

Wojskowa Akademia Techniczna



Temat: Analiza sieci dróg na terenie uczelni WAT

Przedmiot: Analiza sieci społecznościowych

Autorzy:
Radosław Relidzyński
Maciej Kawka

Prowadzący:
mgr inż. Marcin Perka

Contents

I.	Pomysł na analizę	3
II.	Dobór narzędzi.....	5
III.	Budowa środowiska badawczego	5
1.	Opis środowiska badawczego.....	5
2.	Reprezentacja danych i narzędzia analityczne.....	6
3.	Plan analizy	6
IV.	Pozyskiwanie danych	7
V.	Eksperymenty.....	7
VI.	Podsumowanie.....	13

I. Pomysł na analizę

Tematem pracy jest analiza jakości połączeń drogowych na terenie Wojskowej Akademii Technicznej w porównaniu do innych terenów.

W ramach analizy wykorzystane zostaną dane z OpenStreetMap.

Na podstawie pozyskanych danych utworzone zostaną grafy, gdzie wierzchołki będą reprezentować wszelkie skrzyżowania, a krawędzie to będą drogi (wraz z informacją o ich nazwie oraz długości). W ramach tak utworzonych grafów będzie przeprowadzona analiza.

Analizę oprzemy o najważniejsze cechy które muszą spełniać drogi komunikacyjne, którymi są:

- Dostępność i sprawność komunikacyjna – sieć powinna umożliwiać szybki i łatwy dostęp do wszystkich punktów obszaru, co oznacza że powinny być możliwie krótkie ścieżki między węzłami i brak ślepych zaułków czy izolowanych fragmentów.
- Lokalna gęstość i elastyczność przejazdów – użytkownicy powinni mieć wiele alternatywnych tras do wyboru na poziomie lokalnym, czyli dużo skrzyżowań i połączeń między punktami, oraz możliwość wykorzystania objazdów
- Odporność na zakłócenia i awarie – sieć powinna być odporna na utratę kluczowych elementów, np. przez korki, remonty czy awarie. Usunięcie kilku węzłów nie powinno prowadzić do rozpadu sieci, ponieważ będą istnieć alternatywne przejazdy.

W związku z powyższymi, dla sieci zostaną przeprowadzone obliczenia mające na celu przedstawienie ich podstawowych cech w postaci wskaźników takich jak:

- 1) Średni stopień wierzchołka - Średnia liczba krawędzi wychodzących z wierzchołków. Pokazuje, jak bardzo rozgałęziona jest sieć – im więcej dróg na skrzyżowaniu, tym lepiej.
- 2) Średnia długość najkrótszej ścieżki - Średnia liczba kroków między dowolnymi parami węzłów w grafie (obliczana po najkrótszych ścieżkach). Określa efektywność przemieszczania się po sieci – im mniejsza, tym szybsze dojazdy.
- 3) Współczynnik klasteryzacji - Miara lokalnej spójności – ile z sąsiadów danego węzła jest również ze sobą połączonych. Pokazuje, czy istnieją alternatywne trasy lokalne – pętle, objazdy, „obwody”.
- 4) Centralność pośrednictwa (Betweenness) - Ile najkrótszych ścieżek przechodzi przez dany węzeł – miara „ruchliwości” punktu w grafie. Wskazuje, które skrzyżowania są kluczowe dla ruchu (i ewentualnie przeciążone).

- 5) Centralność bliskości (Closeness) - Odwrotność sumy odległości od danego węzła do wszystkich innych – im bliżej wszystkich, tym wyższa wartość. Pokazuje, jak łatwo dotrzeć z danego miejsca do reszty sieci – miara dostępności.
- 6) Centralność wektorowa (Eigenvector) - Mierzy „jakość” połączeń – węzeł ma wysoką wartość, jeśli łączy się z innymi ważnymi węzłami. Wskazuje główne centra komunikacyjne – miejsca strategiczne w strukturze sieci.
- 7) Gęstość sieci - Stosunek liczby istniejących krawędzi do maksymalnej możliwej liczby w grafie nieskierowanym. Pokazuje, jak „pełna” jest sieć – czy mamy dużo połączeń czy raczej luźną strukturę.

Wskaźniki 1, 3 i 6, czyli odpowiednio: średni stopień wierzchołka, współczynnik klasteryzacji oraz centralność wektorowa, będą wykorzystywane zarówno do porównań między różnymi sieciami, jak i do analizy jednej sieci w ramach eksperymentów. Są to wskaźniki niezależne od rozmiaru sieci – odzwierciedlają wyłącznie strukturę i lokalną organizację połączeń.

Z kolei wskaźniki 2, 4, 5 i 7 – czyli średnia długość najkrótszej ścieżki, centralność pośrednictwa, centralność bliskości oraz gęstość – są wrażliwe na skalę sieci, dlatego ich surowe wartości będą wykorzystywane wyłącznie przy analizie tej samej sieci w różnych warunkach eksperymentalnych.

Jednak spośród wskaźników wrażliwych na skalę, wybrane zostały trzy, które po zastosowaniu odpowiednich metod normalizacji (logarytmicznej dla długości ścieżki, względnej dla centralności betweenness i closeness) mogą zostać wykorzystane również do porównań między sieciami o różnym rozmiarze. Będą to zatem: znormalizowana średnia długość najkrótszej ścieżki, centralność pośrednictwa (normalized betweenness) oraz centralność bliskości.

Taka analiza zostanie przeprowadzona dla każdej z sieci oraz dla przeprowadzonych na nich eksperymentów. Jako że w sieci dróg jedną z najważniejszych cech jest odporność na zakłócenia i awarie (wynikające np. z wypadków) to zostaną przeprowadzone następujące eksperymenty:

- a) Badanie odporności na utratę wierzchołków o największym stopniu, czyli najlepiej połączonych skrzyżowań (wartość degree)
- b) Badanie odporności na utratę wierzchołków o najczęstszych wystąpieniach na najkrótszych trasach (wartość betweenness)
- c) Badanie odporności na utratę wierzchołków o najszybszej dostępności do innych punktów (wartość closeness)
- d) Badanie odporności na utratę wierzchołków o największym wpływie w sieci (wartość eigenvector)

II. Dobór narzędzi

Poniżej znajduje się lista narzędzi, które zostaną wykorzystane do budowy środowiska:

1. Język Python

Ze względu na liczbę dostępnych bibliotek wspomagających analizę sieci grafowych wybór tego języka pozwoli na szybką i skuteczną implementację wybranych narzędzi.

2. Biblioteka networkx

Jest to biblioteka do tworzenia oraz zarządzania strukturami grafowymi. Pozwala na tworzenie grafów różnego rodzaju, walidację poprawności ich tworzenia oraz dostarcza gotowych narzędzi do ich analizy.

3. Biblioteka osmnx

Jest to wyspecjalizowana biblioteka do pobierania i analizowania danych przestrzennych z OpenStreetMap, umożliwia budowanie grafów drogowych, ścieżek pieszych, rowerowych i transportu publicznego. Zapewnia bezpośrednią integrację z networkx.

4. Biblioteka matplotlib

Jedna z najpopularniejszych bibliotek do tworzenia wykresów i wizualizacji danych. Dzięki niej będzie można wyświetlić graf, a posiadając informację o położeniach geograficznych punktów będzie można nawet wyświetlić rzeczywiste położenie skrzyżowań między nimi.

III. Budowa środowiska badawczego

1. Opis środowiska badawczego

W ramach środowiska badawczego przygotowane zostaną trzy niezależne zbiory danych drogowych, reprezentujące trzy różne typy układów urbanistycznych:

1. Układ organiczny – wybrany fragment dzielnicy Wola w Warszawie, charakteryzujący się naturalnym, nieregularnym rozwojem urbanistycznym.
2. Układ hybrydowy – sieć dróg znajdująca się na zamkniętym terenie Wojskowej Akademii Technicznej (WAT) w Warszawie, która łączy elementy planowania i adaptacji do istniejącej przestrzeni.
3. Układ planowy – fragment dzielnicy Eixample w Barcelonie, będącej klasycznym przykładem siatki ulic projektowanej od XIX wieku, zgodnie z zasadami modernistycznego urbanizmu.

Dane dla wszystkich trzech lokalizacji zostaną pozyskane za pomocą funkcji eksportu z serwisu OpenStreetMap (OSM), który umożliwia zaznaczenie dowolnego obszaru oraz pobranie jego struktury drogowej w formacie .osm.

2. Reprezentacja danych i narzędzia analityczne

Pozyskane dane zostaną przekształcone w postać grafów, w których:

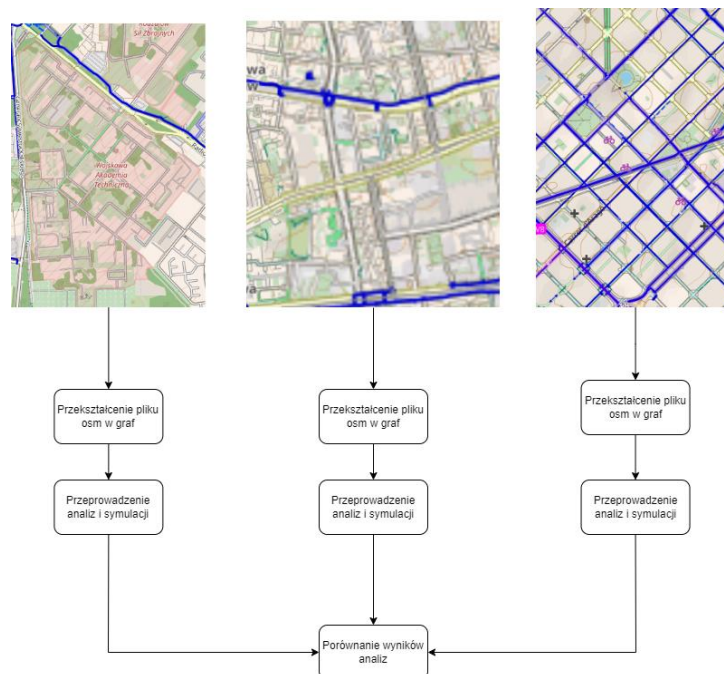
- wierzchołki (nodes) reprezentują skrzyżowania lub końce dróg,
- krawędzie (edges) odpowiadają odcinkom ulic między węzłami.

Do reprezentacji i przetwarzania grafów wykorzystana zostanie biblioteka networkx w języku Python, która oferuje rozbudowane możliwości analizy struktur grafowych oraz wydajne operacje obliczeniowe. Przetwarzanie i wczytywanie danych .osm umożliwi biblioteka osmnx, natywnie zintegrowana z networkx, specjalizująca się w analizie sieci przestrzennych.

Do wizualizacji wykorzystane będą narzędzia takie jak matplotlib, natomiast do przeprowadzania właściwych analiz zostaną zastosowane metody i algorytmy dostępne w bibliotekach analitycznych oraz środowisku osmnx.

3. Plan analizy

Na rysunku 1 przedstawiono ogólny schemat planowanej analizy. Dla wszystkich trzech lokalizacji dane zostaną przetworzone w jednakowy sposób, co pozwoli na spójne i rzetelne porównanie struktur drogowych o różnym charakterze planowania.



Rys. 1. Plan analizy

IV. Pozyskiwanie danych

Dane zostaną pobrane z OpenStreetMap dostępne na tej stronie:

<https://www.openstreetmap.org/export>

Możliwe są zarówno do pobrania danych w zakresie całego kontynentu jak i pojedynczych miast czy wybranych terenów. Nas przede wszystkim będzie interesować funkcjonalność pozwalająca wydzielić prostokątny obszar i pobrać dane o połączeniach tylko z tego obszaru – tak uda się pozyskać informacje o połączeniach na terenie uczelni.

Dla porównania zostaną pobrane dane z innych obszarów, aby porównać jakość sieci dróg na terenie uczelni z innymi drogami na terenie Polski.

V. Eksperymenty

W Tab. 1. Przedstawione zostały wyniki obliczeń współczynników dla 3 podanych lokalizacji.

Wskaźnik	WAT	Wola	Barcelona
Avg Degree	2.635294	2.742251	2.877515
Avg Shortest Path	22.34566	27.00241	43.57902
Normalized Avg Shortest Path	2.901752	3.10773	4.898397
Clustering Coefficient	0.043183	0.047232	0.02375
Avg Betweenness (normalized)	0.009667	0.004382	0.005829
Avg Closeness	0.045823	0.038182	0.02389
Avg Eigenvector	0.004262	0.001687	0.001296
Largest CC Size	2210	5936	7307
Largest CC % of all nodes	1	1	1
Density	0.001193	0.000462	0.000394

Tab. 1. Wyniki obliczeń podstawowych metryk dla 3 lokalizacji

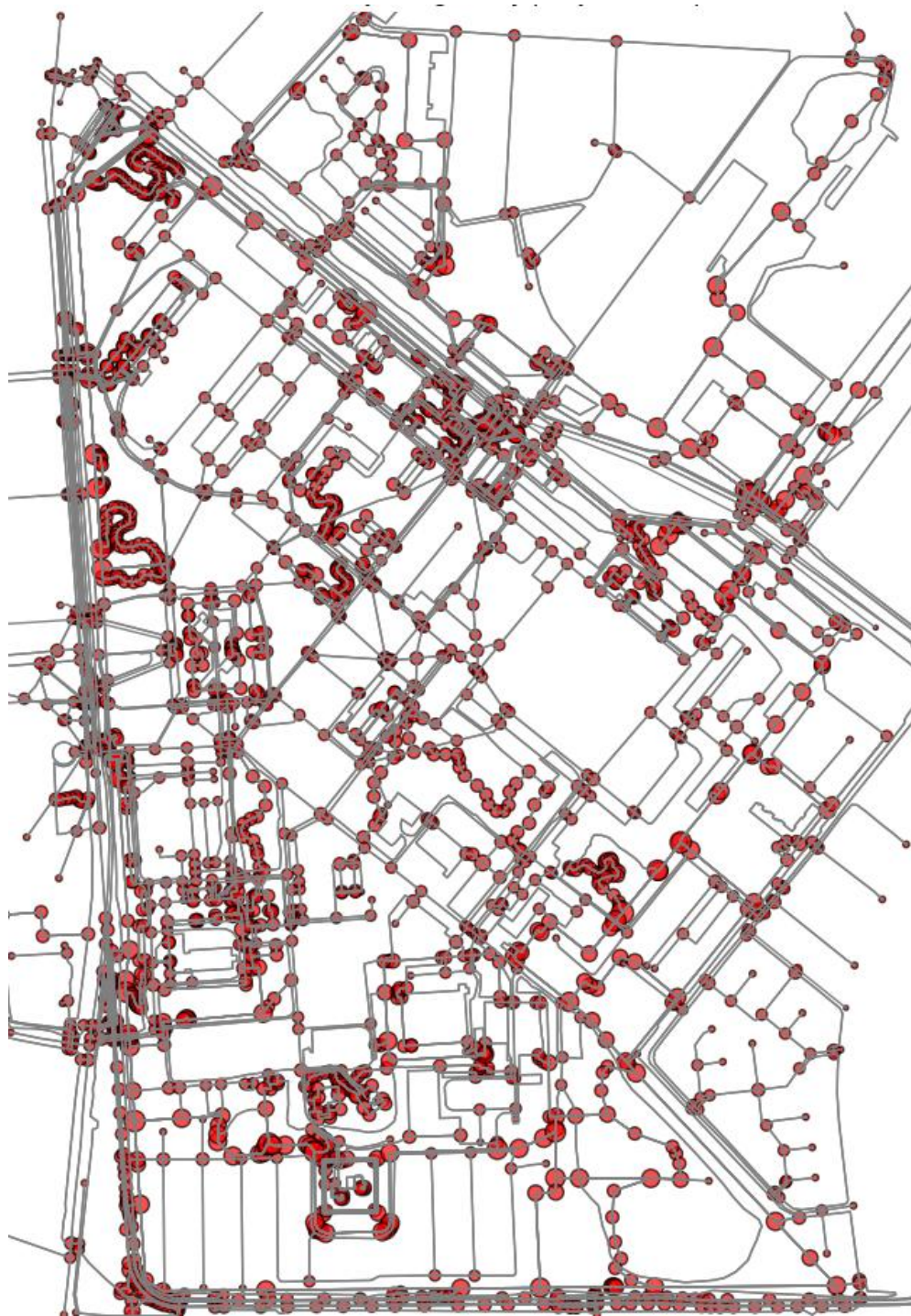
Analiza trzech sieci drogowych – kampusu WAT, warszawskiej dzielnicy Wola oraz dzielnicy Eixample w Barcelonie – została przeprowadzona z uwzględnieniem zarówno surowych, jak i znormalizowanych wskaźników struktury grafowej. Dzięki temu możliwe było sprawiedliwe porównanie sieci o różnych rozmiarach.

Sieć WAT, jako najmniejsza i najbardziej zwarta, wyróżnia się najkrótszą znormalizowaną średnią ścieżką (2.90), najwyższą centralnością bliskości (0.0458) oraz największą gęstością (0.00119), co potwierdza jej wysoką efektywność komunikacyjną. Ponadto, osiąga najwyższy wynik centralności wektorowej, co wskazuje na wyraźne lokalne centra komunikacyjne w strukturze kampusu.

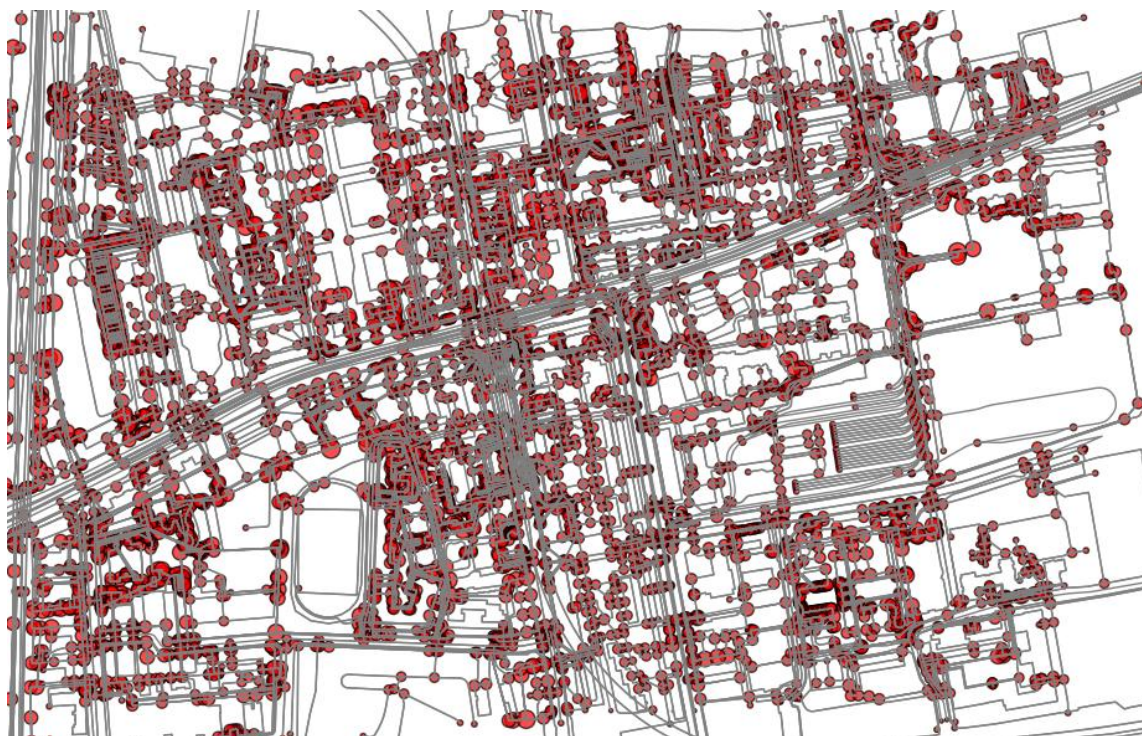
Wola, jako sieć średniej wielkości, uzyskała najwyższy współczynnik klasteryzacji (0.0472), co oznacza, że jest najbardziej rozgałęziona lokalnie – posiada największą liczbę objazdów i połączeń sąsiednich. Pozostałe wartości, w tym średni stopień wierzchołka (2.74) oraz znormalizowana długość ścieżki (3.11), wskazują na zrównoważony układ drogowy typowy dla rozbudowującej się dzielnicy miejskiej.

Z kolei Barcelona, mimo planowego, siatkowego układu ulic, wykazuje najwyższą średnią długość ścieżki (43.57) i najniższą centralność bliskości (0.0239), co jest naturalną konsekwencją największego rozmiaru sieci. Dodatkowo, uzyskano najniższy współczynnik klasteryzacji (0.0237) i najniższą gęstość (0.00039), co oznacza mniejszą lokalną spójność oraz rzadsze połączenia między punktami węzłowymi.

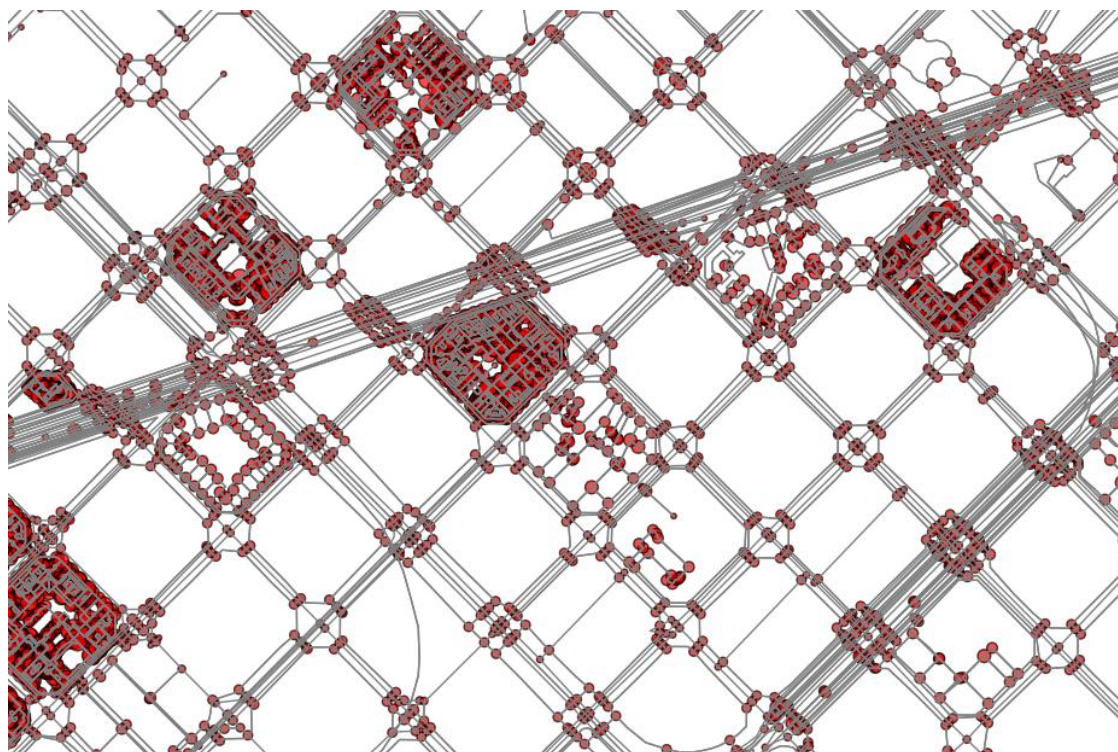
Poniżej przedstawione są również analizy wizualne, gdzie węzły są tym większe im wyższy jest stopień pośrednictwa wierzchołka



Rys. 2. Wizualizacja WAT.



Rys. 3. Wizualizacja Wola.



Rys. 4. Wizualizacja Barcelona.

Z powyższych ilustracji wynika że najbardziej uporządkowaną siecią jest sieć ulic w dzielnicy Barcelony, najmniej natomiast woli. Ulice na kampusie WAT są w miarę uporządkowane i w sposób logiczny rozmieszczone.

Analiza odporności jakości połączeń na utratę skrzyżowań:

WAT:

Rodzaj grafu	Śr. Stopień	Śr. Pośrednictwo	Śr. Bliskość	Śr. Centralność wektorowa	Gęstość grafu
Bazowy	2.6353	0.0097	0.0458	0.0043	0.0012
Usunięty stopień	2.5859	0.0095	0.0425	0.0039	0.0012
Usunięte pośrednictwo	2.6042	0.0116	0.0385	0.0050	0.0012
Usunięta bliskość	2.6115	0.0104	0.0427	0.0040	0.0012
Usunięty wektor własny	2.6152	0.0100	0.0445	0.0040	0.0012

Analiza:

1. Usunięte wierzchołki największego stopnia - niewielki spadek średniego stopnia – 1% wierzchołków o największym stopniu nie ma znacznie większej tej wartości od innych
2. Usunięte wierzchołki największego stopnia - minimalne obniżenie pozostałych wskaźników – niewielkie rozrzedzenie lokalnych powiązań
3. Usunięte wierzchołki największego pośrednictwa – średnia bliskość spadła, a centralność wektorowa wzrosła – pozostałe węzły muszą „nadrabiać” za usuniętymi, stąd wzrost pośrednictwa, ale spadek efektywności.
4. Usunięte wierzchołki o największej bliskości – centralność wektorowa spadła – usunięcie tych wierzchołków silnie wpływa na ich otoczenie

Wola:

Rodzaj grafu	Śr. Stopień	Śr. Pośrednictwo	Śr. Bliskość	Śr. Centralność wektorowa	Gęstość grafu
Bazowy	2.7423	0.0044	0.0382	0.0017	0.0005
Usunięty stopień	2.6806	0.0054	0.0310	0.0012	0.0005
Usunięte pośrednictwo	2.7140	0.0066	0.0251	0.0014	0.0005
Usunięta bliskość	2.7225	0.0057	0.0280	0.0017	0.0005
Usunięty wektor własny	2.7221	0.0053	0.0324	0.0018	0.0005

Analiza:

1. Usunięte wierzchołki o największym stopniu – średni stopień spadł, 1% wierzchołków o najwyższym stopniu nie wyróżnia się znacząco w porównaniu z pozostałymi.
2. Usunięte wierzchołki o największym stopniu – minimalne obniżenie innych wskaźników - świadczy o niewielkim rozrzedzeniu lokalnych powiązań
3. Usunięte wierzchołki o największym pośrednictwie - pośrednictwo wzrasta - pozostałe węzły muszą „nadrabiać” rolę pośredniczącą.
4. Usunięte wierzchołki o największym pośrednictwie - średnia bliskość drastycznie spada - wydłużenie efektywnych ścieżek komunikacyjnych i spadek efektywności sieci.
5. Usunięte wierzchołki o największej bliskości - obniżenie średniej bliskości, centralność wektorowa pozostaje na poziomie bazowym
6. Usunięte wierzchołki o największej centralności wektorowej – minimalne zmiany wartości - wskazuje to na relatywnie równomierny rozkład wpływu mierzony centralnością wektorową

Barcelona

Rodzaj grafu	Śr. Stopień	Śr. Pośrednictwo	Śr. Bliskość	Śr. Centralność wektorowa	Gęstość grafu
Bazowy	2.8775	0.0058	0.0239	0.0013	0.0004
Usunięty stopień	2.8101	0.0052	0.0205	0.0021	0.0004
Usunięte pośrednictwo	2.8490	0.0019	0.0129	0.0013	0.0004
Usunięta bliskość	2.8537	0.0061	0.0226	0.0012	0.0004
Usunięty wektor własny	2.8573	0.0060	0.0235	0.0014	0.0004

Analiza:

1. Usunięte wierzchołki o największym stopniu – średni stopień spadł - wierzchołki o najwyższym stopniu nie wykazują ekstremalnie wyższych wartości niż reszta sieci
2. Usunięte wierzchołki o największym stopniu - centralność wektorowa wzrasta - oznacza minimalne przesunięcia w strukturze sieci
3. Usunięte wierzchołki o największym pośrednictwie - średnie pośrednictwo spada drastycznie, a średnia bliskość gwałtownie obniża się – oznacza to, że w tej sieci kilka węzłów pełniło kluczową rolę łącznikową – ich usunięcie znacznie wydłużyło ścieżki komunikacyjne pogarszając efektywność całego systemu
4. Usunięte wierzchołki o największej bliskości - średnia bliskość spada, centralność wektorowa minimalnie obniża się – miały istotne znaczenie dla efektywności, natomiast nie aż tak istotne jak te wierzchołki o największym pośrednictwie

5. Usunięte wierzchołki o największej centralności wektorowej - zbliżone do wariantu z usunięciem wierzchołków o najwyższej bliskości - świadczy to o dość równomiernym rozkładzie

VI. Podsumowanie

Analiza trzech sieci drogowych – WAT, Woli i Barcelony – wykazała, że najmniejsze i najbardziej zwarte sieci, takie jak kampus WAT, charakteryzują się największą efektywnością komunikacyjną. WAT uzyskał najkrótszą znormalizowaną długość ścieżki, najwyższą bliskość oraz największą gęstość, co przekłada się na łatwy dostęp do każdego punktu. Sieć Woli, choć większa i bardziej nieregularna, wykazuje dobrą lokalną spójność dzięki wysokiemu współczynnikowi klasteryzacji, zapewniając alternatywne trasy przejazdu. Z kolei Barcelona, mimo planowego układu, okazała się najmniej odporna na usunięcia kluczowych węzłów – silnie scentralizowana struktura prowadzi do pogorszenia komunikacji przy zakłóceniach. Wniosek: najlepsze sieci to te, które łączą zwartą strukturę z lokalną elastycznością i nie opierają się na kilku kluczowych punktach.