

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



Sieci złożone

Sprawozdanie z laboratorium

AUTOR

Piotr Matys

nr albumu: **260337**

kierunek: **Sztuczna Inteligencja**

16 kwietnia 2024

Streszczenie

Celem pracy jest zbadanie wspólnych hipotetycznych relacji pomiędzy pięcioma najpopularniejszymi dietami bazując wyłącznie na produktach, które je łączą. Dane do badań zostały pobrane metodą ekstrakcji danych ze stron internetowych - linki do poszczególnych stron znajdują się w repozytorium GitHub. Analizę przeprowadzono z wykorzystaniem metody projekcji grafu dwudzielnego. Stosując metodę projekcji ważonej, Newmana oraz Jaccarda udało się ustalić, że najbardziej uwidaczniające się relacje występują pomiędzy dietą śródziemnomorską, a fleksitariańską. W dodatku zbadano też, że nie da się jednoznacznie ustalić, które diety są ze sobą powiązane w najmniejszym stopniu.

1 Wstęp – sformułowanie problemu

Każda z diet charakteryzuje się określoną listą produktów, które należy spożywać podczas jej stosowania. Część z nich pokrywa się dla różnych diet, dlatego w celu zbadania wspólnych relacji między dietami, utworzono sieć powiązań pomiędzy nimi, a produktami, bazując na danych zebranych ze stron internetowych o danych dietach. Podjęta zostanie próba wywnioskowania z otrzymanej sieci informacji, które z diet są najsilniej powiązane ze sobą, a które najmniej.

2 Opis rozwiązania

Dane o produktach, którymi cechują się odpowiednie diety zostały pobrane ze stron, których linki znajdują się w repozytorium GitHub w pliku tekstowym <https://github.com/MatPatCarry/Diets/blob/main/diety-urls.txt>. W badaniu wzięto pod uwagę 5 najpopularniejszych diet: śródziemnomorską, ketogeczną, wegańską, dash oraz fleksitariańską. Dane zostały pozyskane metodą ekstrakcji danych ze stron internetowych, wykorzystując bibliotekę *BeautifulSoup* dla języka programowania Python. W celu oczyszczenia i odpowiedniego sformatowania danych, zostały wykorzystane autorskie funkcje napisane w języku Python. Sformatowane dane zostały zapisane w pliku tekstowym w postaci dwuelementowych krotek. Plik `dieta-produkt.txt` <https://github.com/MatPatCarry/Diets/blob/main/dieta-produkt.txt> zawiera 336 takich krotek, których pierwszym elementem jest nazwa diety, a drugim nazwa produktu, który się spożywa podczas stosowania takiej diety. Powiązania diet i produktów stworzyły graf dwudzielny, w którym powiązania występują wyłącznie pomiędzy wierzchołkami różnych typów - w badanym przypadku jest to graf skierowany, gdzie powiązanie jest skierowane od diety do produktu. W dalszych rozważaniach i badaniach przyjęto, że wykorzystane zostaną wyłącznie te produkty, które są powiązane z co najmniej dwoma dietami. W celu wizualizacji takiej sieci wykorzystano bibliotekę *networkx* i jej funkcję do wizualizacji grafów dwudzielnych *bipartite_layout()*. Następnie w celu wyznaczenia hipotetycznych relacji pomiędzy dietami, zastosowano metodę projekcji grafu dwudzielnego. Metoda projekcji wierzchołków górnych (produkty) na wierzchołki dolne (diety) wykorzystuje założenie o ich homofili, czyli domniemanej bliskości wynikającej z posiadania krawędzi prowadzących do wspólnych wierzchołków górnych. Zastosowano metody projekcji ważonej, Newmana i Jaccarda - pominięto projekcję prostą, która nie zwraca informacji o sile powiązań, a wyłącznie informację, czy dwa wybrane wierzchołki dolne mają wspólnego sąsiada. Wykorzystana projekcja ważona zwraca graf wynikowy z wagami odpowiadającymi intensywności powiązań między wierzchołkami, gdzie waga odpowiada liczbie wspólnych wierzchołków górnych sąsiadujących z połączonymi wierzchołkami dolnymi. W przypadku metody projekcji Newmana waga pomiędzy wierzchołkami dolnymi reprezentuje *aktywności*, w których wspólnie uczestniczą obiekty. Brana jest pod uwagę siła domniemanych powiązań - jest ona

tym mniejsza im więcej diet charakteryzuje się spożywaniem tego produktu. Zastosowano również metodę projekcji Jaccarda, która wyznaczyła, jak bardzo spożywane produkty badanych diet się pokrywają - im bardziej się pokrywają, tym ta wartość jest bliższa 1. Każda z zastosowanych metod projekcji została użyta przy pomocy biblioteki *networkx* i jej modułu dla grafów dwudzielnych *bipartite*.

3 Rezultaty obliczeń

3.1 Plan badań

Analizę sieci rozpoczęto od wizualizacji grafu dwudzielnego diet i produktów. Następnie zastosowano kolejno metodę projekcji ważonej, Newmana i Jaccarda, wykorzystując odpowiednie funkcje biblioteki *networkx*. W przypadku projekcji ważonej i Newmana, wagę powiązań znormalizowano, resztę argumentów funkcji pozostawiono domyślnych. Dwukrotnie sprawdzono, czy otrzymane wyniki są powtarzalne dla kolejnych uruchomień programu.

3.2 Wyniki

Na rys. 1 pokazana jest wizualizacja grafu dwudzielnego.

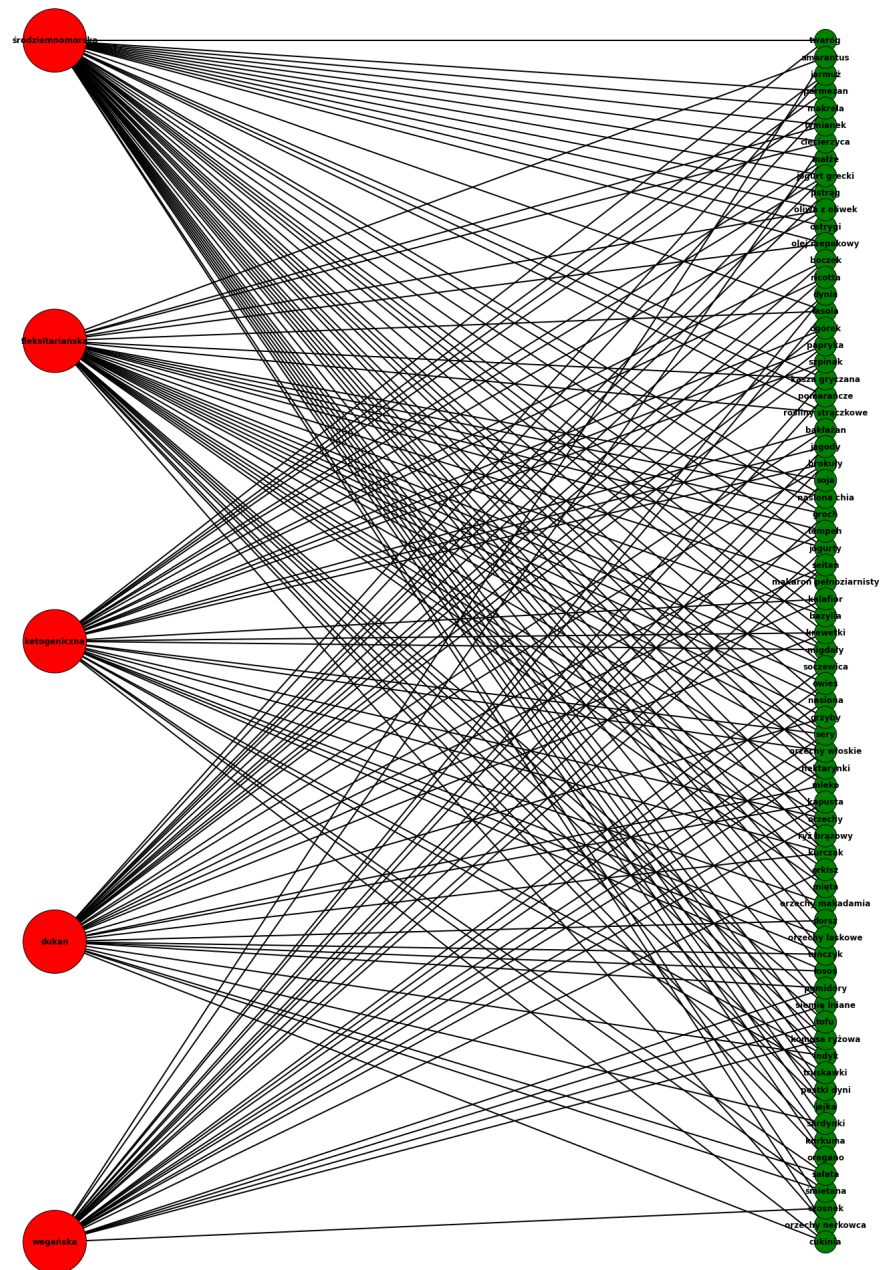
Rezultaty poszczególnych metod projekcji wierzchołków dolnych zostały zwizualizowane za pomocą macierzy, gdzie intensywność koloru zmienia się wraz z wagą połączenia. W każdej z zastosowanych metod projekcji, największą wartość wagi połączenia charakteryzuje się połączenie łączące wierzchołek diety śródziemnomorskiej z fleksitariańską. W przypadku metody projekcji ważonej - rys. 2 i Newmana - rys. 3 najmniejszą wagę ma powiązanie diet wegańskiej i ketogenicznej, a w projekcji Jaccarda - rys. 4 diety dukan i fleksitariańskiej. Niepoprawną metodą byłoby dokładne porównywanie wartości wag dla różnych metod projekcji, jednak aby było to możliwe, w tym celu posłużono się oceną proporcji poszczególnych wag - są one dobrze wizualizowane przez intensywność kolorów. W przypadku projekcji Jaccarda - rys. 4 w większym stopniu (porównywano stosunek wartości danej wagi do wartości wagi maksymalnej danej projekcji) uwidoczniła się relacja pomiędzy dietą dukan, a dietą ketogeniczną oraz dietą fleksitariańską, a wegańską, niż w przypadku pozostałych zastosowanych metod projekcji. W każdej zastosowanej metod projekcji wagi pomiędzy dietą fleksitariańską, a dietami dukan i ketogeniczną charakteryzowały się najmniejszą wartością. Dla wszystkich zastosowanych metod wartość wag pomiędzy dietą ketogeniczną, a dietami śródziemnomorską i dukan były powyżej średniej arytmetycznej wagi z wartością najwyższą i najniższą.

4 Wnioski

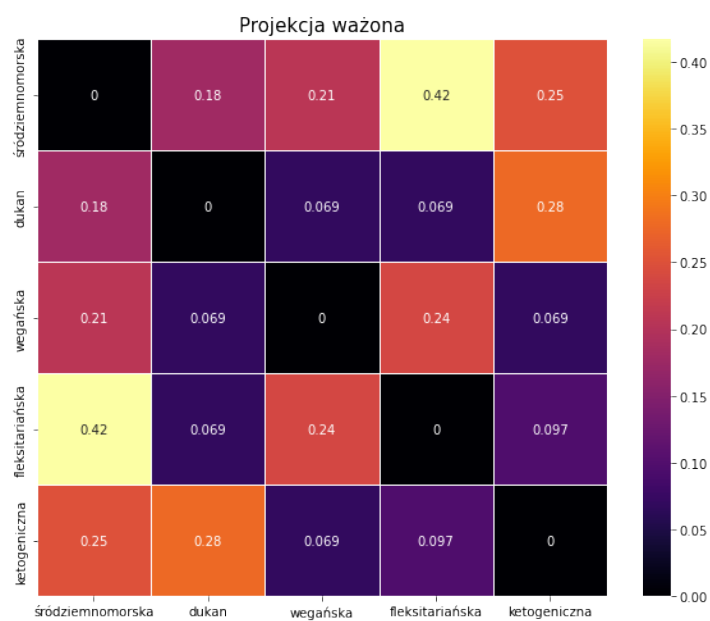
Najwyższe wartości wag dla każdej z zastosowanych metod projekcji wskazują na fakt, że najsilniejszą relację wykazuje dieta śródziemnomorska z fleksitariańską. Wykorzystują one największą liczbę wspólnych produktów, a produkty te dodatkowo są również najbardziej unikalne dla tych dwóch diet. Nie da się jednoznacznie stwierdzić, które dwie diety są ze sobą najsłabiej powiązane. Wyznaczono, że najsłabiej powiązane są ze sobą pary diet dukan i wegańskiej, fleksitariańskiej i dukan, wegańskiej i ketogenicznej oraz ketogenicznej i fleksitariańskiej.

A Dodatek

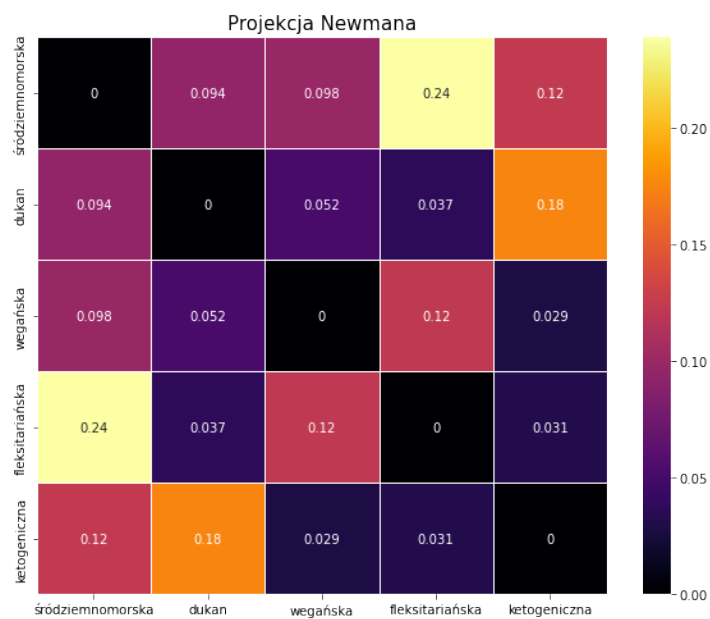
Kody źródłowe umieszczone zostały w repozytorium github:
<https://github.com/MatPatCarry/Diets>.



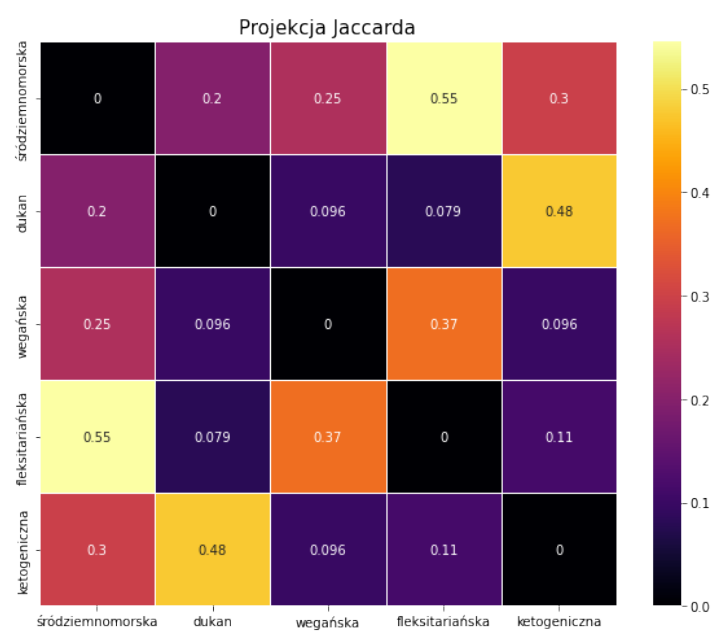
Rysunek 1: Graf dwudzielny diet i produktów



Rysunek 2: Macierz wag po zastosowaniu projekcji ważonej



Rysunek 3: Macierz wag po zastosowaniu projekcji Newmana



Rysunek 4: Macierz wag po zastosowaniu projekcji Jaccarda