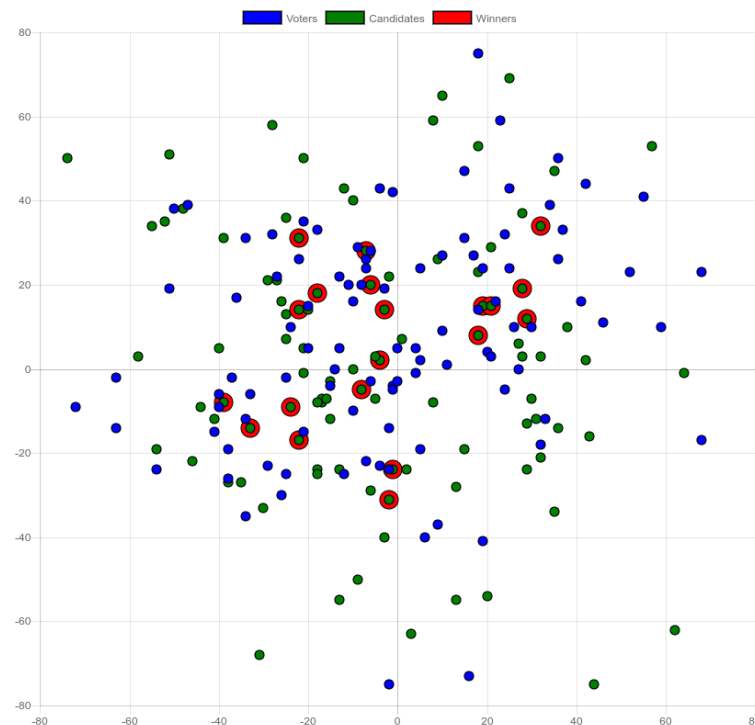


Rysunek 6.1: Porównanie wyników i czasów działania algorytmu genetycznego z różnymi parametrami

Należy pamiętać, że nawet mała zmiana wyniku punktowego komitetu może mieć duży wpływ na faktycznie wybranych kandydatów - poniżej znajdują się mapy preferencji z zaznaczonymi dwoma wybranymi komitetami, których wynik różni się zaledwie o około 19 setnych punkta. Są to dwa najlepsze rezultaty wspomniane wcześniej.

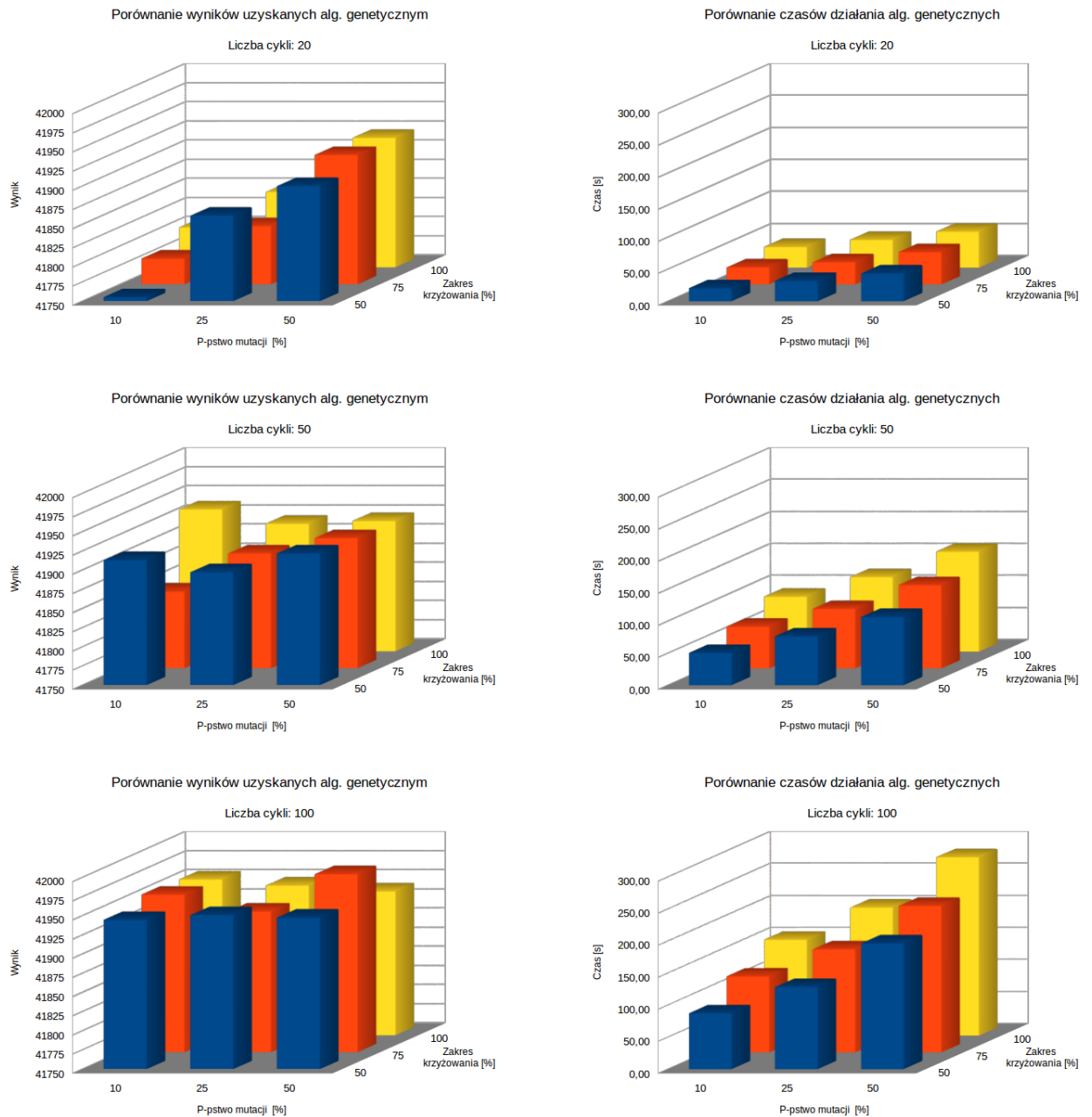


Rysunek 6.2: Wybrany w 150,29s komitet o punktacji 10867,96



Rysunek 6.3: Wybrany w 363,92s komitet o punktacji 10868,15

Podobne zestawienie wykonaliśmy jeszcze dla "większych" wyborów: 200 głosujących na 200 kandydatów, 15 miejsce w Komitecie. Poniżej znajduje się porównanie osiągniętych wyników:



Rysunek 6.4: Porównanie wyników i czasów działania algorytmu genetycznego z różnymi parametrami

W tym wypadku również zwiększenie liczby cykli do 200 pozwoliło na jeszcze lepszy wynik, windując jednak czas wykonania programu do nieco ponad 10 minut. Co ciekawe próba przy 300 cyklach dała wynik gorszy - ujawnia się tutaj losowa natura algorytmu genetycznego.

7. Porównanie algorytmów

7.1. Zrównanie jakości działania algorytmów zachłannych

Ponieważ algorytm zachłanny według zasady *Chamberlina-Couranta* przystosowany jest do działania dla dużych wartości parametru p (im większy parametr p , tym wartość wskazanego przez algorytm komitetu powinna rosnąć), zasadnym jest analiza dla jakiego parametru p algorytm zachłanny według zasady *Chamberlina-Couranta* staje się nie gorszy od algorytmu zachłannego zależnego od parametru p . Algorytm zachłanny według zasady *Chamberlina-Couranta* jest znacznie szybszy od swojego konkurenta, więc wiedząc dla jakiej wartości parametru p staje się nie gorszy pod względem jakości rezultatu, można go z powodzeniem stosować zamiast algorytmu zachłannego zależnego od parametru p . W tym podrozdziale podjęta zostanie próba oszacowania wartości parametru p , dla którego algorytmy zachłanne zrównują się pod względem jakości wyniku oraz zostanie przeprowadzona analiza czynników, które mogą potencjalnie wpływać na wartość tego parametru p . W dalszej części podrozdziału parametr p , dla którego algorytmy zachłanne zrównują się pod względem jakości wyniku będzie nazywany graniczną wartością parametru p .

7.1.1. Działanie algorytmów zachłannych

Dla wszystkich przypadków testowanych, wyniki algorytmów zachłannych zachowywały pewien schemat wraz ze wzrostem parametru p . Dla małych wartości parametru p jakość rezultatu obliczonego przez algorytm zachłanny zależny od wartości parametru p , znacznie przewyższa drugi z algorytmów zachłannych. Wraz ze wzrostem parametru p jakości wyników wyliczanych przez algorytmy są coraz bliższe sobie. Dla pewnej wartości granicznej parametru p , algorytm zachłanny według zasady *Chamberlina-Couranta* staje się nie gorszy od swojego konkurenta pod względem jakości wyniku. Dla wartości parametru p większych od tej wartości granicznej, algorytm zachłanny według zasady *Chamberlina-Couranta* jest co najmniej tak samo dobry jak algorytm zachłanny zależny od parametru p . Dla wartości parametru p większych i równych od wartości granicznej parametru p , najczęściej występuje taka sytuacja, że wyniki obydwu algorytmów są takie same.

Powyższa zasada zachowywania się wyników algorytmów zachłannych względem siebie sprawdziła się dla wszystkich przypadków testowanych. Ponieważ badane algorytmy są algorytmami heurystycznymi, mogą wystąpić odstępstwa od tej reguły.

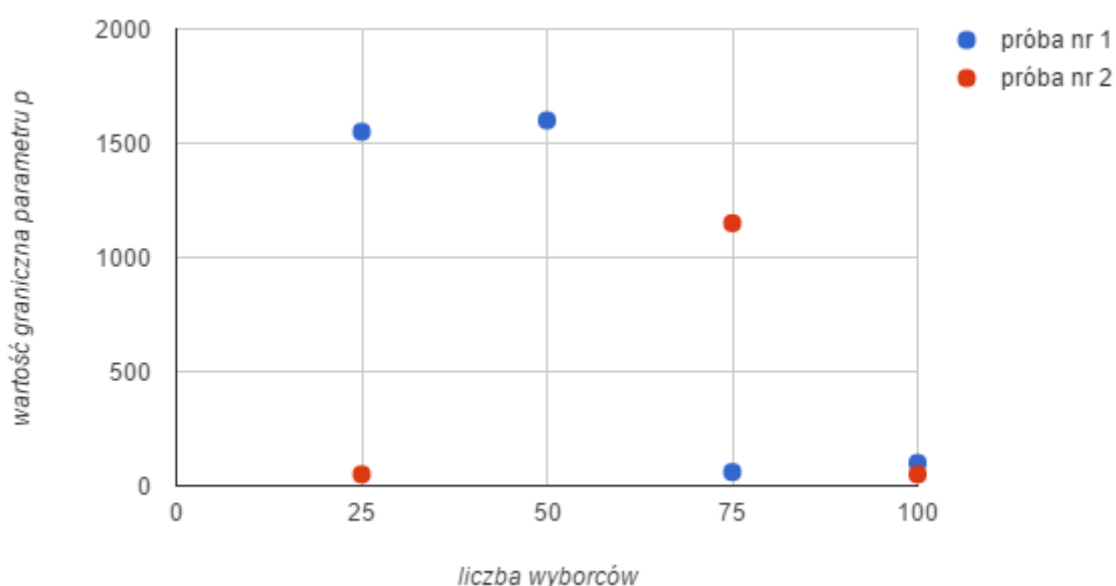
7.1.2. Wpływ liczby wyborców na wartość graniczną parametru p

Pierwszym czynnikiem, który badano pod kątem jego wpływu na wartość graniczną parametru p , była liczba wyborców. W tym celu określono pewne stałe wartości dla innych parametrów definiujących wybory, a następnie szacowano wartość graniczną parametru p . Szacowanie polegało na poszukiwaniu

wartości granicznej w ten sposób, że porównywano dla pewnej wybranej początkowo wartości parametru p wyniki obydwu algorytmów i w zależności, który algorytm okazał się lepszy pod względem jakości wyniku, dalsze poszukiwanie ograniczano do wartości parametru p mniejszych lub większych wybranej początkowo wartości. Jeżeli lepszy okazywał się algorytm zachłanny zależny od wartości parametru p , wtedy dalsze poszukiwanie ograniczano do wartości parametru p większych od badanej początkowo wartości parametru p . W przeciwnym przypadku szukano dla wartości parametru p mniejszych od wybranej na początku wartości. Postępowanie w celu oszacowania wartości granicznej parametru p miało cechy wyszukiwania binarnego. Wartość graniczna parametru p nie była określana dokładnie, lecz jedynie szacowana. Błąd bezwzględny określonej wartości wynosił dla tego testu 50. Jeżeli przykładowo stwierdzono, że wartość graniczna znajduje się między wartością 1500 a 1600, to szacowano, że wartość graniczna parametru p wynosi 1550. Dla tego testu przyjęto, że rozmiar komitetu będzie wynosił 15, a liczba kandydatów będzie wynosiła 50. Zbadano zależność wartości granicznej parametru p dla liczby wyborców wynoszącej 25, 50, 75 oraz 100. Próbę przeprowadzono dwukrotnie dla każdej wartości liczby wyborców w celu określenia czy wynik dla danej liczby wyborców jest podobny niezależnie od innych parametrów. Testy przeprowadzono dla wyborów generowanych z rozkładu normalnego. Wyniki przedstawione są w formie tabeli i wykresu.

liczba wyborców	25	50	75	100
wartość graniczna parametru p - próba nr 1	1550	1600	50	100
wartość graniczna parametru p - próba nr 2	50	ponad 5000	1150	50

Tablica 7.1: Graniczne wartości parametru p dla różnej liczby wyborców



Rysunek 7.1: Zależność wartości granicznej parametru p od liczby wyborców

Na wykresie nie zaznaczono wyniku drugiej próby dla liczby wyborców równej 50. Nie oszacowano jej dokładnie, gdyż wiadomo że wynosiła ponad 5000. Nie jest istotne jak duża jest ta wartość. Badanie to pokazało, że nie ma zależności między wartością graniczną parametru p a liczbą wyborców. Rozrzut wyników jest bardzo duży nawet dla dwóch różnych prób dla tej samej liczby wyborców.

7.1.3. Wpływ odchylenia standardowego na wartość graniczną parametru p

Zdecydowano się na sprawdzenie wpływu rozproszenia wyborców i kandydatów na wartość graniczną parametru p . Wartości parametru były takie same jak w poprzednim teście poza wartością odchylenia standardowego. Wcześniej wartość odchylenia standardowego dla wyborców i kandydatów wynosiła domyślne 30. W tym teście podniesiono tę wartość do 100. Sposób przeprowadzenia testu był analogiczny do poprzedniego. Wyniki przedstawione są w formie tabeli.

liczba wyborców	25	50
wartość graniczna parametru p - próba nr 1	ponad 5000	1500
wartość graniczna parametru p - próba nr 2	50	50

Tablica 7.2: Graniczne wartości parametru p dla różnej liczby wyborców i $\sigma = 100$

Ponieważ po dwukrotnym sprawdzeniu wyniku dla dwóch różnych liczb wyborców, rozrzut wyników okazał się równie duży jak w przypadku pierwszego testu, stąd porzeczono na tych wynikach. Wpływu odchylenia standardowego w ten sposób badanego na graniczną wartość parametru p nie odnotowano.

7.1.4. Wpływ liczby kandydatów na wartość graniczną parametru p

Kolejnym zbadanym czynnikiem był wpływ liczby kandydatów na wartość graniczną parametru p przy założeniu, że pozostałe parametry definiujące wybory są stałe. Na potrzeby testu przyjęto, że rozmiar komitetu będzie równy 15, liczba wyborców wyniesie 50, a odchylenie standardowe zarówno dla wyborców, jak i kandydatów wyniesie domyślne 30. Sposób przeprowadzenia testu był analogiczny do poprzednich. Błąd bezwzględny oszacowanych wartości granicznych, dla wartości rzędu kilku tysięcy wynosił 250, dla rzędu kilkaset lub kilkadziesiąt 50. Jeżeli przykładowo stwierdzono, że wartość graniczna znajduje się między wartością 3000 a 3500, to szacowano, że wartość graniczna parametru p wynosi 3250. Większa dokładność okazała się nieistotna. Wyniki przedstawione są w formie tabeli.

liczba kandydatów	20	40
wartość graniczna parametru p - próba nr 1	850	3250
wartość graniczna parametru p - próba nr 2	250	50

Tablica 7.3: Graniczne wartości parametru p dla różnej liczby kandydatów

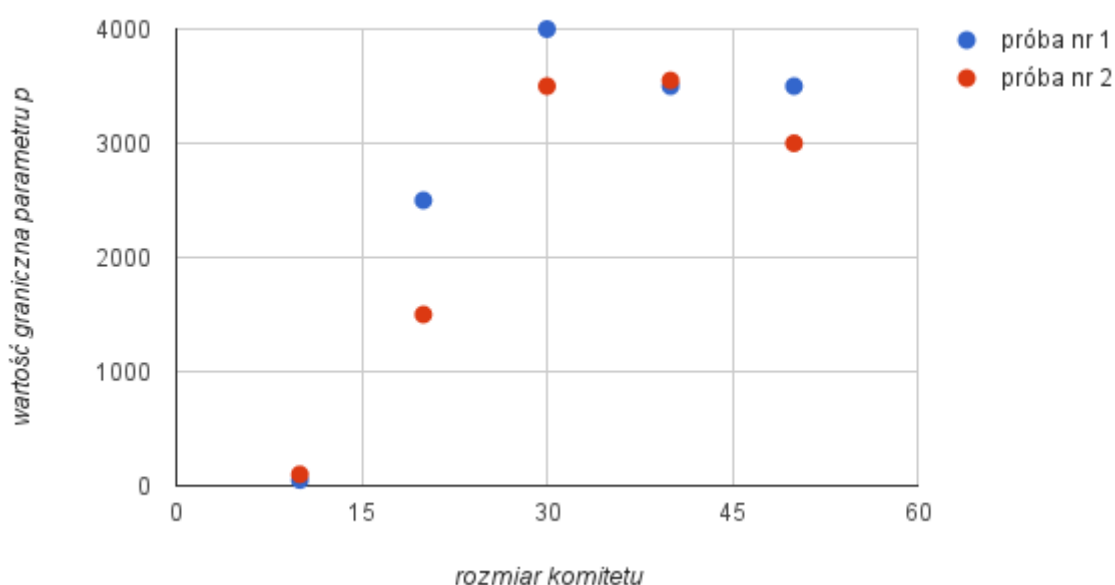
Ponieważ podobnie jak w poprzednim teście rozrzut wyników okazał się bardzo duży, zdecydowano porzucić na dwukrotnym sprawdzeniu dwóch różnych liczb kandydatów. Dla takiej samej liczby kandydatów, wartość graniczna parametru p potrafi być jednocześnie i bardzo duża i bardzo mała. Wpływu w ten sposób badanej liczby kandydatów na graniczną wartość parametru p nie znaleziono.

7.1.5. Wpływ rozmiaru komitetu na wartość graniczną parametru p

Kolejnym badanym czynnikiem był rozmiar komitetu. Metodyka badania była taka sama jak w poprzednim teście. Błąd bezwzględny oszacowanych wartości granicznych, dla wartości rzędu kilku tysięcy wynosił 500, dla rzędu kilkaset lub kilkadziesiąt 50. Parametry definiujące wybory poza rozmiarem komitetu przyjęto jako stałe. Liczbę kandydatów określono na wartość 60, liczbę wyborców na 50, a odchylenie standardowe wyborców i kandydatów na domyślne 30. Wyniki przedstawiono w formie tabeli i wykresu.

rozmiar komitetu	10	20	30	40	50
wartość graniczna parametru p - próba nr 1	50	2500	4000	3500	3500
wartość graniczna parametru p - próba nr 2	100	1500	3500	3500	3000

Tablica 7.4: Graniczne wartości parametru p dla różnego rozmiaru komitetu



Rysunek 7.2: Zależność wartości granicznej parametru p od rozmiaru komitetu

Rozrzut wyników dwóch różnych prób dla badanych rozmiarów komitetów był w porównaniu do poprzednich badań stosunkowo niewielki. Dla najmniejszego rozmiaru komitetu wynoszącego 10, wartość graniczna parametru p była rzędu kilkadziesiąt. Dla rozmiaru komitetu równego 20 rozrzut wyników był największy. Dla większych rozmiarów komitetu wartość graniczna była na poziomie kilka tysięcy i zdawała się mieć pewien rodzaj stabilizacji. Wciąż jednak trudno stwierdzić czy istnieje jakkolwiek zależność między badanymi wielkościami. Można jedynie zaobserwować, że dla mniejszych rozmiarów komitetu wartość graniczna przybiera dość często małą wartość rzędu kilkadziesiąt, a dla większych rozmiarów komitetu wartość graniczna jest raczej rzędu kilka tysięcy.

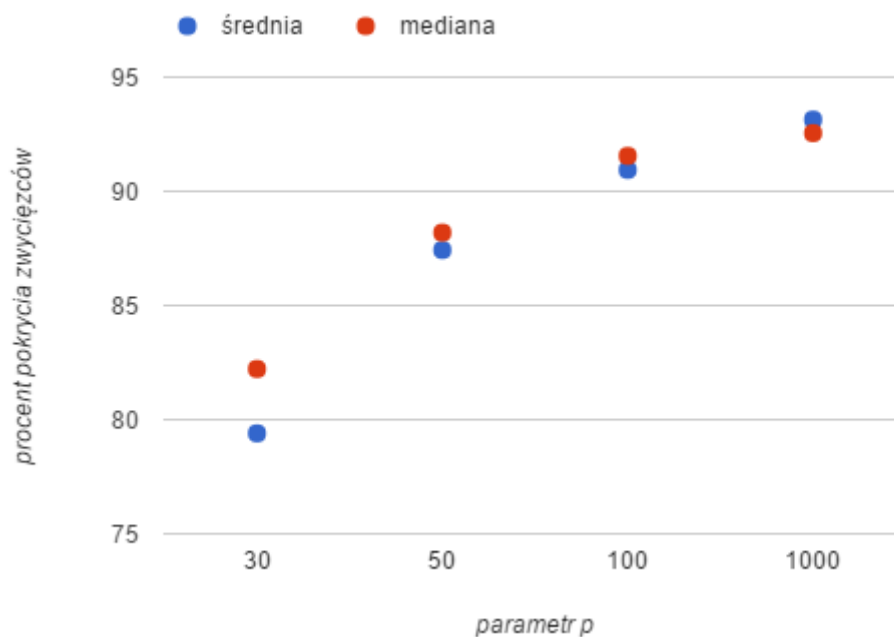
Przeprowadzając wszystkie powyższe testy zauważono jednak inną prawidłowość. Otóż zaobserwowano, że już dla niewielkiej wartości parametru p większość zwycięzców wskazanych przez obydwa algorytmy pokrywa się ze sobą. O innej wartości wskazanego komitetu dla większych wartości parametru p decyduje niewielka część zwycięzców. Tworzą się pary lub trójki kandydatów, którzy znajdują się blisko siebie i w zależności od tego który z nich zostanie wybrany do zwycięskiego komitetu, wartości zadowolenia wyborców się nieco zmieniają. Przykładowo może się tak zdarzyć, że istnieje para kandydatów spośród której zawsze do zwycięskiego komitetu trafia jeden z kandydatów w tej parze. Przy czym istnieje graniczna wartość parametru p , poniżej której zawsze jeden z kandydatów w tej parze wchodzi do zwycięskiego komitetu i jednocześnie powyżej tej granicznej wartości parametru p drugi kandydat z pary wchodzi do zwycięskiego komitetu. Jeżeli ta graniczna wartość jest duża, to wtedy algorytm zachłanny zależny od parametru p długo będzie wybierał pierwszego kandydata z pary przewyższając tym samym algorytm według zasady *Chamberlina-Couranta*, który zawsze wybierze drugiego kandydata. Stąd graniczna wartość parametru p dla tego przypadku jest duża, bo algorytm zachłanny zależny od parametru p będzie długo minimalnie lepszy od drugiego algorytmu zachłannego. Jednak trzon zwycięzców dla obydwu algorytmów szybko jest taki sam. Stąd na wartość graniczną parametru p głównie ma wpływ lokalne rozmieszczenie kandydatów. Jeżeli znajdują się pary i trójki, spośród których dla dużych wartości parametru p następuje zamiana miejsc w zwycięskim komitecie, wtedy wartość graniczna parametru p jest dla takich przypadków duża. Jeżeli takich zaciętych par nie ma i zwycięzcy są wyłaniany w zdecydowany sposób pod kątem reprezentatywności dla pewnej części wyborców, wtedy wartość graniczna parametru p jest mała. Poniekąd taka interpretacja tłumaczy tendencję do małej wartości granicznej parametru p dla małych zwycięskich komitetów i dużą wartość graniczną parametru p dla dużych zwycięskich komitetów. Dla małych rozmiarów komitetów, każdy ze zwycięzców klarownie reprezentuje pewną część wyborców przez co nie tworzą się dla większych p pary rywalizujące o reprezentację dla pewnej części wyborców. Dla dużych komitetów, w związku z wieloma miejscami w komitecie o zwycięstwo rywalizują również mniej popierani w danej grupie wyborców kandydaci, przez co jest większa szansa na stworzenie się opisanej wcześniej pary lub trójki kandydatów.

7.1.6. Stopień pokrywania się zwycięzców dla różnych wartości parametru p

W związku z zaobserwowaniem faktu dużego pokrywania się zwycięskich kandydatów wyłanianych przez algorytmy zachłanne dla odpowiednio dużych wartości parametru p , postanowiono dokładniej zbadać tę kwestię. Sprawdzone jaki procent kandydatów ze zwycięskich komitetów wyłonionych przez obydwa algorytmy pokrywa się ze sobą. Testy przeprowadzono dla różnych wartości definiujących wybory. Dla każdej ze sprawdzonych wartości parametru p sprawdzono pokrycie się zwycięzców dla 10 różnych wyborów. Dla każdej wartości parametru p policzono następnie średnią i medianę z wyliczonych wartości procentowych pokrycia się komitetów. Wyniki poszczególnych prób i obliczonych średnich oraz median przedstawiono w formie tabeli. W tabeli przyjęto następujące oznaczenia: k - rozmiar komitetu, c - liczba kandydatów, v - liczba wyborców. Dla poszczególnych wyborów przedstawiono ilu zwycięskich kandydatów z całego komitetu pokryło się z wynikami drugiego algorytmu dla danej wartości parametru p . Średnie i mediany przedstawiono również w formie wykresu.

parametr p	30	50	100	1000
$k = 10, c = 70, v = 100$	6/10	8/10	8/10	9/10
$k = 15, c = 55, v = 75$	13/15	15/15	15/15	15/15
$k = 20, c = 50, v = 40$	15/20	15/20	15/20	17/20
$k = 25, c = 65, v = 80$	20/25	23/25	24/25	24/25
$k = 30, c = 80, v = 45$	26/30	26/30	27/30	27/30
$k = 35, c = 45, v = 75$	30/35	34/35	34/35	34/35
$k = 40, c = 75, v = 60$	35/40	35/40	36/40	36/40
$k = 45, c = 85, v = 50$	38/45	40/45	41/45	41/45
$k = 50, c = 150, v = 150$	34/50	39/50	46/50	47/50
$k = 55, c = 95, v = 110$	44/55	49/55	54/55	54/55
średnia	79.4 %	87.43 %	90.94 %	93.14 %
mediana	82.22 %	88.19 %	91.55 %	92.55 %

Rysunek 7.3: Procentowe pokrycie się zwycięskich kandydatów dla różnych wartości parametru p



Rysunek 7.4: Średnie i mediany pokrycia się zwycięskich komitetów

Uzyskane wyniki pokazują, że już dla wartości parametru $p = 50$ średnio ponad 87% zwycięskich kandydatów wśród obydwu wyników pokryło się. Dla większych wartości parametru p średnie i mediany nieznacznie wzrosły. Dla wartości $p = 50$ pokrycie się wyników poza jednym przypadkiem nie schodziło poniżej 80%. Zwycięzcy kandydaci, którzy nie powtórzyli się w drugim wyniku zazwyczaj stanowili dla parametru p wynoszącego co najmniej 50 około 10%. Te przypadki to opisane wcześniej pary bądź trójki kandydatów, którzy wymieniają się na pozycji zwycięzcy dla dużych wartości parametru p .