



Wydział Matematyki i Nauk Informacyjnych

POLITECHNIKA WARSZAWSKA

Teoria algorytmów i obliczeń

Projekt zaliczeniowy

Piotr Jacak
Jakub Kindracki
Wiktor Kobielski
Ernest Mołczan

Koordynator: prof. dr hab. inż. Władysław Homenda

Semestr zimowy 2025/2026

Spis treści

1 Wstęp	3
2 Definicje pojęć	4
3 Rozmiar multigrafu	5
4 Metryka w zbiorze wszystkich multigrafów	6
5 Minimalne rozszerzenie multigrafu	7
5.1 Algorytm aproksymacyjny do problemu izomorfizmu podgrafu	7
5.1.1 Transformacja multigrafu	7
5.1.2 Aproksymacja algorytmem LeRP	7
6 Bibliografia	10

1 Wstęp

Niniejsza praca stanowi sprawozdanie z projektu zrealizowanego w ramach przedmiotu **Teoria algorytmów i obliczeń**. Przedmiotem badań są algorytmy operujące na multigrafach, ze szczególnym uwzględnieniem problematyki izomorfizmu podgrafów oraz minimalnych rozszerzeń grafów.

Głównym celem projektu jest opracowanie, analiza teoretyczna oraz implementacja algorytmów rozwiązujących dwa ściśle powiązane problemy. Pierwszym z nich jest weryfikacja, czy dany multigraf H jest izomorficzny z podgrąfem multigrafu G . Drugim, kluczowym zagadnieniem, jest wyznaczenie *minimalnego rozszerzenia* multigrafu G do postaci G' , która zawiera co najmniej jeden podgraf izomorficzny z H .

Realizacja powyższych celów wymagała formalnego zdefiniowania oraz uzasadnienia kilku fundamentalnych pojęć. W pracy zaproponowano autorskie lub bazujące na literaturze definicje:

- *rozmiaru multigrafu*,
- *metryki* w zbiorze multigrafów,
- *minimalnego rozszerzenia* multigrafu.

Pojęcia te stanowią podstawę do dalszej analizy algorytmicznej oraz oceny kosztu operacji.

W ramach pracy przeprowadzono analizę złożoności obliczeniowej opracowanych algorytmów. Zgodnie z założeniami projektu, w przypadku gdy algorytmy dokładne charakteryzują się złożonością wykładniczą, przedstawiono również propozycje algorytmów aproksymacyjnych o złożoności wielomianowej.

Niniejszy raport, oprócz formalnych definicji i analizy algorytmów, zawiera także opis przeprowadzonych testów obliczeniowych, dokumentację techniczną implementacji oraz wnioski końcowe.

2 Definicje pojęć

Definicja 1 (Graf). Grafem nazywamy parę $G = (V, E)$, gdzie V jest zbiorem wierzchołków, a $E \subseteq V \times V = \{(u, v) : u, v \in V \wedge u \neq v\}$ jest zbiorem krawędzi. Dla każdej pary wierzchołków $u, v \in V$ istnieje co najwyżej jedna krawędź łącząca wierzchołki u i v .

Definicja 2 (Multigraf). Multigrafem nazywamy graf, w którym pomiędzy dowolnymi dwoma różnymi wierzchołkami $u, v \in V$ może istnieć więcej niż jedna krawędź.

Definicja 3 (Izomorfizm grafów). Dwa grafy $G_1 = (V_1, E_1)$ i $G_2 = (V_2, E_2)$ są izomorficzne, wtedy i tylko wtedy, gdy istnieje bijekcja $f : V_1 \rightarrow V_2$, taka że dla każdej krawędzi $(u, v) \in E_1$ zachodzi $(f(u), f(v)) \in E_2$. Definicja ta jest analogiczna dla multigrafów.

Definicja 4 (Podgraf). Graf $H = (V_H, E_H)$ nazywamy podgrafem grafu $G = (V_G, E_G)$, wtedy i tylko wtedy, gdy $V_H \subseteq V_G$ oraz $E_H \subseteq E_G$. Definicja ta jest analogiczna dla multigrafów.

Definicja 5 (Graf atrybutowy). Graf $G = (V, E, f)$ nazywamy grafem atrybutowym, gdzie V jest zbiorem wierzchołków, a $E \subseteq V \times V = \{(u, v) : u, v \in V \wedge u \neq v\}$ jest zbiorem krawędzi. Dla każdej pary wierzchołków $u, v \in V$ istnieje co najwyżej jedna krawędź łącząca wierzchołki u i v . $f : E \rightarrow \Sigma_E$ jest funkcją, przypisującą etykiety wszystkim krawędziom w grafie G .

Definicja 6 (Macierz sąsiedztwa). Macierzą sąsiedztwa multigrafu $G = (V, E)$ nazywamy macierz A , której pole $A_{uv} = k$, wtedy i tylko wtedy, gdy istnieje k krawędzi $(u, v) \in E$. W przypadku gdy nie istnieje żadna krawędź pomiędzy wierzchołkami u i v , to $A_{uv} = 0$. Dla zwykłych grafów, macierz sąsiedztwa jest macierzą binarną.

3 Rozmiar multigrafu

Definicja 7 (Rozmiar multigrafu). Rozmiarem S multigrafu $G = (V, E)$ nazywamy sumę liczby wierzchołków $|V|$ oraz liczby krawędzi $|E|$ grafu G :

$$S(G) = |V| + |E|$$

Definicja 8 (Porządek w zbiorze wszystkich multigrafów). Niech G_1 i G_2 będą dwoma multigrafami. Mówimy, że G_1 jest mniejszy, lub równy G_2 wtedy i tylko wtedy, gdy suma liczb wierzchołków i krawędzi grafu G_1 jest mniejsza, lub równa sumie liczb wierzchołków i krawędzi grafu G_2 , czyli $S(G_1) \leq S(G_2)$.

Żeby udowodnić poprawność powyższej definicji porządku wykazujemy, że spełnia ona trzy wymagane własności:

- **Zwrotność:**

Dla każdego multigrafu G , $S(G) = S(G)$. Jest to prawda, ponieważ suma liczby wierzchołków i krawędzi multigrafu G jest równa samej sobie.

- **Przechodniość:**

Dla dowolnych multigrafów G_1 , G_2 oraz G_3 , jeśli $S(G_1) \leq S(G_2)$ oraz $S(G_2) \leq S(G_3)$, to $S(G_1) \leq S(G_3)$. Jest to oczywiście prawda.

- **Antysymetryczność:** Dla dowolnych multigrafów G_1 oraz G_2 , jeśli $S(G_1) \leq S(G_2)$ oraz $S(G_2) \leq S(G_1)$, to G_1 jest równy w sensie wcześniejszej zdefiniowanego rozmiaru z G_2 , czyli $S(G_1) = S(G_2)$.

4 Metryka w zbiorze wszystkich multigrafów

5 Minimalne rozszerzenie multigrafu

5.1 Algorytm aproksymacyjny do problemu izomorfizmu podgrafu

Do problemu można zastosować algorytm LeRP (Length-R Paths) opracowany przez Freda W DePiero oraz Davida Krouta [?]. Autorzy bezpośrednio stwierdzają, że nie zajmują się multigrafami. Zaznaczają natomiast, że multigrafy nie są ograniczeniem, ponieważ można stworzyć reprezentację multigrafu opisującą istnienie równoległych krawędzi za pomocą schematu kolorowania krawędzi. Zatem kompletny algorytm aproksymacyjny będzie składał się z dwóch kroków:

1. Redukcja: Transfomacja multigrafów wejściowych G_1 oraz G_2 na grafy atrybutowe G'_1 oraz G'_2 .
2. Aproksymacja: Wykorzystanie algorytmu LeRP (który obsługuje grafy atrybutowe) do stwierdzenia, czy graf G_1 jest izomorficzny z pewnym podgrafem P grafu G_2 .

5.1.1 Transformacja multigrafu

Niech G_1 i G_2 będą multigrafami. Odpowiadające im grafy atrybutowe G'_1 i G'_2 konstruowane są w następujący sposób:

- Zbiory wierzchołków pozostają bez zmian ($V'_1 = V_1, V'_2 = V_2$).
- Dla każdej pary wierzchołków (u, v) w G_1 (i analogicznie w G_2): Jeśli między u a v w G_1 istnieje k równoległych krawędzi, to w G'_1 tworzona jest pojedyncza krawędź (u, v) z atrybutem $k \in \mathbb{N}^+$.

5.1.2 Aproksymacja algorytmem LeRP

Fundamentalna zasada LeRP różni się od algorytmów backtrackingu (przykładowo algorytm Ullmanna). Zamiast próbować dopasować pełną strukturę sąsiedztwa krok po kroku, LeRP opiera się na założeniu, że o podobieństwie strukturalnym dwóch wierzchołków można wnioskować na podstawie porównania liczby ścieżek (*sgnatur*) o różnej długości (r - parametr algorytmu) w ich sąsiedztwie. Działanie algorytmu LeRP można podzielić na kilka etapów, realizowanych na grafach atrybutowych G'_1 i G'_2 .

1. Obliczanie liczby ścieżek (pre-processing)

Dla obu grafów G'_1 i G'_2 obliczane są potęgi ich macierzy sąsiedztwa, odpowiednio A^r i B^r aż do maksymalnej długości R . Wartość A_{ij}^r w macierzy A^r reprezentuje liczbę ścieżek o długości dokładnie r z wierzchołka i do wierzchołka j . W kontekście omówionej transformacji, macierz A jest macierzą, gdzie A_{ij} przechowuje atrybut k - liczbę równoległych krawędzi wcześniejszego multigrafu G_1 . Obliczenie A^r polega na r -krotnym mnożeniu macierzy.

2. Porównanie strukturalne

Dla każdej pary wierzchołków $g_{1i} \in G'_1$ i $g_{2k} \in G'_2$ algorytm porównuje ich struktury: sprawdza, do jakiej długości ścieżki ($r = 1, \dots, R$) grafy wyglądają tak samo. Im dłużej są podobne, tym wyższy wynik podobieństwa r_{max} .

3. Mapping

Algorytm zaczyna od pojedynczego ziarna mapowania - wstępnie przypisuje wierzchołek g_{1i} do wierzchołka g_{2k} . Następnie iteracyjnie dodaje do tego mapowania sąsiadów już zamapowanych wierzchołków, wybierając te, które maksymalizują wskaźnik podobieństwa i są spójne ze znalezionymi wcześniej.

LeRP opiera się na założeniu, że dopasowanie sygnatur ścieżek (taka sama liczba ścieżek o długości $1, 2, \dots, R$ jest wystarczającym dowodem podobieństwa, aby utworzyć mapowanie.

Złożoność pesymistyczna algorytmu LeRP jest wielomianowa i wynosi $O(N^3 \cdot D^2 \cdot R)$, gdzie:

- N to liczba wierzchołków w grafach (zakładając, że oba grafy mają rozmiar rzędu N).
- D to średni stopień wierzchołków w grafach.
- R to maksymalna długość ścieżki brana pod uwagę.

Uzasadnienie złożoności: porównanie par wierzchołków wymaga N^2 porównań. Każdy wierzchołek ma średnio D sąsiadów, więc daje to D^2 dodatkowych operacji. Analiza różnych długości ścieżek to czynnik R . Podczas budowania dopasowania, algorytm jeszcze raz przechodzi po wszystkich kandydatach, a więc N jest kolejnym czynnikiem.

W kontekście tego algorytmu aproksymacyjnego nie rozważano formalnego dowodu

poprawności. Dostarczono empiryczną gwarancję jakości - algorytm testowany na dużych zbiorach danych, wykazując, że algorytm konsekwentnie zwraca wyniki bliskie optimum w praktyce.

6 Bibliografia