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

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 podgrafem 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 $f : V_1 \to V_2$, 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. $f : E \to \Sigma_E$ 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:

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

Zakładamy, ze liczby wierzchołków i krawędzi są zapisanymi wcześniej stałymi, więc obliczenie rozmiaru multigrafów jest operacją o złożoności czasowej $O(1)$.

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 wcześniej zdefiniowanego rozmiaru z G_2, czyli $S(G_1) = S(G_2)$.

4 Metryka w zbiorze wszystkich multigrafów

Definicja 9 (Metryka w zbiorze multigrafów)**.** Niech $\mathcal{G}$ będzie zbiorem wszystkich multigrafów. **Metryką** w zbiorze $\mathcal{G}$ nazywamy funkcję:

$$d : \mathcal{G} \times \mathcal{G} \to \mathbb{N}_0$$

Wartość $d(G_1, G_2)$ nazywamy **odległością** między multigrafami G_1 i G_2, a definiujemy ją, jako **minimalną** liczbę operacji dodawania lub usuwania pojedynczej krawędzi lub wierzchołka, za pomocą których można przekształcić graf G_1 w graf izomorficzny z G_2.

Powyższa definicja spełnia następujące własności metryki:

- **Identyczność nierozróżnialnych**:
  Dla dowolnych multigrafów G_1 oraz G_2, $d(G_1, G_2) = 0$ wtedy i tylko wtedy, gdy G_1 jest izomorficzny z G_2. Wynika to bezpośrednio z definicji naszej metryki.

- **Symetria**:
  Dla dowolnych multigrafów G_1 oraz G_2, $d(G_1, G_2) = d(G_2, G_1)$. Dodawanie i usuwanie krawędzi lub wierzchołków jest operacją odwracalną, więc liczba operacji potrzebnych do przekształcenia G_1 w G_2 jest równa liczbie odwrotnych operacji potrzebnych do przekształcenia G_2 w G_1.

- **Nierówność trójkąta**:
  Dla dowolnych multigrafów G_1, G_2 oraz G_3, $d(G_1, G_3) \leq d(G_1, G_2) + d(G_2, G_3)$. Oznacza to, że najkrótsza droga między dwoma multigrafami nie może być dłuższa niż droga przechodząca przez trzeci multigraf. Jest to prawda, ponieważ każda sekwencja operacji przekształcających G_1 w G_2 oraz G_2 w G_3 może być złożona w jedną sekwencję przekształcającą G_1 w G_3.

5 Minimalne rozszerzenie multigrafu

6 Aproksymacyjne minimalne rozszerzenie multigrafu

Do problemu można zastosować pewną modyfikację algorytmu LeRP (Length-R Paths), opracowanego przez Freda W DePiero oraz Davida Krouta [1]. Algorytm opiera się na założeniu, że o podobieństwie strukturalnym dwóch wierzchołków można wnioskować na podstawie porównania liczby ścieżek (*sygnatur*) o długości r w ich sąsiedztwie.

Modyfikacja algorytmu jako argumenty przyjmuje dwa multigrafy G_1 i G_2, maksymalną długość ścieżki R oraz liczbę szukanych kopii grafu G_1 - k. Im większa długość ścieżki R, tym algorytm jest dokładniejszy. Zwraca natomiast przekształcenie $f(g_{1i}) = g_{2j}$, które opisuje najlepszy (w rozumieniu aproksymacji) wspólny podgraf G_1 oraz G_2.
Oznaczmy przez H najlepszy wspólny podgraf G_1 oraz G_2. Do multigrafu H należy dołożyć zbiór krawędzi X, tak aby grafy G_1 oraz $H' = (V_H, E_H \cup X)$ były izomorficzne. Wówczas multigraf $G_3 = (V_2, E_2 \cup X)$ będzie minimalnym rozszerzeniem G_2, aby ten zawierał G_1 jako podgraf. Następnie usuwamy z początkowego grafu G_2 wierzchołki podgrafu H i powtarzamy proces, aby znaleźć k rozdzielnych kopii G_1 jako podgrafu G_2.

6.1 Opis algorytmu

Algorytm można podzielić na kilka etapów.

1. W pierwszym kroku, należy wykonać transformację multigrafów wejściowych G_1 oraz G_2 na grafy atrybutowe G'_1 oraz G'_2. Transformacja przebiega 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 $k \in \mathbb{N}^+$.
2. W etapie drugim, 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^r_{ij} w macierzy A^r reprezentuje liczbę ścieżek o długości dokładnie r z wierzchołka i do wierzchołka j (na ścieżkach mogą się powtarzać wierzchołki i krawędzie). W kontekście omówionej transformacji, macierz A jest macierzą, gdzie A_{ij} przechowuje atrybut k - liczbę równoległych krawędzi między wierzchołkami i i j w multigrafie G_1.

3. Każdy wierzchołek $g_{1i} \in G'_1$ można porównać z każdym wierzchołkiem $g_{2k} \in G'_2$, stosując przykładowo podobieństwo cosinusowe między histogramem wartości w wierszu macierzy A_i^r a histogramem wartości w wierszu macierzy B_i^r. Następnie tworzymy macierz z wartościami podobieństw między wierzchołkami. Wierzchołki najbardziej podobne zostają ziarnem mapowania.

4. Następnie iteracyjnie, algorytm porównuje sąsiadów wierzchołków g_{1i} i g_{2k}. Do następnego mapowania, wybiera tych sąsiadów, do których liczba ścieżek o długości co najwyżej R jest największa i jest równa dla obu wierzchołków z dwóch grafów. Algorytm sprawdza również, czy wybrane wierzchołki nie zostały już zmapowane.

5. Zmapowane wierzchołki w grafie G'_2 są usuwane, tworząc nowy graf G''_2 i proces jest powtarzany dla grafów G'_1 oraz G''_2. W ten sposób znajdowane jest k rozdzielnych kopii.

Złożoność pesymistyczna algorytmu LeRP (i omówionej modyfikacji jest wielomianowa i wynosi $O(N^3 \cdot D^2 \cdot R \cdot k)$, 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.
- k to liczba szukanych kopii w minimalnym rozszerzeniu

Uzasadnienie złożoności: porównanie par wierzchołków wymaga N^2 porównań. Każdy wierzchołek ma średnio D sąsiadów, więc porównywanie sąsiadów daje 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. Czynnik k odpowiada za znalezienie k kopii mniejszego multigrafu G_1.

W kontekście tego algorytmu aproksymacyjnego nie rozważano formalnego dowodu poprawności. Dostarczono empiryczną gwarancję jakości - algorytm testowano na dużych zbiorach danych, wykazując, że algorytm konsekwentnie zwraca wyniki bliskie optimum w praktyce.

7 Bibliografia

Literatura

[1] Fred DePiero and David Krout. An algorithm using length-r paths to approximate subgraph isomorphism. *Pattern Recognition Letters*, 24(1):33–46, 2003.