

Wydział Budownictwa Lądowego i Wodnego
Politechniki Wrocławskiej

Raport serii PRE nr 4/2026
Rozprawa doktorska
Metoda agregacji danych z dużych
wolumenów, zróżnicowanego
pochodzenia i struktury w celu
udoskonalenia i tworzenia modeli
transportowych

mgr inż. Michał Zawodny

Słowa kluczowe:

modelowanie transportu, symulacja,
inżynieria ruchu,
big data, łączenie, macierz
podróży

Promotor:

dr hab. inż. Maciej Kruszyna, prof. uczelni

Wrocław, kwiecień 2026

Autor:

1. mgr inż. Michał Zawodny

Politechnika Wrocławska
Wydział Budownictwa Lądowego i Wodnego
Katedra Dróg, Mostów, Kolei i Lotnisk
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

Raport został złożony w Redakcji Wydawnictw Wydziału Budownictwa Lądowego i Wodnego Politechniki Wrocławskiej w kwietniu 2026 r.

Lista odbiorców:

Recenzenci	3 egz.
Promotor	1 egz.
Autor	1 egz.
Archiwum W-2	1 egz.
Razem	6 egz.

Spis treści

Spis treści	3
Wykaz podstawowych skrótów i oznaczeń.....	5
Podstawowe skróty użyte w rozprawie	5
Podstawowe oznaczenia użyte w rozprawie.....	6
1. Wprowadzenie.....	9
2. Przegląd literatury	12
2.1. Modelowanie podróży	12
2.2. Aktualny stan wiedzy o wykorzystaniu Big Data w modelowaniu	14
2.3. Charakterystyka baz danych big data i symulowanie baz danych	25
2.4. Identyfikacja luki badawczej	31
3. Cele i układ pracy.....	33
3.1. Uwagi wstępne, przedmiot pracy	33
3.2. Cele pracy.....	33
3.3. Układ pracy	34
4. Metoda agregacji danych z dużych wolumenów, zróżnicowanego pochodzenia i struktury w celu udoskonalenia i tworzenia modeli transportowych	37
4.1. Metoda łączenia macierzy podróży z Big Data.....	37
4.2. Uwzględnienie źródeł Big Data w metodzie.....	42
5. Metoda symulacji ośrodka miejskiego i macierze podróży z Big Data	46
5.1. Podstawy teoretyczne metody symulacji ośrodka miejskiego	46
5.2. Model poglądowy symulowanego ośrodka miejskiego	49
5.3. Macierze podróży ze źródeł Big Data	58
6. Walidacja metody.....	68
6.1. Prezentacja wyników.....	68
6.2. Wykorzystanie metody i analiza rezultatów	69
6.2.1. Uzupełnienie wszystkimi dostępnymi źródłami	70
6.2.2. Wariant 2 - Uzupełnienie źródłami symulowanymi na bazie map ciepła i charakterystyki źródła danych	75
6.2.3. Wariant 3 – uzupełnienie danymi z wrocławskiego KBR z 2018.....	80
6.2.4. Wariant 4 – sam GSM	85
6.2.5. Wariant 5 – Bluetooth + APC + SmartCard Data + ANPR	90
6.2.6. Wariant 6 – uzupełnienie danymi z mediów społecznościowych.....	95

6.2.7.	Wariant 7 – uzupełnienie danymi z Taxi	100
6.2.8.	Wariant 8 – dane z rowerów miejskich i Urzędzeń Transportu Osobistego	105
6.2.9.	Wariant 9 – Dane z GSM, mediów społecznościowych i Taxi.....	110
6.3.	Analiza rezultatów i skuteczności uzupełniania danych	115
6.4.	Podsumowanie	117
7.	Dyskusja	119
7.1.	Ogólny opis rezultatów	119
7.2.	Ocena metody.....	119
8.	Wnioski końcowe i kierunki dalszych prac.....	121
8.1.	Realizacja celu.....	121
8.2.	Kierunki dalszych prac	122
Spis rysunków		124
Spis tabel		126
Spis załączników		127
Literatura		128
Streszczenie		142
Abstract		143
ZAŁĄCZNIKI.....		144

Wykaz podstawowych skrótów i oznaczeń

Podstawowe skróty użyte w rozprawie

AFC – Automated Fare Collection – System Automatycznego Biletowania

ANPR – Automatic Number-Plate Recognition – Automatyczny System Rozpoznawania Numerów Rejestracyjnych

APC – Automatic Passenger Counter – System Automatycznego Zliczania Pasażerów

AVL – Automatic Vehicle Location – Systemy automatycznej lokalizacji pojazdu

BIM – Building Information Modelling

BLE – Bluetooth Low Energy

BTS – Base Transceiver Station – Stacja Bazowa

CCTV – Closed Circuit Television – Monitoring wizyjny

CDR – Call Detail Record/Call Data Record

DT – Digital Twin – Cyfrowy Bliźniak

FCD – Floating Car Data

FTP – Fuzzy Transportation Problem

GIS – Geographic Information System – System Informacji Geograficznej

GNSS – Global Navigation Satellite System – System Globalnej Nawigacji Satelitarnej

GPS – Global Positioning System

GSM – Global System for Mobile Communications – Sieć Telefonii Komórkowej

ITS – Intelligent Transportation Systems – Systemy Inteligentnego Transportu

IoT – Internet of Things – Internet Rzeczy

iOS – iPhone Operating System

KBR – Kompleksowe Badania Ruchu

kNN – k Nearest Neighbours – Algorytm k najbliższych sąsiadów

MAC – Media Access Control

MC – symulacja Monte Carlo

MCMC – Markov Chain Monte Carlo

O-D – Origin-Destination (Matrix) – Macierz podróży źródłowo-docelowa

POI – Point of Interest – Punkt Zainteresowania

QR – Quick Response (Code)

RSSI – Received Signal Strength Indicator

UTO – Urządzenia Transportu Osobistego

V2I – Vehicle-to-Infrastructure – komunikacja pojazdu z drogą

Podstawowe oznaczenia użyte w rozprawie

$K0_i$ – liczba dzieci do lat 7 w rejonie i-tym

$K1_i$ – liczba mieszkańców pracujących w rejonie i-tym

$K2_i$ – liczba mieszkańców uczących się w szkołach ponadpodstawowych oraz studentów w rejonie i-tym

$K3_i$ – liczba mieszkańców uczących się w szkołach podstawowych w rejonie i-tym

$K4_i$ – liczba emerytów w rejonie i-tym

KX_i – liczba grupy mieszkańców w rejonie i-tym

L – wygenerowana liczba pseudolosowa w zakresie liczb rzeczywistych od 0 do 1

M_{ij} – Liczba podróży pomiędzy rejonem i-tym i j-tym w utworzonej metodą macierzy podróży

M_{ij}^m – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z źródła m

M_{ij}^{ANPR} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z automatycznych systemów rozpoznawania tablic rejestracyjnych

M_{ij}^{APC} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z systemów zliczania pasażerów

M_{ij}^{BT} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z ze skanerów Bluetooth

M_{ij}^{GPS} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z GPS lub innego źródła opartego na GNSS

M_{ij}^{GSM} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z sieci telefonii komórkowej

M_{ij}^{SC} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z systemów inteligentnych kart miejskich

M_{ij}^{SM} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z mediów społecznościowych

M_{ij}^T – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z taxi

M_{ij}^{UTO} – Liczba podróży pomiędzy rejonem i-tym i j-tym w macierzy podróży pochodzącej z flot urządzeń transportu osobistego lub rowerów miejskich

N – Liczba wykorzystanych we wzorze źródeł Big Data

R_{DN2} – Współczynnik ruchotwórczości w motywacji dom-nauka dla grupy K2

R_{DN3} – Współczynnik ruchotwórczości w motywacji dom-nauka dla grupy K3

R_{DP1} – Współczynnik ruchotwórczości w motywacji dom-praca dla grupy K1

R_{DI1} – Współczynnik ruchotwórczości w motywacji dom-inne dla grupy K1

R_{DI2} – Współczynnik ruchotwórczości w motywacji dom-inne dla grupy K2

R_{DI3} – Współczynnik ruchotwórczości w motywacji dom-inne dla grupy K3

R_{DI4} – Współczynnik ruchotwórczości w motywacji dom-inne dla grupy K4

R_{NZD1} – Współczynnik ruchotwórczości w motywacji niezwiązane z domem dla grupy K1

R_{NZD2} – Współczynnik ruchotwórczości w motywacji niezwiązane z domem dla grupy K2

R_{NZD3} – Współczynnik ruchotwórczości w motywacji niezwiązane z domem dla grupy K3

R_{NZD4} – Współczynnik ruchotwórczości w motywacji niezwiązane z domem dla grupy K4

S – stała we wzorze na generowanie liczby grupy mieszkańców

W^{ANPR} – Współczynnik przeliczeniowy dla źródła z automatycznych systemów rozpoznawania tablic rejestracyjnych

W^{APC} – Współczynnik przeliczeniowy dla źródła z systemów zliczania pasażerów

W^{BT} – Współczynnik przeliczeniowy dla źródła ze skanerów Bluetooth

W^{GPS} – Współczynnik przeliczeniowy dla źródła z GPS lub innego źródła opartego na GNSS

W^{GSM} – Współczynnik przeliczeniowy dla źródła z sieci telefonii komórkowej

W^{SC} – Współczynnik przeliczeniowy dla źródła z systemów inteligentnych kart miejskich

W^{SM} – Współczynnik przeliczeniowy dla źródła z mediów społecznościowych

W^T – Współczynnik przeliczeniowy dla źródła z taxi

W^{UTO} – Współczynnik przeliczeniowy dla źródła z flot urządzeń transportu osobistego lub rowerów miejskich

W^m – Współczynnik przeliczeniowy dla źródła m

W_w^m – Współczynnik wiarygodności dla źródła m

Y_{ij} – ruchotwórczość pomiędzy rejonem i -tym i j -tym

Z – zmienna we wzorze na generowanie liczby grupy mieszkańców

$\%AI_j$ – procentowa atrakcyjność rejonu j -tego motywacji inne

$\%ANPod_j$ - procentowa atrakcyjność rejonu j -tego motywacji szkoła podstawowa

$\%AP_j$ – procentowa atrakcyjność rejonu j -tego motywacji praca

$\%AŚr_j$ – procentowa atrakcyjność rejonu j -tego motywacji szkoła średnia

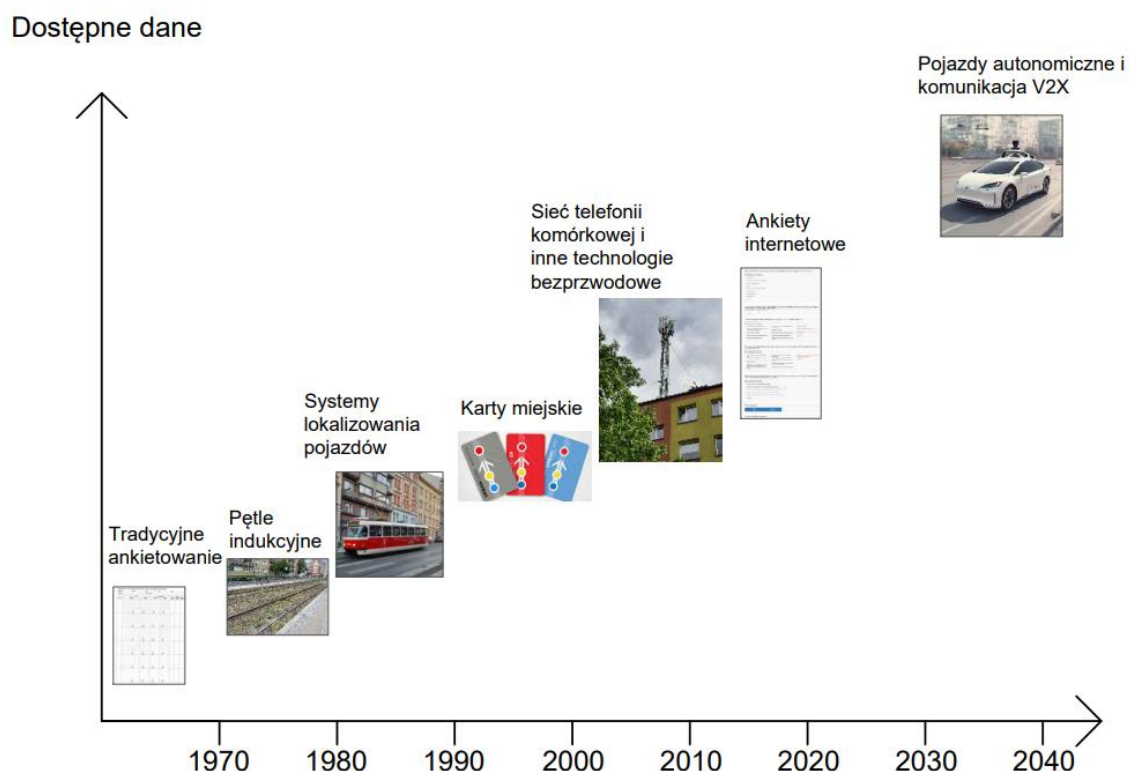
$\%AWyż_j$ – procentowa atrakcyjność rejonu j -tego motywacji szkoła wyższa

1. Wprowadzenie

W budownictwie związanym z transportem podobnie jak w przypadku konstrukcji kubaturowych należy przyjąć wstępne założenia i zestawić oddziaływania zewnętrzne. Jednymi z oddziaływań, istotnymi dla infrastruktury transportowej, będą potoki podróźnych. W przypadku remontów lub przebudowy infrastruktury istniejącej można dokonać pomiarów ruchu i przeliczyć je na liczbę osi które będą oddziaływać w założonym horyzoncie czasowym na konstrukcję jezdni. Sytuacja jednak komplikuje się, gdy wykracza się poza zwykłe remonty dróg i dokonuje się modyfikacji całej sieci transportowej nowymi połączeniami lub budową nowych tras tramwajowych. Wtedy zwykłe pomiary ruchu nie wystarczą. Trzeba oszacować ile podróźnych wybierze planowany środek transportu, który wariant przyciągnie najwięcej mieszkańców do korzystania z publicznego transportu zbiorowego. W jaki sposób zaprojektować infrastrukturę, by ta najlepiej spełniła swoje działania, więc była najbardziej korzystna z punktu analiz kosztów i korzyści. W tym celu potrzebne są modele transportowe. Są one również niezbędne do współczesnej metody oceny infrastruktury, która nie polega tylko na osiągnięciu maksymalnej przepustowości, często kosztem innych użytkowników sieci drogowej, a poprawieniem mobilności mierzalnej na przykład czasem podróży od drzwi do drzwi [94]. Modele dzielą się na makroskopowe, mezoskopowe i mikroskopowe. Modele mikroskopowe uwzględniają szczegółowy ruch pojazdów na małym elemencie infrastruktury, w postaci np. skrzyżowania, ronda, węzła drogowego i służą do weryfikacji zastosowanych rozwiązań. Umożliwiają, także zbadanie przepustowości konkretnych elementów, takich jak łącznice, czy relacje skrajne. Modele mikroskopowe opisujące ruch drogowy są modelami stochastycznymi, ponieważ muszą odwzorować zachowania ludzkie, które są trudne do przewidzenia. Wyjątkiem będą modele ruchu kolejowego, który ze względu na poruszanie się pojazdów po określonym torze i odcinku blokady liniowej, będzie modelem deterministycznym. Wykorzystywanie tego typu modeli pozwala zweryfikować warianty rozwiązań infrastrukturalnych. To podejście często wykorzystywane do weryfikacji rozwiązań infrastrukturalnych w Wielkiej Brytanii [132]. Transport for London posiada nawet poradnik do tworzenia modeli na ich potrzeby [14]. Rząd Wielkiej Brytanii wydał również poradnik do tworzenia modeli makroskopowych [161]. Podobne poradniki wydały również rządy Nowej Zelandii [162], oraz Australii [166]. Modele makroskopowe służą do opisanie i prognozowania zachowań transportowych z perspektywy sieci transportowych. Ruch pojazdów, czy innych użytkowników ruchu nie jest opisany szczegółowo, a ilościowo. Modele mogą zakładać wycinek sieci miejskiej, np. dzielnicy, całego ośrodka miejskiego, aglomeracji, lub nawet

całych państw, jak w przypadku Pasażerskiego Modelu Transportowego [164] stworzonego na potrzeby Centralnego Portu Komunikacyjnego, a w szczególności inwestycji towarzyszących. Model mezoskopowy to model makroskopowy z elementami mikroskopowego takimi jak na przykład wprowadzone do węzłów cykle sygnalizacji świetlnej na skrzyżowaniach, które ze względu na straty czasu, wpływają na decyzje podróżnych do wyboru konkretnego połączenia na sieci transportowej. W niniejszej rozprawie poruszane będzie modelowanie makroskopowe na poziomie aglomeracji miejskiej.

Modele transportowe potrzebują danych wstępnych o przemieszczeniach podróżnych. Nie są to zawsze wartości łatwo mierzalne jak na przykład liczba pojazdów konkretnej kategorii na konkretnym odcinku drogi. Pierwotnie w tym celu używano badań ankietowych, choć w Polsce przy niskiej responsywności badania są trudne do przeprowadzenia. W celu pozyskania szczegółowych danych o przemieszczeniach podróżnych zaczęto wykorzystywać źródła Big Data. Szacowany przyrost liczby dostępnych źródeł w dziedzinie pokazano na rys. 1.



Rys. 1.1 Poglądowy wzrost liczby źródeł Big Data wraz z upływem lat

Źródeł danych jest wiele, najbardziej popularne w Polsce to sieć GSM. Inne źródła to na przykład przemieszczenia podróżnych możliwe do zbadania z używanych kart miejskich w publicznym transporcie zbiorowym, systemy zliczania pasażerów, skanery wykrywające urządzenia Bluetooth itd. Wykorzystywać można również dane z aplikacji na telefony, na przykład do zamawiania przewozu taksówką. Liczba przydatnych do modelowania ruchu źródeł danych przyrasta, więc pojawia się problem braku ich kompatybilności.

2. Przegląd literatury

2.1. Modelowanie podróży

W wielu dyscyplinach wykorzystuje się modele będące odwzorowaniem rzeczywistości. Mogą być to modele fizyczne lub obliczeniowe. W przypadku modelowania transportu korzystamy z modeli matematycznych, które pozwalają na odtwarzanie zjawisk ważnych z punktu widzenia transportu za pomocą opisu atrybutów w postaci zmiennych matematycznych i działaniu opisanego wzajemnemu powiązaniu funkcjami matematycznymi. O ile możliwe jest wykonanie wszystkich obliczeń za pomocą kartki i papieru, jest to czasochłonne i nie pozwala uszczegółowić modeli do satysfakcjonującego poziomu.

Modelowanie transportu miało swoje początki w Detroit w latach 50. XX wieku [108] po obserwacjach pierwszych oznak kongestii spowodowanej wzmożonym ruchem pojazdów samochodowych indywidualnych. Pierwszą polską pozycją literaturową o zastosowaniu symulacji komputerowej w inżynierii ruchu jest [82], gdzie symulacje przeprowadzano w języku ALGOL m. in. na komputerach ODRA-1204 i ODRA-1305. Aktualnie przeprowadzanie symulacji jest możliwe z poziomu komputera osobistego za pomocą oprogramowania służącego do modelowania ruchu takiego, jak wykorzystany w rozprawie PTV Visum, lub jego zamienniki. Program jest powszechnie wykorzystywany, wykonano w nim m. in. Pasażerski Model Transportowy CPK czy Model Ruchu Wrocławia i wielu innych polskich miast. Program pozwala na modyfikacje za pomocą skryptów pisanych w języku programowania Python, co pokazano w [79] i [83] i pozwala to na elastyczne wykorzystanie w badaniach. Narzędzie jest również kompatybilne z programami do mikromodelowania ruchu, takimi jak PTV Vissim od tego samego producenta, ale również SUMO [13] dostępny na licencji open-source. Oprogramowanie ma jednak limity. W [61] pokazano jak szczegółowość modeli ogranicza możliwości obliczeniowe komputerów. W rozprawie stosowano nomenklaturę zgodną z [81] oraz [46].

W Rozdziale pierwszym opisano różnice pomiędzy modelem makroskopowym wykorzystywanym w rozprawie, a modelem mezoskopowym i mikroskopowym. Jest to jeden z wymiarów modeli ruchu, czyli poziom odtworzenia szczegółów charakterystyki ruchu. Drugą formą kategoryzacji modeli będzie jego konstrukcja. W rozprawie badano elementy

klasycznego czterostadiowego modelu ruchu. Pierwsze stadium modelu to generacja podróży. W tym etapie określa się potencjał rejonów transportowych do generowania i absorbowania podróży w określonym interwale czasowym np. godzinowym lub dobowym użytym w rozprawie. Atrakcyjność rejonów determinuje m. in. liczba miejsc pracy czy nauki. Podróże generować będą mieszkańcy należący do konkretnych grup społecznych np. uczniowie będą odbywać podróże dom-nauka i nauka-dom, a pracujący analogicznie dom-praca i praca-dom. W rozprawie wykorzystano ten krok modelu w pierwszej z dwóch metod, opisanej w Rozdziale czwartym – metody symulowania ośrodka miejskiego.

Drugim stadium modelu klasycznego jest rozkład przestrzenny podróży. Określa on liczbę podróży pomiędzy poszczególnymi parami rejonów transportowych. Wykorzystuje go „główna” metoda niniejszej rozprawy polegająca na łączeniu zróżnicowanych źródeł Big Data. Rozkład podróży może określać prostszy model proporcjonalny, w którym większy rejon przyciąga większą liczbę podróży proporcjonalnie do rozmiarów rejonu i jego atrakcyjności. Modelem bardziej zaawansowanym jest model grawitacyjny, który uwzględnia odległości pomiędzy rejonami oraz kosztowność podróży, zmniejszając liczbę podróży pomiędzy parami rejonów, które są gorzej skomunikowane.

Trzecim stadium jest wybór środka transportu. Określa się w nim podział modalny dzielący podróże na danej parze źródłowo-docelowej na poszczególne środki transportu. Wybór jest zależny od dostępnej infrastruktury i najczęściej podróżnych dzieli się na korzystających z transportu indywidualnego samochodowego, publicznego transportu zbiorowego, transportu pieszego oraz rowerowego.

Ostatnim stadium - czwartym jest rozkład ruchu na sieć transportową. Na bazie znanej sieci transportowej oraz zachowań transportowych podróżnych ustala się jaką drogę wybiorą. Wybór jest zależny od kosztów (finansowych lub czasowych) odbycia podróży danym połączeniem. Na te decyzje wpływa na przykład przepustowość połączeń.

Innym typem modeli transportowych są modele oparte o aktywności. W tym podejściu modeluje się w oparciu o plan aktywności w ciągu dnia osób stanowiących gospodarstwa domowe, co bezpośrednio przekłada się na czas i miejsce podróży. W anglojęzycznych badaniach tego typu dane, które da się uzyskać za pomocą Big Data, określane są jako „spatio-temporal data”. Typami motywacji podróży są najczęściej:

- Dom-Praca
- Praca-Dom
- Dom-Nauka
- Nauka-Dom
- Dom-Inne
- Inne-Dom
- Niezwiązane z domem

W nowszych pozycjach rozróżnione na:

- Niezwiązane z domem zawodowe (Non-Home-Based Business)
- Niezwiązane z domem inne

Nie jest to pierwsza rozprawa doktorska dotycząca modelowania ruchu w dyscyplinie Inżynieria Lądowa, Geodezja i Transport. Rozprawa doktorska [18] porusza temat wyboru środka lokomocji przez użytkowników ruchu i tym jak działania zarządców infrastruktury wpływają na zmianę przemieszczeń poszczególnymi środkami lokomocji. Rozprawa wykorzystuje metody heurystyczne, w postaci logiki rozmytej, co pozwoliło zbadać wpływ niestosowanych powszechnie zmiennych na prognozowanie przemieszczeń podróży.

2.2. Aktualny stan wiedzy o wykorzystaniu Big Data w modelowaniu

Big Data jest stosunkowo nowym elementem badań w modelowaniu ruchu, jednak jest przedmiotem istotnej części aktualnych badań w tej gałęzi nauki. W tym podrozdziale dokonano przeglądu stanu wiedzy na temat wykorzystania wybranych źródeł danych w modelowaniu podróży. Podczas przeglądu szczególnie zwracano uwagę na informacje niezbędne do zrealizowania badań w niniejszej rozprawie, takie jak parametry techniczne baz danych pozyskiwanych z różnych źródeł. Szukano również macierzy podróży wraz z ich opisami w formach liczbowej oraz map ciepła. Na bazie przeglądu wytypowano lukę badawczą jaką jest brak uniwersalnej metody łączenia wielu źródeł i zaproponowano autorską metodę.

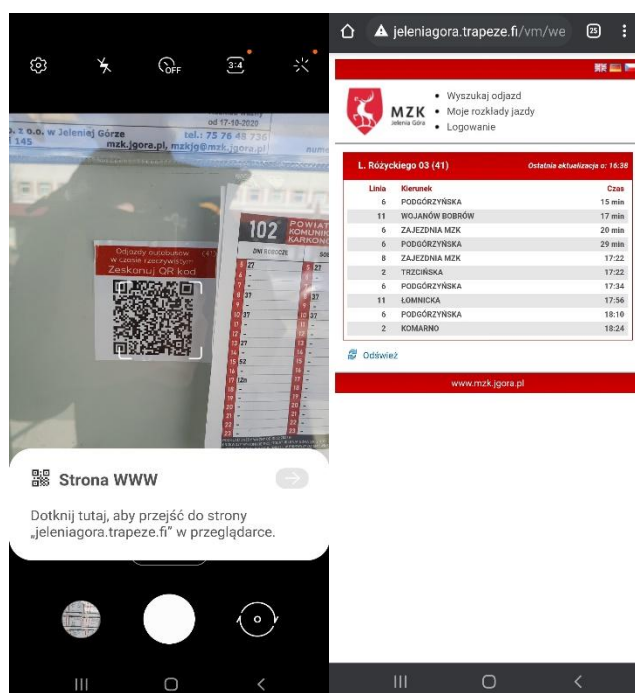
W [57] skupiono się na problemie tendencyjności danych pochodzących z dużych baz danych. Publikacja jest skupiona na temacie bezpieczeństwa w ruchu, jednak podczas tworzenia modeli ruchu wykorzystuje się dane tego samego typu, co powoduje zbliżone problemy jak w tematyce dotyczącej przedmiotu tej pozycji. Poruszono tam również aspekt wykorzystywania konkretnych źródeł danych w modelowaniu. Metodą badawczą był wywiad z ekspertami ze

środowisk akademickich, komercyjnych i urzędniczych. W przypadku źródła tendencyjności w danych demograficznych autorzy [57] jako rozwiązanie wskazali uwzględnienie badań tradycyjnych. Problem tendencyjności danych był również poruszany w książce [125]. [119] wykorzystało ANPR do redukcji tendencyjności danych pochodzących z floty taksówek.

Pierwszymi omawianymi źródłami Big Data są kody QR i NFC. Te dwa źródła zestawiono razem, ze względu na podobne zastosowanie w publicznym transporcie zbiorowym. Są to technologie, które mogą być wykorzystywane alternatywnie. Na początek należy wyjaśnić różnice pomiędzy tymi dwoma technologiami. Kod QR (Quick Response Code – Kod Szybkiej Odpowiedzi) jest to biało-czarny znak, który podróżny skanuje aparatem swojego telefonu. Zeskanowany kod zawiera informację zawartą w bajtach, na przykład odnośnik do danej strony internetowej. NFC (Near Field Communication – Komunikacja Bliskiego Zasięgu) przekazuje dane za pomocą komunikacji smartfonu z nadajnikiem za pomocą fal radiowych na niewielkim, około dwudziestocentymetrowym dystansie. Jedno z urządzeń może również być pasywne, komunikacja może wtedy odbyć się za pomocą indukcji magnetycznej. Wtedy nie zachodzi potrzeba zasilania obydwu urządzeń.

Liczba osób korzystających z QR Kodów stale wzrasta według [67] i trend wzrostowy ma się utrzymać. Według przekazów medialnych wzrasta liczba podmiotów korzystających z tej technologii. W wielu krajach technologia jest również wykorzystywana do śledzenia kontaktów przez aplikacje instytucji ochrony zdrowia [163]. Działania te wskazują na możliwe poszerzenie liczby osób korzystających z tego typu technologii. Rośnie więc uzasadnienie wykorzystania tych technologii w publicznym transporcie zbiorowym.

W Polsce kody QR nie są powszechnie stosowane w transporcie publicznym. Jeżeli już się je stosuje, to służą one podawaniu podróżnemu linku do dynamicznego rozkładu jazdy. Jest to tańsza opcja niż wyświetlacze na przystankach o mniejszych potokach pasażerskich. Tego typu systemy są wykorzystywane w np. MPK Lublin czy MZK Jelenia Góra.



Rys. 2.1 Skanowanie kodu QR na przystanku

Na rys. 2.1 przedstawiono zrzuty ekranu z telefonu skanującego kod QR umieszczony w gablocie przystankowej (po lewej) oraz rozkład jazdy dostępny po wejściu na stronę WWW, do której link został zapisany w kodzie (po prawej). Zastosowanie to, jednak nie daje dużej szansy na wykorzystanie danych do modelowania potoków podróży. Jednakże, można rozszerzyć funkcjonalność kodów. W Porto pojęto próbę zastosowania kodów do systemu biletowania celem dematerializacji biletów [44]. Sposób działania aplikacji pozwala na rejestrowanie podróży. Podróżny musi się zameldować (check-in) skanując kod QR w pojeździe przewoźnika oraz po skończonej podróży wymeldować (check-out) za pomocą zeskanowania kodu umieszczonego na przystanku końcowym. Rozwiązanie to może pozwolić na rejestrowanie przystanku początkowego i końcowego indywidualnego podróżnego, co może być wykorzystane w modelowaniu ruchu. W przypadku NFC nie ma dostępnych wielu źródeł literaturowych dotyczących modelowania ruchu. W [49] wskazano na duży wzrost popularności tej technologii jako metody płatności, a w [8] badano jak technologia jest postrzegana przez użytkowników w kontekście bezpieczeństwa. Aktualnie ma zastosowanie w transporcie jako jedna z technologii wykorzystywana w kartach miejskich (smart card) [111]. Duży potencjał do wykorzystywania w systemach biletowania zauważono również w raporcie [135].

Alternatywą może być Beacon Bluetooth. Jest to urządzenie podobne rozmiarami do urządzeń działających w technologii NFC i jest zasilane baterią. Emituje ono sygnał zawierający informację, którą mogą odebrać inne urządzenia skanujące w poszukiwaniu sygnałów Bluetooth [32]. BLE (Bluetooth Low Energy) oznacza urządzenia technologii Bluetooth od wersji 4.0 wzwyż. Wersje te charakteryzują się mniejszym zużyciem energii od poprzedników, co zwiększa możliwości zastosowania. Szczegóły technologii opisano w [56]. Jak wskazuje nazwa Beacon BLE nie będzie rejestrował danych, a je nadawał. Ciężar rejestracji podróznego i jego obsługa (np. w postaci zakupu biletu) będzie spoczywała na barkach aplikacji w telefonie podróznego. Połączenie z beaconem może zameldować podróż w aplikacji. Zespół badawczy z Porto opisał w [140] oraz [24] ich implementację, za to inną implementację opisano w [129]. W odróżnieniu od kodów QR i NFC, przy użyciu beaconów BLE możliwy jest również schemat be-in/be-out lub mieszany. Zasady działania różnią się tym, że początek i koniec podróży przy włączonej aplikacji może być rejestrowany bez interakcji z użytkownikiem. Zameldowanie i/lub wymeldowanie podróznego może odbyć się automatycznie w momencie nawiązania i zakończenia połączenia z podróznym. Szczegóły w jaki sposób można wykryć podróznego za pomocą beacona w szczegółowej lokalizacji opisano w [31]. W [36] wykorzystano w tym celu sieć neuronową i symulację Monte Carlo.

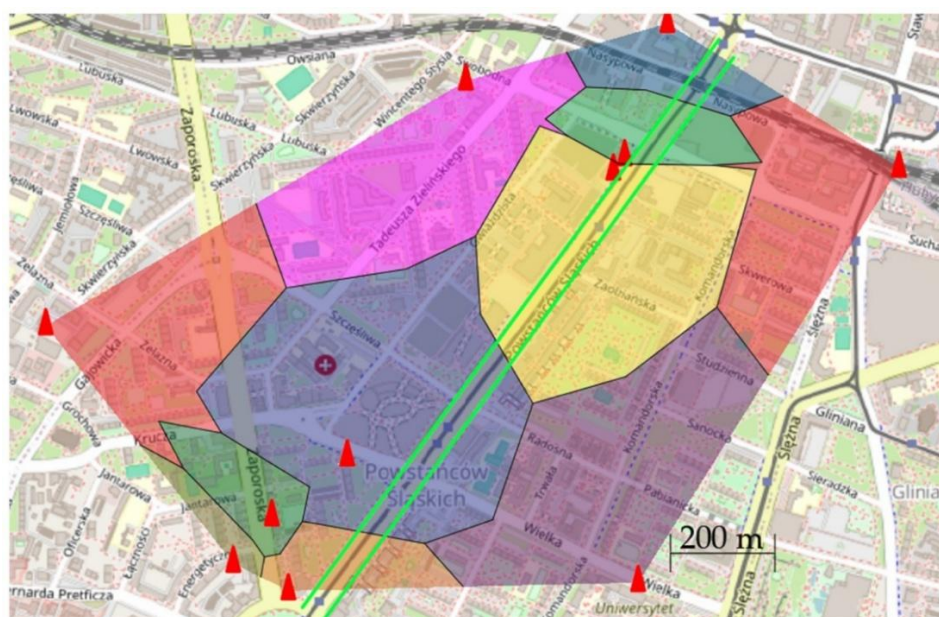
Beacony BLE wykorzystywano również do badań wewnątrz budynków. Jest to kolejne źródło danych, które można wykorzystać w modelowaniu transportu. Badania przemieszczeń podróжных w sklepach za pomocą Bluetooth mają podobne zagadnienia jak w przypadku modelowania transportu, w [113] badano na przykład długość podróży „zakupowej”. W wykorzystaniu Beaconów BLE w podejściu aktywnościowym modelowania ruchu kluczowe mogą być badania takie jak [53], gdzie wykorzystano Beacon do zbierania danych w konkretnych generatorach, określanych jako POI (point of interest). W tym przypadku było to wykonane w skali mikro, gdzie poszczególne z 10 beaconów było umiejscowionych przy 10 dziełach sztuki. Badacze udostępnili również pozyskaną bazę danych [52].

Jako źródło Big Data można traktować również dane pochodzące z systemów zliczania pasażerów - APC (Automatic Passenger Count). Za pomocą ankiety i systemów zliczania można oszacować źródło i cel podróży [72]. Podobnym źródłem informacji związanym z publicznym transportem zbiorowym będzie AFC (Automated Fare Collection). Są to źródła danych oparte na systemie biletowania w transporcie publicznym. W literaturze to źródło danych może być przedstawiane również jako Smart Card Data, gdy w systemach publicznego

transportu zbiorowego używa się kart miejskich, takich jak na przykład wrocławski Urban Card. Przegląd dotyczący wykorzystywanych technologii jest dostępny w [17]. W [78] zaproponowano zaawansowany algorytm Weighted Stop Density Based Scanning Algorithm with Noise (WS-DBSCAN), w celu analizy przestrzennej danych. Natomiast w [159] udało się zbadać motywacje podróży. Przy dodaniu metod opartych na GIS możliwe jest ustalenie całych łańcuchów motywacji podróży [88]. Stworzono również aplikację, która za pomocą systemów opisanych w niniejszym akapicie przekazuje w czasie rzeczywistym dane o zajętości pojazdów transportu publicznego dla podróżnych [71]. W przypadku źródeł opartych na GPS i CDR problematyczne jest metro. Dane można uzupełnić używając baz danych ze smart card [160]. Można więc wnioskować, że zachodzi potrzeba stworzenia metody uzupełniania tych baz danych. Kluczowy z punktu widzenia części badawczej rozprawy był odsetek użytkowników tej technologii, przykładowo w Seulu jest to 99% podróżnych [87]. Pozycja [157] wykorzystwała dane z 19 milionów podróży MTR (system kolei miejskiej) w Hong Kongu. Analizę przeprowadzono siecią neuronową skupiając się na relacji podróżnych z konkretną stacją. Zbadano motywacje podróży i macierze podróży. Badanie dotyczące metra w Shenzhen analizowało wielkość potoków podróżnych. Badanie wykorzystywało tylko dane z wejścia i wyjścia z bramek metra i badało np. wewnątrz których stacji dochodziło do przesiadek [150]. W [143] profilowano typy podróżnych. W [158] badano charakterystykę podróżnych odbywających podróże do pracy oraz podróżnych niepodróżujących do pracy. Osoby korzystające z publicznego transportu zbiorowego w Pekinie odbywały podróże, które były dłuższe i częstsze, a w weekendy pokrywały większy obszar niż w dni pracujące, będąc jednak krótsze niż w dni pracujące. Podróżujący niedojeżdżający publicznym transportem zbiorowym do pracy ogólnie podróżowali mniej w dni pracujące w celu uniknięcia kongestii, ale w weekendy podróżowali dłużej, częściej i bardziej zróżnicowanie.

Lokalizację posiadacza karty SIM szacuje się na podstawie BTS (Base Transceiver Station – stacja bazowa), do której wysyłany jest sygnał. W momencie połączenia ze stacją rejestruje się wystąpienie połączenia i czas, w którym do niego doszło. Pozwala to oszacować rejon, w którym znajduje się użytkownik sieci. Według [20] granice rejonów teoretycznie będą wyznaczone przez połowę odległości pomiędzy stacjami bazowymi. Rejony przybiorą więc kształt wielokątów na płaszczyźnie podzielonej na kształt Diagramu Woronoja. Wielokąty te będą mniejsze w centrach miast, gdzie zagęszczenie stacji bazowych jest większe oraz większe na przedmieściach i obszarach wiejskich, ze względu na duże odległości pomiędzy nielicznymi stacjami bazowymi. Autor wykorzystał tę metodę w pozycji [155]. Realne granice będą się

różnić od przybranych w powyższy sposób ze względu na ukształtowanie terenu, zabudowę i inne przeszkody na drodze fal elektromagnetycznych. Wyznaczanie stref transportowych za pomocą algorytmu centroidów (k-means clustering) szczegółowo badano w [40]. Poszczególne badania mają również mocno zróżnicowaną liczbę rejonów transportowych. Przykładowo w badaniu przeprowadzonym w Sztokholmie [26] miasto podzielono na 32 rejonów. W znacznie mniejszym Wrocławiu KBR z 2018 roku [167] wykorzystując dane z sieci telefonii komórkowej podzielił miasto na 475 rejonów macierzy podróży pozyskanej z tego źródła Big Data, by dane te pasowały do podziału miasta na klasyczne rejonów transportowe. Pogłębiona analiza wrocławskiego KBR została dokonana w [80]. W [41] za pomocą sieci telefonii komórkowej udało się dokonać podziału na klasyczne typy podróży związane z domem, pracą i innymi. Artykuł [137] zaproponował oprogramowanie, które łączy dane z sieci telefonii komórkowej z danymi klasycznymi w postaci ankiet, spisów ludności i powszechnie dostępnych danych geolokalizacyjnych. Kluczowy w wykorzystaniu tego typu danych jest odsetek podróżnych. W Polsce udział operatorów sieci telefonii komórkowej jest mocno rozdrobiony, jednak w innych krajach przy większym monopolu, udało się przeprowadzić badania na bazach danych obejmujących większość podróżnych. W przypadku badania z Bangladeszu jest to 65% [68]. W Holandii, gdzie sytuacja również jest lepsza niż w Polsce, odsetek ten już spada do 32% [149]. Przy uwzględnianiu odsetka użytkowników należy mieć na uwadze strukturę wiekową użytkowników, ponieważ z rosnącym wiekiem, liczba użytkowników spada [168].



Rys. 2.2 Diagram Woronoja – rejonów stworzone przez granice oddziaływania stacji bazowych (czerwone trójkąty) w obszarze ulicy Powstańców Śląskich (zaznaczonej na zielono) (Opracowanie własne na podkładzie mapowym [169])

Aktualizacja obszaru w którym znajduje się podróżny jest zależna od tego z jakiej bazy danych się korzysta. W [20] zwrócono uwagę na istnienie dwóch rodzajów baz danych. Pierwsza to baza bilingowa, która jest jednak tendencyjna, ponieważ zależy od częstości wykonywania połączeń za pomocą telefonu komórkowego. Będzie to generowało tendencyjność skorelowaną np. z wiekiem podróżnego. Wykorzystano więc bazę sygnałową, która rejestruje zmianę stacji bazowej, w której zasięgu znajduje się podróżny. Lokalizacja sprawdzana jest co 6 godzin, lub co 3 godziny w przypadku nowszych systemów. Analiza tego typu danych pozwala oszacować drogę podróżnego. Omówiony jest tam również efekt ping-ponga i „bias” związany z charakterystyką socjodemograficzną użytkowników telefonów komórkowych. Porównano również dane pochodzące od operatorów sieci komórkowej z tradycyjnymi badaniami ankietowymi. W [85] podjęto ten sam obszar badawczy jak w [20], którym jest aglomeracja Paryża i zwrócono szczególną uwagę na problem wykrywania podróżnych przemieszczających się metrem. Jest to charakterystyczny obszar, gdzie dwie bazy danych mogą się uzupełniać. Podróżnych w metrze można badać za pomocą źródła pochodzącego z systemów płatności lub w oderwaniu od konkretnego właściciela baz danych, przeprowadzić zewnętrzne badania za pomocą skanerów WiFi/Bluetooth opisanych poniżej. W [107] przedstawiono syntezę badań na temat wykorzystania sieci telefonii komórkowej w modelowaniu ruchu i zaprezentowano 6 źródeł danych dostępnych publicznie.

Alternatywną metodą wobec pomiarów za pomocą sieci komórkowej jest wykrywanie urządzeń Bluetooth podróżnych. W tej metodzie nie ma potrzeby pozyskiwania danych od podmiotów komercyjnych. Trzeba jednak rozmieścić skanery, które wykrywają urządzenia z włączonym modułem Bluetooth w trybie widoczności. Każde urządzenie ma swój unikalny adres MAC. Skanery rejestrują miejsce i czas, w którym urządzenia posiadające dany adres zostały wykryte. Pozwala to, na odtworzenie podróży pomiędzy skanerami. Mniej zaawansowanym wykorzystaniem mogą być pomiary ruchu wraz z określeniem średniego czasu i prędkości na odcinku [96]. W nowszej pozycji wykazano, że w porównaniu do użycia radaru, skuteczność Bluetooth to zaledwie 62% [50]. Według [128] już 10 lat temu 85% podróżnych korzystało z tej technologii, z czego 35% korzysta z niej często, a 10% używało jej stale w trybie widoczności. W [21] za pomocą urządzenia Ubertooth One użyto nowej metody zbierającej i analizującej ruch urządzeń Bluetooth w scenariuszu zamkniętego pomieszczenia na uczelni. Metoda ta przechwyciła 263 unikalne adresy MAC w pięciu sesjach, sklasyfikowała poziomy zaufania ruchu, odfiltrowała szum za pomocą progów RSSI i zlokalizowała urządzenia za pomocą trilateracji i map cieplnych.

Alternatywą może być pozyskanie adresu MAC za pomocą skanera Wi-Fi, jednak pomimo dużych nadziei w [96], w [1] wykazano już znacząco mniejszą użyteczność danych w porównaniu do wykrywania adresów za pomocą Bluetooth. Skuteczność samej technologii Bluetooth wykorzystywanej w tym celu wzrasta z wersją 5.0 [30].

Użyteczność tej metody może się zwiększyć w przypadku, gdy do ruchu dołączą pojazdy autonomiczne. Bluetooth jest jedną z proponowanych technologii komunikacji pojazdów z infrastrukturą (V2I – vehicle to infrastructure) [48]. Dane pochodzące od flot pojazdów autonomicznych określane są w literaturze jako FCD (Floating Car Data) [89]. Aktualne badania na temat Bluetooth pozwalają grupować wykrywanych podróżnych, co może być kluczowe w wykorzystaniu danych pochodzących ze skanerów w badaniu zachowań transportowych podróżnych [33].

W [33] zwrócono, również uwagę na brak użyteczności danych ze smartfonów z systemem iOS 8.0 wzwyż, gdyż ze względu na politykę prywatności producenta mają one zmienne adresy MAC. Według artykułu na stronie Orange około 90% mobilnych urządzeń w Polsce to urządzenia z systemem Android, który umożliwia rejestrowanie adresów MAC w różnych punktach pomiarowych [134]. Zastosowanie skanerów Bluetooth w badaniu tłumów ludzi badano w [2]. W przypadku ustalania macierzy podróży wciąż nie są precyzyjne, problem adresowano w [138]. Artykuł [19] badał dedykowaną grupę studentów. Określona grupa badawcza pozwoliła na skupienie się na wykrywaniu środka transportu za pomocą skanerów Bluetooth/WiFi. W badaniu różnicowano 5 środków transportu. W [62] pokazano jak za pomocą Bluetooth wyznaczać rozkład podróży na sieć transportową. Raport [58] pokazał, że tradycyjne metody są korzystniejsze ekonomicznie w przypadku prowadzenia ciągłych pomiarów w okresie roku. Gdy pomiary kongestii planowane są na dłużej, wykorzystanie skanerów Bluetooth jest zasadne. Może więc być to metoda częściej stosowana w przyszłości. Obszarem miejskim, w którym zlecono wykorzystanie skanerów Bluetooth i prowadzi się szereg badań jest Brisbane w Australii [2], [15] wspomina, że znajduje się tam sieć 400 skanerów Bluetooth. W [66] wykorzystano sieć skanerów z Brisbane łącząc dane z danymi pochodzącymi ze smartcard data. Układ sieci skanerów Bluetooth w Brisbane jest pokazany w pozycjach [16], [60], oraz [101]. O możliwej tendencyjności danych na bazie socjoekonomicznej wspomniano w [98]. Artykułem, gdzie wykorzystywano zarówno skanery Bluetooth i dane pochodzące z GPS w celu weryfikacji rozkładu podróży jest [10].

O ile temat wykorzystania GPS przewijał się w niniejszym przeglądzie wiele razy przy okazji omawiania innych źródeł danych jako źródło pomocnicze, są również badania, które wykorzystują tylko i wyłącznie to źródło danych, np. [64]. Podobnie jak w przypadku sieci telefonii komórkowej, GPS również ma problem z lokalizacją podróżnych w metrze czy ogólnie pod ziemią [25]. Badanie to poruszyło wątek problemów związanych z pomiarami pasażerów pod ziemią, jednak jego głównym celem było zestawienie danych pochodzących z GPS telefonów komórkowych podróżnych z danymi pochodzącymi z systemów AVL (Automatic Vehicle Location) umiejscowionych w pojazdach samochodowych, które również opierają się na GPS. Badanie [110] skupiło się na określeniu jak dane pochodzące z GPS muszą być dokładne, by można było oszacować prędkość pojazdów. Można również ustalić prędkości pojazdów na poziomie całych sieci transportowych [154]. Jako odniesienie zastosowano skanery Bluetooth, pomiary prędkości były ich pierwszym zastosowaniem związanym z inżynierią ruchu. Ogólnie dane z GPS potrafią być na tyle szczegółowe, że możliwe jest ich zastosowanie w modelowaniu mikroskopowym. W [145] dane te posłużyły do oceny zmienności stylu jazdy kierowców. Nawet w przypadku małej ilości danych, na przykład w przypadku braku połączenia z różnymi stacjami bazowymi, opracowano metody oparte na GIS, które zwiększają dokładność określania lokalizacji podróżnego [45]. Za pomocą sieci neuronowej możliwe jest dokładne ustalenie rodzaju pojazdu [35].

Szczególnym typem danych GPS są dane z flot taksówek. W gęstych ośrodkach miejskich artykuł [90] proponuje metodę CWGAN-GP do przewidywania macierzy podróży przy użyciu danych z flot taxi z Chengdu i Nowego Jorku. Następny przykład zaawansowanej metody ze świeżej literatury to Randomized K-Nearest Neighbor Regression with Spherical Distance. W [116] tą metodą analizowano dane z Nowego Jorku, Porto i San Francisco w czasie rzeczywistym, przewidując czasy podróży taksówkami i wykrywając, czy przekraczano dopuszczalną prędkość w celu zwiększenia bezpieczeństwa podróży. Pozycja [148] analizowała dane GPS z Chicago Data Portal, takie jak czas trwania podróży, dystans, opłata za przejazd i miejsca odbioru i celu podróżnego, które zagregowano w rejony tożsame z cenzusem. Badanie zaproponowało szereg rekomendacji w celu ulepszenia polityki transportowej taksówkami poprzez dodatkową analizę danych socjodemograficznych podróżnych. Badanie [142] wykazało za pomocą danych z taksówek, że ten środek lokomocji odgrywa kluczową rolę w obsłudze grup społecznych znajdujących się w niekorzystnej sytuacji i ulepszaniu połączeń multimodalnych, szczególnie w przypadku krótkich podróży, obejmujących pierwszą/ostatnią milę. W [121] przeanalizowano ponad miliard podróży

obejmujący Manhattan w Nowym Jorku, a dane z GPS pozwoliły stworzyć więźbę podróży, z identyfikacją kluczowych stref wprowadzając „OD rank”. W celu weryfikacji wyników symulacji wykorzystano metodę Monte Carlo, co jest kluczowe z uwagi na wykorzystanie metody symulacji opartej na metodzie Monte Carlo w niniejszej rozprawie. Surowa baza danych dostępna jest w [170].

Innym źródłem Big Data w modelowaniu transportu jest ANPR (Automatic Number Plate Recognition). Rozpoznawanie tablic odbywa się za pomocą kamer, z których obraz przetwarzają algorytmy sztucznej inteligencji wykrywające na obrazie sam pojazd i numer jego tablicy rejestracyjnej. W zaawansowanej formie możliwe jest również rozpoznanie klasy pojazdu. Zbadanie macierzy podróży było możliwe np. w Teheranie, gdzie każdy wjazd do strefy centralnej miasta rejestruje tablice samochodów za pomocą tej metody [102]. Jest to również rozwiązanie rozważane we Wrocławiu jako metoda weryfikacji i sankcjonowania pojazdów niespełniających wymagań planowanej Strefy Czystego Transportu [165]. Rozwiązanie to może być okazją do badania tej metody w polskich warunkach i daje potencjał do wykorzystywania jej w Kompleksowych Badaniach Ruchu, do których już wykorzystywano dane z sieci telefonii komórkowej. Potrzebne więc będą metody umożliwiające łączenie danych z nich pochodzących i implementacja ich w modelach transportowych. W badaniu [141] na bazie danych z ANPR stworzono macierz podróży. W [103] dostępna jest macierz podróży z Teheranu z wartościami liczbowymi. Możliwe jest również wykorzystanie kamer CCTV w badaniu ruchu bez konieczności zaczytywania numerów tablic rejestracyjnych [133]. Oprócz Stref Czystego Transportu do wykorzystania są inne dane pochodzące z systemów ITS, takie jak dane z czujników opartych na mikrofalach do badania kongestii [130]. W [4] wykorzystano dane z systemu ITS w celu stworzenia godzinowych macierzy podróży. Źródła danych Big Data w modelowaniu transportu mogą pochodzić również z social mediów [54], mogą być to dane tendencyjne, jednak są to źródła przydatne w identyfikowaniu POI (points of interest). W badaniu [156] przeanalizowano przyswajanie informacji podawanych za pomocą mediów społecznościowych na bazie ankiet przeprowadzonych wśród podróżnych. Konstruowanie macierzy podróży za pomocą danych z mediów społecznościowych jest możliwe [29], należy jednak oddzielić dane od mieszkańców od danych od turystów, które mogą negatywnie wpłynąć na wyniki. W [28] badano jak zmieniają się zachowania podróżnych w kwestii wyboru środka transportu podczas pandemii SARS-CoV-2. Mapa ciepła macierzy podróży wraz z opisem wykorzystana w metodzie niniejszej rozprawy znajduje się w [73] oraz w rozprawie doktorskiej [152].

Kolejnym źródłem danych są dane z flot wypożyczanych hulajnóg elektrycznych, według polskiego ustawodawstwa [171] nazywane Urządzeniami Transportu Osobistego (UTO). Do tego typu danych należeć mogą również floty rowerów miejskich. Jest to środek transportu wpływający w znaczący sposób na przesunięcie modalne, badane np. w [77], gdzie wykazano chęć podróżnych do zmiany środka transportu na hulajnogę elektryczną w przypadku podróży krótkich o 46%. Zmiana jest szczególnie wysoka w grupie młodych mężczyzn z wysokimi dochodami. Badanie przeprowadzono w dwóch szwedzkich miastach. Temat dalej badano w [153], gdzie wykazano wzrost odsetka podróży odbywanych hulajnogami elektrycznymi o 3,13% za każdy 1% większej ceny za paliwo. Proporcje do współczynników z podrozdziału 4.2. pozyskano z badania [65] przy wsparciu danymi historycznymi o mobilności w Minneapolis [172]. W kwestii rowerów miejskich dane do wyliczenia współczynnika pozyskano z [136]. Przesunięcie modalne badano w [97] analizując dane GPS z trzech flot w Delft. Rezultaty na temat przesunięcia modalnego po wprowadzeniu floty rowerów miejskich w Vancouver przedstawiono również na plakacie konferencyjnym [63]. Badanie z Barcelony [126] porównujące rowery miejskie i floty hulajnóg elektrycznych wykazało, że mikromobilność w postaci hulajnóg elektrycznych wpływa znikomo na zmniejszenie liczby podróży samochodowych, a zastępuje inne formy podróży aktywnych takich jak piesze i rowerowe oraz podróże publicznym transportem zbiorowym. Do innych wniosków doszli badacze w [117], udowadniając, że rozbudowa sieci publicznego transportu zbiorowego poprawia podział modalny katalizując podróże flotą rowerów miejskich. Dane dostępne w tym badaniu również zostały wykorzystane do ustalenia współczynników w metodzie niniejszej rozprawy. Raport [99] wskazuje, że floty współdzielonych rowerów mają duży potencjał na rozrost. Elektryczne wersje rowerów pomimo 21% udziału we flocie potrafią generować 30% podróży rowerowych. Czyste bazy danych znaleziono w [173], [174] oraz [175]. Wykorzystując dane z GPS udało się pozyskać dokładne trasy przejazdów wypożyczanymi rowerami [7].

Dotychczas w literaturze były poruszane tematy łączenia źródeł. Szczególnie interesująca jest kwestia łączenia źródeł klasycznych i nowych. Artykuł [34] skupia się na problemie łączenia danych z nieobszernych źródeł (Small Data) ze źródłami Big Data. Przegląd dotyczący Big Data i Small Data przeprowadzono w [27]. Łączenie różnych źródeł w celu badania występowania kongestii poruszono w [3]. W [91] stworzono metodę łączenia danych w celu przewidywania potoków ruchu.

2.3. Charakterystyka baz danych big data i symulowanie baz danych

W tym podrozdziale zestawiono przykłady surowych danych z różnych źródeł Big Data, na które natrafiono wykonując przegląd literaturowy. Ze względu na duże rozmiary, wykorzystane macierze podróży wraz z odnośnikami literaturowymi pokazano w postaci map ciepła w podrozdziale 5.3. opisując w jaki sposób zaimplementowano je do metody. W tab. 2.1 jest przykład macierzy podróży dostępnej w literaturze, a w podrozdziale 5.3. korzystano z podobnych przykładów. W tabeli 2.2 białymi polami w tabeli przedstawiono liczbę przypadków niewykrycia podróży, która się odbyła. Niedokładność pomiaru została uwzględniona w metodzie. Tabele 2.1 do 2.10. przedstawiają surowe dane odczytane ze źródeł danych. Ich forma jest niemożliwa do implementacji do modeli transportowych i są agregowane do postaci z tab. 2.1. Ze względu na znaczną pracochłonność przetwarzania tych danych i fakt, że metoda ich przetwarzania jest już odkryta, w rozprawie wykorzystano ich przetworzoną formę do postaci macierzy źródłowo-docelowej. Pozycją dokładnie opisującą bazę danych transportowych oraz ich konstrukcję jest na przykład [106].

Tab. 2.1. Macierz podróży oparta na bazie danych pochodzącej ze skanerów Bluetooth [62]

	BMS-11	BMS-12	BMS-13	BMS-14	BMS-15	BMS-16	BMS-17	BMS-18	BMS-19	BMS-20	BMS-21	BMS-22
BMS-11	—	4773	789	437	89	95	138	131	97	77	108	73
BMS-12	4766	—	944	555	137	113	114	182	67	76	181	84
BMS-13	646	788	—	929	259	124	40	286	17	87	368	76
BMS-14	311	404	837	—	778	324	137	963	65	258	1168	190
BMS-15	85	139	307	1149	—	392	455	183	456	483	2557	176
BMS-16	52	77	118	440	309	—	1164	1353	692	580	248	940
BMS-17	58	49	16	108	287	309	—	117	846	197	158	228
BMS-18	128	165	290	1129	79	1300	215	—	77	237	108	329
BMS-19	66	44	23	161	484	768	612	111	—	551	358	510
BMS-20	29	41	125	316	349	765	383	261	460	—	325	295
BMS-21	89	149	380	1476	2064	273	296	116	269	288	—	160
BMS-22	30	33	52	190	214	1036	296	324	470	450	162	—

Tab. 2.2 Przykład luk w detekcji za pomocą skanera Bluetooth, białe pola to brak wykrycia, czerwone to wykrycie [10]

I. niewykrytych	Skuteczność detekcji	I. wykrytych urządzeń	Moto-Zoom	HTC_HD2	HTC_740	HTC_s730	HTC_s710	SE_W880i	SE_W580i	Nokia_2960	Nokia_cl_10	Nokia_cl_09	Nokia_cl_08	Nokia_cl_07	Nokia_cl_06	Nokia_cl_05	Nokia_cl_04	Nokia_cl_03	Nokia_cl_02	Nokia_cl_01	Trip ID
0	100%	18																			1
3	83%	15																			2
0	100%	18																			3
0	100%	18																			4
2	89%	16																			5
2	89%	16																			6
3	83%	15																			7
2	89%	16																			8
0	100%	18																			9
0	100%	18																			10
0	100%	18																			11
0	100%	18																			12
5	72%	13																			13
5	72%	13																			14
5	72%	13																			15
3	83%	15																			16
0	100%	18																			17
5	72%	13																			18
0	100%	18																			19
16	11%	2																			20
2	89%	16																			21
5	72%	13																			22
0	100%	18																			23
0	100%	18																			24
11	39%	7																			25
9	50%	9																			26
0	100%	18																			27
6	67%	12																			28
0	100%	18																			29
4	78%	14																			30
0	100%	18																			31
0	100%	18																			32
2	89%	16																			33
3	83%	15																			34
0	100%	18																			35
2	89%	16																			36
1	94%	17																			37
3	83%	15																			38
3	83%	15																			39
9	50%	9																			40
6	67%	12																			41
3	83%	15																			42
4	78%	14																			43
6	67%	12																			44
3	83%	15																			45
7	61%	11																			46
0	100%	18																			47
6	67%	12																			48
3	83%	15																			49
4	78%	14																			50
5	72%	13																			51
11	39%	7																			52
1	94%	17																			53
6	67%	12																			54
1	94%	17																			55
6	67%	12																			56
0	100%	18																			57
0	100%	18																			58
2	89%	16																			59
4	78%	14																			60
3	83%	15																			61
2	89%	16																			62
6	67%	12																			63
6	67%	12																			64
1	94%	17																			65
1	94%	17																			66
3	83%	15																			67
7	61%	11																			68
2	89%	16																			69
9	50%	9																			70
2	89%	16																			71
3	83%	15																			72
7	61%	11																			73
6	67%	12																			74
Średnia			12	18	16	20	23	6	11	8	15	19	9	10	13	20	14	11	22	6	Liczba niewykrytych na urządzenie
3	81%	15	16%	24%	22%	27%	31%	8%	15%	11%	20%	26%	12%	14%	18%	27%	19%	15%	30%	8%	Procent niewykrytych
			5	4	5	3	3	7	5	7	6	5	5	7	7	7	5	5	4	7	Średnia wykryć

Tab. 2.3 Przykład bazy danych ze Smart Card Data [78]

SC ID	Dzień	ID podróży	Czas źródła (min od 0 h)	Czas celu (min od 0 h)	Sekwencja zatrzymań ID	Czas pełnej podróży (min)	Czas przesiadki (min)	Całkowity czas (min)	Sekwencja ID trasy
X1	1.kwi	1697	680,2	686,54	5 → 4	6,34	0	6,34	999
X1	2.kwi	1083	557	566,9	54,371 → 24,653	9,9	0	9,9	726
X2	3.kwi	1415	898,08	905,24	12,861 → 2452 → 15,212	7,16	41,42	48,58	550 → 562
X2	4.kwi	1412	887,45	944,63	10,730 → 499 → 88	57,18	32,7	89,88	690 → 999

Tab. 2.4 Pierwszy przykład bazy danych z sieci telefonii komórkowej [27]

X	Y	ID	Czas	Czas trwania (sec)
195925	32464	J000001	82141	81
195925	32464	J000001	82456	75
195018	31555	J000002	82100	140

Tab. 2.5 Drugi przykład bazy danych z sieci telefonii komórkowej [27]

ID	Czas	Miejsce
3X35E90	1319242582	34.044162 -112.454400
3X35E90	1319242583	34.044059 -112.455550
3X35E90	1319031785	34.044392 -112.453519

Tab. 2.6 Trzeci przykład bazy danych z sieci telefonii komórkowej [68]

ID	Data rozmowy	Czas rozmowy	Czas trwania	Szerokość	Długość
AH03JAC8AAAbXtAId	20120701	09:34:19	18	23,8153	90,4181
AAH03JABiAAJKnPAA5	20120707	06:15:20	109	23,8139	90,3986
AAH03JABiAAJKnPAA5	20120707	09:03:06	109	23,7042	90,4297
AAH03JABiAAJKnPAA5	20120707	10:34:19	16	23,6989	90,4353
AAH03JABiAAJKnPAA5	20120707	18:44:53	154	23,6989	90,4353
AAH03JABiAAJKnPAA5	20120707	20:00:08	154	23,8092	90,4089
AAH03JAC5AAAdAYAE	20120701	09:15:05	62	23,7428	90,4164
AAH03JAC+AAAcVKAC	20120707	08:56:34	42	23,7908	90,3753
AAH03JAC+AAAcVKAC	20120701	18:03:06	36	23,93	90,2794
AAH03JAC5AAAdAYAA	20120701	11:15:55	12	23,7428	90,4164

Tab. 2.7 Czwarty przykład bazy danych z sieci telefonii komórkowej. Tabela dzieli się na trzy części, pierwsza dotyczy podróży od stacji bazowej do stacji bazowej, druga jest pośrednia pomiędzy rejonem transportowym a stacją bazową, trzecia to podróże zakwalifikowane do konkretnych rejonów transportowych [68]

ID	Data rozmowy	Czas rozmowy	BTS Źródłowy	BTS Celowy	Przyp. Rejon Źr.	Przyp. Rejon Cel	Ostateczny rejon Źr.	Ostateczny rejon Cel
AAH03JA	20120718	15:54	6	1	3	1	3	1
AAH03JA	20120718	16:13	1	2	1	2 lub 1	1	1
AAH03JA	20120718	16:15	2	1	2 lub 1	1	1	1
AAH03JA	20120718	18:53	1	6	1	3	1	3
AAH03JA	20120718	20:49	6	1	3	1	3	1
AAH03JA	20120718	23:41	1	6	1	3	1	3

Tab. 2.8 Przykład bazy danych z AFC [139]

	Wartość 1	Wartość 2	Wartość 3
ID	30,315	30,33	30,379
Dana	20,150,602	20,150,602	20,150,602
Numer linii	4	4	13
Numer Stacji	29	47	43
Czas	70	20	52
Pasażerowie wchodzący	385	0	74
Pasażerowie wychodzący	147	2	105
Nazwa linii	Linia 4	Linia 4	Linia 13

Tab. 2.9 Piąty przykład bazy danych z CDR [139]

Numer SIM	Kontroler Stacji Bazowej	ID komórki	ID Sektora	Rodzaj połączenia	Data i czas
4600357544****4	13	200	2	0	2013/08/19 05:45:58
4600357544****1	18	1009	2	0	2013/08/19 23:58:04
4600357544****0	14	131	1	0	2013/08/19 18:50:34

Tab. 2.10 Przetworzone dane z taksówek w Nowym Jorku [124]

Ż_Dł	Ż_Szer	C_Dł	C_Szer	Odbiór	Odległość	Napiwek	Częstość	Numer
-74	40,7	-74	40,7	Noc	Średnia	Wysoki	Niska	6074
-74	40,7	-74	40,7	Nic	Średnia	Niski	Niska	2844
-73,9	40,7	-73,9	40,7	Dzień	Mała	Wysoki	Średnia	16
-73,9	40,7	-73,9	40,7	Rano	Mała	Wysoki	Niska	14
-73,9	40,7	-74	40,7	Rano	Mała	Wysoki	Wysoka	3

Zazwyczaj bazy danych posiadają braki. W [147] pokazano jak można uzupełniać brakujące dane w Inteligentnych Systemach Transportowych za pomocą ulepszonej przez autorów wersji algorytmu kNN, który uwzględnia m.in. stochastyczną właściwość ruchu drogowego. Autorzy stworzyli algorytm, który wykazuje się największą średnią procentową dokładnością wyników oraz jedną z najlepszych wariancji procentowej zgodności wyników wśród alternatywnych rozwiązań. Możliwe więc jest wykorzystanie metody przy uzupełnianiu brakujących danych pochodzących z dużych wolumenów. W celu uzupełniania danych można zastosować również sieci Bayesowskie, które są popularne w niniejszych zagadnieniach, np. [51].

Korzystając z możliwości metody pozwalającej wykluczać poszczególne źródła danych w rozprawie zasymulowano trzy bazy danych Big Data. Bazy danych można symulować na bazie metod Monte Carlo [11], również dodając łańcuchy Markowa [23]. Przy dodaniu łańcuchów Markowa w symulacji Monte Carlo metoda ma nazwę Markov Chain Monte Carlo (MCMC), a w pozycji użyto jej do stworzenia systemu simSQL na bazie języka zapytań SQL. W [9] przedstawiono symulację związaną z Internetem rzeczy (IoT) w formacie Big Data, co jest szczególnie istotne ze względu na wykorzystanie urządzeń IoT w badaniach ruchu.

Symulacje Monte Carlo są szeroko wykorzystywane do celów symulacyjnych w badaniach w różnych dyscyplinach. Beacons Bluetooth są proponowaną metodą biletowania w publicznym transporcie zbiorowym, co pokazano w rozdziale 3.2 niniejszej rozprawy. W [37] wykorzystano sekwencyjną metodę Monte Carlo, by wykorzystać Beacons BLE do śledzenia przemieszczenia podróżnych w pomieszczeniach zamkniętych, gdy nie ma połączenia z GPS, np. na stacjach metra. Podobne podejście zostało pokazane w [115], gdzie w badaniach wykorzystano również czujniki kinetyczne jak np. żyroskopy. W pozycji [42] wykorzystano metodę MCMC do modelowania ryzyka wypadków w transporcie morskim. Tę samą metodę wykorzystano również do badania symulacji syntetycznych danych finansowych [22] oraz do stworzenia modelu prognozowanych trzęsień ziemi [75]. W literaturze są przykłady generowania danych zupełnie syntetycznych bez wykonania rzeczywistych pomiarów. W przypadku badań prętów paliwowych pod wpływem różnych temperatur, dzięki użyciu danych symulowanych metodą Monte Carlo na bazie efektu Dopplera, brak jakichkolwiek rzeczywistych pomiarów temperatury nie był problemem dla badaczy w [38]. W [120] wykorzystano model propagacji światła i metodę Monte Carlo by stworzyć bazy danych. Metoda Monte Carlo może również być użyta w przypadku weryfikacji danych, w [93] symulowana baza danych dostarczyła 300 modeli do oceny metody szybkiego obliczania odpowiedzi z detektorów. Symulacje Monte Carlo sprawdzają się również w przypadku występowania niepewnych danych [70].

W różnych dyscyplinach stosuje się również syntetyczne bazy danych, które nie zawsze wykorzystują metodę Monte Carlo. W tym miejscu warto zaznaczyć, że poniższy paragraf powstał bez używania słów kluczowych związanych z Monte Carlo, pomimo tego że w części badań metoda została wykorzystana, co może wskazywać na słuszność wybrania tej metody w niniejszej rozprawie doktorskiej. W [92] pokazano przykład budowania syntetycznej bazy danych. Tego typu bazy pozwalają rozwiązać nie tylko problem dostępności danych, lecz

również mogą rozwiązać problemy związane z ograniczeniami związanymi z kwestiami prywatności [118]. Jest to szczególnie istotne w dyscyplinie autora rozprawy, ze względu na dane które często zawierają informacje wrażliwe, takie jak na przykład adresy zamieszkania. Syntetyczną bazę danych wykorzystano na przykład w celu wyszkolenia sieci neuronowej stworzonej do łamania kodów Captcha [86]. Tam również wykorzystano metodę MCMC. Inny przykład syntetycznej bazy danych opartej na metodach Monte Carlo wykorzystano w [127], gdzie autorzy wygenerowali struktury chemiczne oraz nowe cząsteczki podobne do leków. W [146] użyto symulacji bazującej na metodzie Monte Carlo do komputerowego wspomagania planowania syntezy.

W powyższych akapitach pokazano szerokie zastosowanie symulowanych baz danych z wyszczególnieniem badań gdzie do tego celu wykorzystano metodę Monte Carlo i jej pochodne. W niniejszej rozprawie przedstawiono jej nowe zastosowanie w inżynierii lądowej, geodezji i transporcie. Do tej pory w transporcie używano metody Monte Carlo w celu przeprowadzenia analizy systemu transportowego z brakiem danych historycznych [59]. Metoda MCMC może z powodzeniem uzupełniać luki w danych systemów ITS [105], [109]. Symulacja Monte Carlo może przydać się również w badaniu kosztów podróży. Wykorzystano ją do rozwiązywania Fuzzy Transportation Problem (FTP) w [100]. [112] wykorzystuje symulację Monte Carlo do symulacji zróżnicowanych scenariuszy mających na celu rozwiązanie niepewności związanej z czasem podróży i pobytu, aby umożliwić obliczenie wyboru najkrótszej ścieżki w stochastycznych warunkach pogodowych. W [6] badacze testowali amerykańską sieć kolejową i autostradową pod kątem grafów typu „Small-world”. Metody MCMC można również użyć w celu przewidywania występowania gołoledzi na autostradach [39]. Wykorzystanie metody Monte Carlo pozwoliło w [74] na zidentyfikowanie wrażliwych elementów sieci wirtualnego miasta. Co ważne, podobnie jak w niniejszej rozprawie sieć transportowa ośrodka miejskiego również była symulowana. W [144] podejmowano racjonalne decyzje dotyczące infrastruktury publicznego transportu zbiorowego wykorzystując symulację Monte Carlo do weryfikacji solidności rozwiązania optymalizacyjnego i w celu przeprowadzenia testów losowych. Przykładem zastosowania metod Monte Carlo bezpośrednio związanym z infrastrukturą transportową, jednak nie związanym z modelowaniem podróży, może być na przykład wykorzystanie tej metody wraz z metodami uczenia maszynowego do oceny stateczności skarp co pokazano w [5].

Symulacje tego typu mogą pozwolić na badanie przyszłych transportowych źródeł Big Data pochodzących z IoT, BIMu, cyfrowych bliźniaków (Digital Twins – DT), czy GISu 3D [47] w celu poprawnego zamodelowania sieci transportowej ośrodków miejskich. DT mogą również dostarczać dane w czasie rzeczywistym [84]. Bliźniaki sieci transportowej są zaproponowane również w [69], a ich użycie w modelowaniu transportu podsumowuje „świeży” przegląd [104]. DT mogą być wykorzystywane również w systemach ITS. Bezpośrednie wykorzystanie zaproponowano w [12]. Tego typu systemy wpisujące się w ideę Smart City w przyszłości mogą korzystać z danych pochodzących od pojazdów autonomicznych [122], [123], gdy te osiągną w ruchu publicznym wyższe poziomy automatyzacji. Systemy V2I są w stanie dostarczyć wiele znaczących informacji do modeli transportowych ze względu na to, że zawierają dane zawierające czas i lokalizację (spatiotemporal data) [151].

2.4. Identyfikacja luki badawczej

Przegląd literatury zidentyfikował, że badania na temat Big Data dotyczące poszczególnych źródeł są zaawansowane w różnym stopniu. Niemniej każde z nich prezentuje wysoki poziom szczegółowości. Na bazie przeglądu aktualnych badań zauważono, że liczba źródeł znacząco przyrasta. Jednak w przypadku łączenia źródeł stosowano metody ad hoc, dostosowane do konkretnych ośrodków miejskich, lub źródeł danych.

Ze względu na brak usystematyzowanej metodyki łączenia baz danych pochodzących ze zróżnicowanych źródeł Big Data, opierających się głównie na współczesnych technologiach (sieć komórkowa, Bluetooth, Smart Card etc.) należy stworzyć metodę ich łączenia, która nie będzie rozwiązaniem ad hoc jak w aktualnych badaniach. Te dane będą możliwe do implementacji w modelach transportowych. Niepokój może również budzić coraz większa dokładność prywatnych danych, jednak aktualne badania proponują systemy zabezpieczeń danych oparte na Blockchain [95].

Aktualnie badania ruchu, oraz modele transportowe mogą mieć braki. Obecne metody w niektórych przypadkach są nieskuteczne. We wnioskach z Warszawskich Badań Ruchu z 2015 roku [55] wykazano małą wiarygodność badań ankietowych przeprowadzanych przez Internet. Początkowa weryfikacja odrzuciła aż 38% formularzy. Dalsza analiza wykazała rozbieżność

między badaniem wiosennym i jesiennym i ostatecznie nie wykorzystano badań internetowych do budowy modelu ruchu.

Niniejszym zauważono potrzebę stworzenia usystematyzowanej metody łączenia zróżnicowanych źródeł Big Data w celu wykorzystania jak największej liczby dostępnych informacji w modelach transportowych i pozbycia się tendencyjności danych, jeśli pochodzą tylko z jednego źródła. Ze względu na obszerność prac rozprawa skupia się tylko na rozkładzie przestrzennym ruchu, nawiązując do drugiego etapu z klasycznego czterostadiowego modelu ruchu. W rozprawie poruszano również elementy generowania ruchu przez rejony transportowe oraz atrakcyjność rejonów, więc rozprawa w części dotyczy również pierwszego kroku klasycznego modelu.

3. Cele i układ pracy

3.1. Uwagi wstępne, przedmiot pracy

Przedmiotem niniejszej rozprawy doktorskiej jest analiza dziewięciu źródeł danych wpisujących się w definicję Big Data. Są to: dane z sieci telefonii komórkowej, sondowania pojazdów kołowych, automatycznych systemów liczenia pasażerów (Automatic Passenger Counting – APC), kart miejskich (Smart Card data), GNSS (najczęściej wykorzystując system GPS), automatycznej rejestracji numerów tablic rejestracyjnych (Automatic Number Plate Recognition – ANPR), mediów społecznościowych (Social Media), Taxi oraz flot rowerów miejskich i Urzędzeń Transportu Osobistego (UTO). Analiza powyższych źródeł danych ma na celu stworzenie metody łączenia zróżnicowanych źródeł i ich implementacji w modelach transportowych. Ze względu na pracochłonność rozprawy zdecydowano się na poruszenie w badaniach jednego etapu w ujęciu klasycznego modelu czterostadiowego – rozkładu przestrzennego podróży. Z tego powodu badania dotyczą przede wszystkim macierzy podróży źródłowo-docelowych. Pozostałe etapy zostaną przebadane w ramach kontynuacji badań w przyszłości, co opisano szerzej w Rozdziale 8. Ze względu na brak wszystkich dziewięciu źródeł danych w jednym istniejącym ośrodku miejskim skorzystano z sześciu istniejących baz danych i przeskalowano je do rozmiarów zasymulowanego ośrodka miejskiego. Trzy następne bazy danych zasymulowano na bazie map ciepła, przykładów i opisów literaturowych. W ramach niniejszej rozprawy doktorskiej stworzono również drugą metodę. Jest to metoda symulowania ośrodka miejskiego oparta na sposobie symulowania danych w metodzie Monte Carlo. Zasymulowanie ośrodka zapobiegnie wpływowi charakterystycznych cech istniejącego, wybranego ośrodka na badania. Z tego powodu celowo nie wykorzystano aglomeracji wrocławskiej jako „naturalnego” przykładu, choć mogła by być oczywistym ze względu na jej znajomość przez autora i dostęp do wyników Kompleksowych Badań Ruchu. Metoda ta ma również potencjał na dalszy rozwój badawczy, pozwala na dosymulowanie brakujących elementów w przypadku tworzenia Kompleksowych Badań Ruchu dla miast z niewystarczającą dostępną liczbą danych wejściowych.

3.2. Cele pracy

Celem pracy jest stworzenie metody agregacji danych ze zróżnicowanych źródeł danych Big Data w celu umożliwienia implementacji do modeli transportowych. Mając na uwadze redukcję możliwych tendencyjności danych obrano za cel stworzenie drugiej metody,

polegającej na zasymulowaniu realistycznego (pod kątem rozkładu podróży) ośrodka miejskiego.

Sformułowano dwie tezy badawcze.

1. Wykorzystanie źródeł Big Data w modelowaniu podróży pozwala uzyskać dokładniejsze wyniki.
2. Wartościowe jest łączenie wielu źródeł Big Data, ponieważ pozwala to uzyskać efekt synergiczny w uzupełnianiu brakujących lub eliminacji nieprawidłowych danych.

3.3. Układ pracy

Rozprawa składa się z ośmiu rozdziałów oraz spisu literaturowego. Poprzedzona jest wykazem podstawowych pojęć i skrótów, oraz zakończona streszczeniem w języku polskim i angielskim. Do rozprawy dodano w formie załączników algorytm metody i macierze podróży.

Rozdział pierwszy jest wprowadzeniem do badanego zagadnienia, wprowadza w tematykę modelowania podróży oraz Big Data. Zawarto w nim również motywację do podjęcia badań.

Rozdział drugi przedstawia aktualny przegląd literaturowy. Na wstępie opisano na czym polega modelowanie podróży odnosząc się do „klasycznych” pozycji literaturowych. W tym rozdziale opisano aktualny stan wiedzy dotyczący wykorzystania Big Data w modelowaniu podróży ze szczególnym uwzględnieniem pozycji, które dotyczą dziewięciu źródeł wykorzystanych do weryfikacji metody. W rozdziale zawarto również opis literatury zawierającej zagadnienia związane z drugą metodą rozprawy, czyli symulowaniem oraz wykorzystaniem symulacji Monte Carlo w tworzeniu i symulowaniu baz danych. Pod koniec rozdziału opisano luki badawcze zidentyfikowane podczas wykonywania przeglądu.

W rozdziale trzecim przedstawione są uwagi wstępne, założenia i przedmiot pracy. Zawarto w nim także cele pracy i tezy badawcze oparte na przeglądzie literaturowym. Następnie opisano układ i treść rozprawy

Rozdział czwarty przedstawia pierwszą z dwóch metod stworzonych w rozprawie. Jest to metoda uwzględniania wielu źródeł Big Data w modelach transportowych. Pokazano, w jaki

sposób metoda łączy macierze podróży z Big Data oraz zaprezentowano wzory. Dokładnie opisano sposób uwzględnienia w metodzie każdego z dziewięciu wykorzystanych źródeł. Pokazano sugerowane uniwersalne wartości współczynników potrzebnych do uwzględnienia macierzy podróży z konkretnych źródeł danych.

Rozdział piąty opisuje drugą metodę stworzoną w rozprawie. W celu minimalizacji wpływu tendencyjności danych na wyniki zdecydowano się stworzyć metodę symulacji ośrodka miejskiego bazowaną na sposobie symulowania wartości liczbowych w metodzie Monte Carlo. Najpierw opisano podstawy teoretyczne symulacji, czyli w jaki sposób i przy jakich założeniach symulowano konkretne wartości liczbowe w utworzonej macierzy podróży o rozmiarze 100x100. Następnie pokazano poglądowy model zasymulowanego środka miejskiego stworzony w oprogramowaniu PTV Visum. Model stworzono przede wszystkim w celu wizualizacji i porównania wyników badań w różnych wariantach. Na końcu rozdziału pokazano w jaki sposób i skąd pozyskano macierze podróży użyte w badaniach. Oprócz sześciu macierzy, których dokładne wartości pochodzą ze źródeł literaturowych, stworzono trzy macierze na bazie opisów w literaturze i zawartych w niej map ciepła.

W rozdziale szóstym przedstawiono wyniki rozprawy. Rozpoczęto od opisanie wyników wykorzystania wszystkich dziewięciu źródeł Big Data za pomocą pierwszej metody. Metoda jest elastyczna, więc umożliwia wykorzystanie tylko części z dostępnych źródeł. W celu sprawdzenia, czy łączenie wielu źródeł jest zasadne stworzono warianty wykorzystujące źródła danych, które powinny dawać najlepsze wyniki. Za pomocą menedżera scenariuszy w programie PTV Visum dokonano porównania wariantów do modelu bazowego. Jest on użyciem pełnej metody wykorzystującej dziewięć dostępnych źródeł z ośmioma wariantami „konkurencyjnymi” wykorzystującymi część ze źródeł danych. Warianty tworzą korzystając z metody pierwszej, stworzonej w rozprawie. Ostatni z wariantów stworzono na bazie doświadczeń z poprzednich wariantów, by wyselekcjonować najlepsze źródła i porównać z wynikami modelu bazowego. Każdy z wariantów składa się z opisu i czterech zrzutów ekranu z PTV Visum pokazujących wartości w obszarze centrum aglomeracji oraz na odnogach aglomeracji w wartościach podróży, a następnie wartościach procentowych zmian w stosunku do modelu bazowego. Warianty podsumowane są mapą ciepła powstałej macierzy podróży. Następnie zestawiono porównanie wariantów w tabelach w wartościach sumy liczby podróży, różnicy procentowej oraz odchylenia standardowego. Na sam koniec rozdziału przeanalizowano wyniki pod kątem spełnienia tez badawczych rozprawy.

Rozdział siódmy to rozdział zawierający dyskusję. Omawia zalety metody i elementy które zdaniem autora wymagają dalszych prac. Rozdział zawiera również ogólny opis rezultatów.

Rozdział ósmy to wnioski końcowe wraz z podsumowaniem realizacji celów rozprawy. Omówiono w nim również kierunki dalszych planowanych prac badawczych związanych z wykorzystaniem stworzonych metod w przyszłych badaniach.

Na końcu rozprawy zawarto spis literatury, rysunków, tabel i załączników. W załącznikach dodano pseudokod pierwszej metody, wersję zapisu w MS Excel oraz pozyskane za jej użyciem macierze podróży z danymi liczbowymi. Te same macierze są również przedstawione w postaci map ciepła w rozdziale szóstym.

4. Metoda agregacji danych z dużych wolumenów, zróżnicowanego pochodzenia i struktury w celu udoskonalenia i tworzenia modeli transportowych

4.1. Metoda łączenia macierzy podróży z Big Data

Rozprawa doktorska skupia się na badaniu rozkładu przestrzennego podróży. Badania ograniczono więc do macierzy podróży. W ramach rozprawy zbadano jak wpływ wielu macierzy ze zróżnicowanych źródeł Big Data wpłynie na dokładność danych. Problemem jest połączenie zróżnicowanych macierzy wypełniających różne części całej sieci połączeń. W celu łączenia zróżnicowanych jakościowo i objętościowo macierzy podróży stworzono metodę łączenia danych. W rozprawie niniejsza metoda tworzy całą nową macierz podróży, jak i pozwala na punktowe uzupełnianie brakujących połączeń w macierzy lub tylko na zastępowanie danych o nierealistycznych wartościach – znacząco zaniżonych lub zawyżonych liczb podróży na danych relacjach.

Metoda polega na stworzeniu osobnej macierzy podróży na bazie dostępnych danych. Uwzględnia ona dane z dowolnej liczby źródeł Big Data. Może uwzględniać również dane ze źródeł tradycyjnych. W rozprawie poruszono zróżnicowane źródła, więc niektóre ze źródeł nie posiadały wartości w dużej liczbie relacji. Jeżeli takie braki wartości występowały, uwzględniono taki przypadek w sposób niewpływający na zaburzenie wyniku zaniżając go wartością zerową. Wersja wzoru bez uwzględniania przypadku zerowego i ze źródłami wykorzystanymi w rozprawie została przedstawiona we wzorze 4.1.

$$M_{ij} = (M_{ij}^{GSM} * \frac{1}{W_{GSM}} + M_{ij}^{SM} * \frac{1}{W_{SM}} + M_{ij}^{BT} * \frac{1}{W_{BT}} + M_{ij}^{APC} * \frac{1}{W_{APC}} + M_{ij}^{SC} * \frac{1}{W_{SC}} + M_{ij}^{GPS} * \frac{1}{W_{GPS}} + M_{ij}^{ANPR} * \frac{1}{W_{ANPR}} + M_{ij}^T * \frac{1}{W^T} + M_{ij}^{UTO} * \frac{1}{W_{UTO}}) / 9 \quad (4.1)$$

Gdzie M_{ij} to liczba podróży z rejonu i-tego do rejonu j-tego w stworzonej macierzy. M_{ij}^{GSM} to liczba podróży z macierzy podróży pozyskanej z sieci telefonii komórkowej. Analogicznie następne indeksy górne przedstawiają kolejne źródła użyte w rozprawie. SM to dane z mediów społecznościowych, BT ze skanerów Bluetooth, APC to systemy automatycznego zliczania pasażerów w pojazdach publicznego transportu zbiorowego, SC to systemy inteligentnych kart miejskich, GPS to dane z aplikacji, z których da się pozyskać dane geolokalizacyjne z systemów

opartych na GNSS, ANPR to systemy rejestracji tablic samochodowych, T to dane z Taxi i aplikacji firm świadczących usługi taksówkowe, UTO to uproszczenie skrótu, w tym przypadku ta macierz zawierała syntezę danych z rowerów miejskich i floty wypożyczanych hulajnóg. Każdą z macierzy przemnożono przez odwrotność współczynnika przeliczeniowego. Uwzględnia on wszystkie kluczowe parametry danego źródła danych, takie jak na przykład podział modalny lub dane socjodemograficzne próby osób, których podróże zostały wykryte i ekstrapoluje tę liczbę podróży na całą populację. Jest to odwrotność ze względu na formę uwzględniania danych w programie z którego korzystano, współczynnik to odsetek podróży uwzględnianych w danym źródle. Niezastosowanie odwrotności zaniżyłoby wyniki. Jeśli jest to komórka pusta lub zerowa to źródło nie jest uwzględniane w równaniu. Konkretnie wartości współczynników opisano szczegółowo w podrozdziale 4.2. Forma ułamkowa została wyciągnięta we wzorze celowo ze względu na opisane poniżej dalsze rozbudowanie wzoru i jego pierwotną formę w postaci kodu programowania. Metoda zakłada, że dla różnych ośrodków miejskich i państw odsetki te będą różne. W przypadku poszczególnych źródeł różnić się będzie np. udział w rynku od dostarczyciela danych. We współczynnikach obecnych we wzorze użytym w rozprawie, uwzględniony jest również podział modalny, odsetek użytkowników danej kategorii, odsetek danych możliwych do zarejestrowania i wykorzystania, odsetek taboru wyposażonego w urządzenia dające dane oraz odsetek wykrywanych obiektów. Szczegółowo, z podaniem wartości liczbowych dla każdego ze źródeł, opisano to w podrozdziale 4.2. O ile dla symulowanego ośrodka przyjęto konkretne wartości współczynników, metoda jest uniwersalna i pozwala na wprowadzenie danych dla dowolnego ośrodka miejskiego.

Uniwersalną postać wzoru do wdrożenia w innych przypadkach, rozumianych jako inny zbiór dostępnych źródeł Big Data, pokazano jako prostą sumę we wzorze 4.2. Gdzie m odnosi się do dowolnego ze źródeł danych, a N to liczba wykorzystanych źródeł. We wzorze kluczowe jest uwzględnianie wykorzystanej liczby źródeł w postaci N , jeżeli uwzględniono konkretne źródło to mianownik zwiększany jest o jeden. Pozwoliło to uniknąć sytuacji, w której doszłoby do zaniżenia wyniku przez uwzględniane w rozprawie źródła cząstkowe, ponieważ zawierają relacje bez wykrytych podróży w komórkach macierzy i bez uwzględnienia tego we wzorze do średniej, byłyby wliczane jako zero podróży. W przypadkach gdy np. tylko trzy źródła zawierały dane dotyczące konkretnej pary rejonów transportowych, wynik byłby zaniżony trzykrotnie, co w rzeczywistości przełożyłoby się na niedoszacowanie liczby podróży,

niedowymiarowanie infrastruktury lub ustalenia zbyt małej liczby kursów publicznym transportem zbiorowym.

$$M_{ij} = \sum_{m=1}^N \frac{M_{ij}^m}{N * W^m} \quad (4.2)$$

Podczas badań komórki niezawierające danych wykluczono formułą IF. W zapisie matematycznym dodano wykluczenie i przedstawiono we wzorze 4.3.

$$M_{ij} = \sum_{m=1}^N \frac{M_{ij}^m}{N * W^m} \text{ dla } M_{ij}^m \in \mathbf{R}_+ \quad (4.3)$$

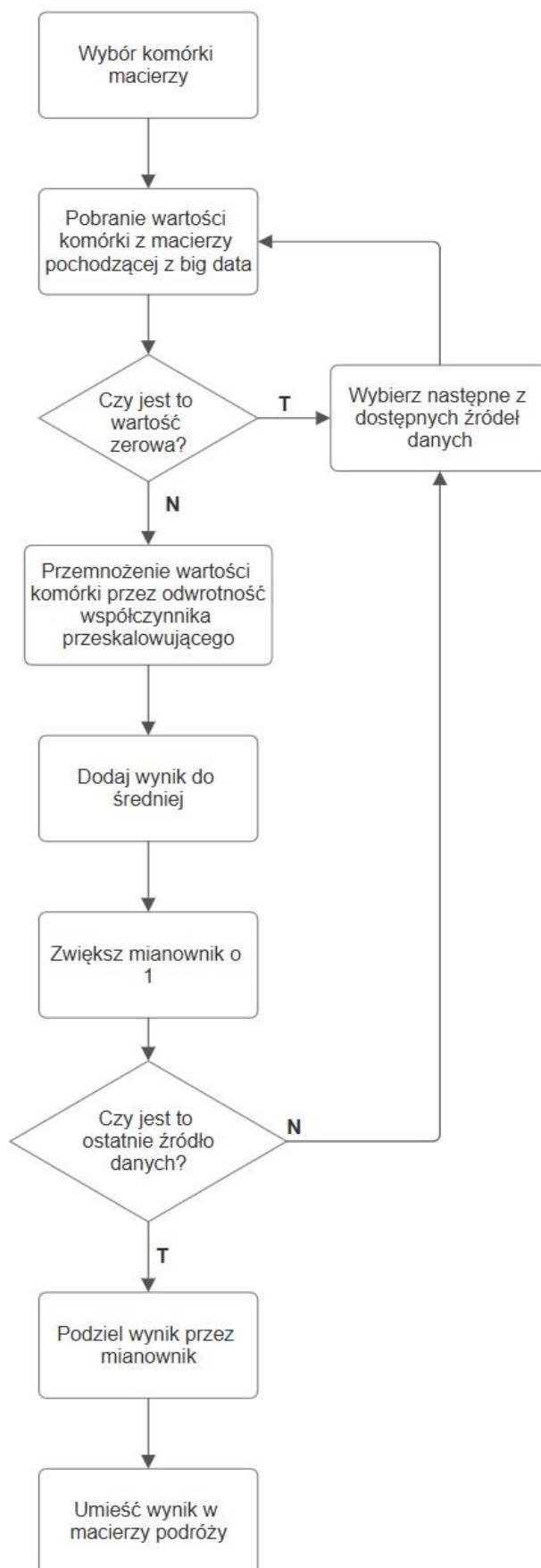
Rozsądne wydaje się uwzględnienie źródeł nadając im dodatkowe wagi. Niektóre ze źródeł będą dawać dane znacznie niższej jakości, niż te lepiej przebadane i znane w literaturze. Wymagałoby to, jednak dokonania kolejnych założeń mogących zaburzać wyniki. Metodę przetestowano w wersji bez nadania wag dla poszczególnych źródeł. Ze względu na obiecujące wyniki, opisane w Rozdziale 6., zrezygnowano z planowanego dodania wag dla poszczególnych macierzy. Źródła danych użyte w rozprawie potrafią być też niekompatybilne. Niektóre źródła, czyli jak w przypadku niniejszej rozprawy, dane pochodzące z publicznego transportu zbiorowego takie jak np. systemy zliczania pasażerów i dane z systemów automatycznego rejestrowania numerów tablic rejestracyjnych dają tylko wycinek macierzy podróży lub tylko część jej zawartości, a jej większa część nie daje rezultatów. Ponadto w części szczytkowych źródeł, takich jak np. ANPR występowały dane mocno zawyżone, które widać w poszczególnych komórkach w macierzy podróży wykorzystującej pełną metodę, przedstawioną w formie mapy ciepła w rozdziale 6.2.1 na rys. 6.5, gdzie szerzej omówiono wyniki. W tym przypadku to źródło danych powinno mieć zaniżoną wagę by zminimalizować wpływ zawyżający rezultat. Jak powyżej wspomniano, na etapie wstępnej analizy, całościowe wyniki były na tyle bliskie w stosunku do modelu bazowego aglomeracji, że zrezygnowano z wprowadzenia wag, niemniej takie prace planuje się w przyszłych badaniach. W przypadku badania danych rzeczywistych zakłada się kontynuację badań i uwzględnianie macierzy z Big Data w ten sam sposób. Źródła mniej wiarygodne będą mieć nadawaną niższą wagę, by ograniczyć występowanie relacji pomiędzy rejonami z zawyżoną lub zaniżoną liczbą podróży.

W przypadku niniejszej rozprawy każda z macierzy jest uwzględniana z tą samą wagą, jednak proporcjonalnie do spodziewanego odsetka podróży z każdego ze źródeł. „Pełną” wersję wzoru wraz z uwzględnieniem wiarygodności źródeł pokazano we wzorze 4.4. będącym rozbudowaną formą wzoru 4.3. We wzorze pojawiła się waga w postaci współczynnika wiarygodności W_w .

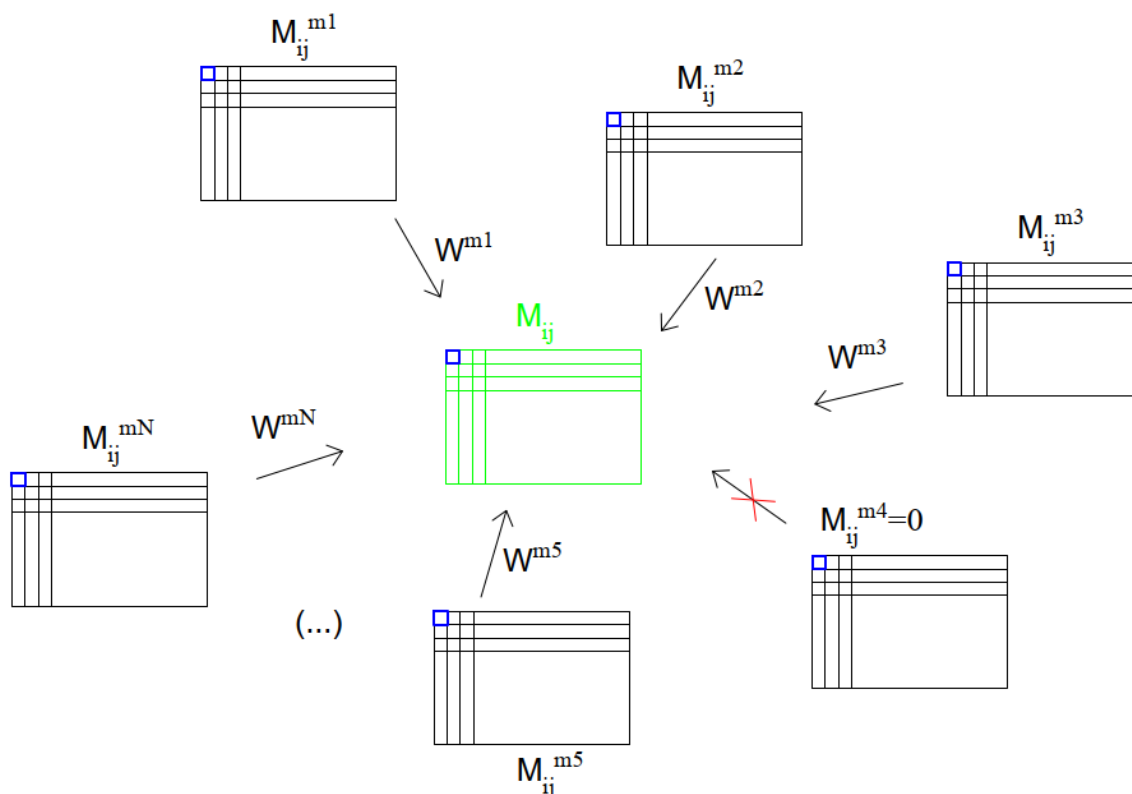
$$M_{ij} = \sum_{m=1}^N \frac{M_{ij}^m}{N * W^m} * W_w^m \text{ dla } M_{ij}^m \in R_+ \quad (4.4)$$

W celu zwiększenia czytelności wzory 4.1 do 4.3 przepisano w formie jednego pseudokodu i załączono do rozprawy w Załączniku 1. Metoda została wykorzystana w arkuszu kalkulacyjnym MS Excel, ale można ją zaimplementować w innych środowiskach i programach. W Załączniku 2 przedstawiono formuły z Excela w formie jednowierszowej oraz pod spodem rozbijając na osobne formuły IF. We wzorach załącznika 2 komórka B2 oznacza podróże z rejonu 1 do rejonu 1, ponieważ to pierwsza komórka macierzy podróży. Wartości liczbowe macierzy zaczynają się od drugiej kolumny i wiersza arkusza. Fragment wywoływanych arkuszy rozpoczynających się od „OD”, np. „OD GSM” oznacza odwołanie do arkusza zawierającego macierz podróży pozyskaną ze źródła Big Data z sieci telefonii komórkowej GSM. Analogicznie „OD SocialM”, macierz z mediów społecznościowych, „OD BT” ze skanerów Bluetooth, „OD APC” z systemów zliczania pasażerów w komunikacji miejskiej, „OD SmartCard” z kart miejskich używanych w publicznym transporcie zbiorowym, „OD GPS” danych z GPS, „OD ANPR” z systemów automatycznego rejestrowania numerów tablic rejestracyjnych, „OD Taxi” danych z taksówek, „OD UTO” danych z Urzędzeń Transportu Osobistego oraz rowerów miejskich. Odwołania do różnych komórek w arkuszu o nazwie „Założenia i wsp” to uwzględnianie współczynnika przeliczeniowego zawartego w wywoływanych poszczególnych komórkach.

Formułę uzupełniania danych przedstawiono również w formie schematu blokowego na rys. 4.1. W tym przypadku nie pokazywano ścieżki postępowania dla każdego ze źródeł Big Data z osobna, a zrobiono pętlę podobnie jak pokazano we wzorze 4.3. W odróżnieniu od poprzednich opisów zastosowano formę opisu słowną. Formę graficzną metody przedstawiono na rys. 4.2.



Rys. 4.1 Formuła uzupełniania danych w formie schematu blokowego



Rys. 4.2 Graficzna forma metody. Utworzona macierz w kolorze zielonym. Uzupełniana komórka w danym kroku kolorem niebieskim

4.2. Uwzględnienie źródeł Big Data w metodzie

W poprzednim podrozdziale we wzorach użyto współczynnika przeliczeniowego. W niniejszym podrozdziale pokazano w jaki sposób go przyjęto i jak powinno się go przyjmować, stosując metodę w innych przypadkach. W rozprawie doktorskiej użyto dużej liczby możliwych do wykorzystania źródeł Big Data, wytypowano ich dziewięć. W literaturze na czas wykonywania przeglądu literaturowego dla jednego ośrodka miejskiego znaleziono maksymalnie trzy źródła Big Data w połączeniu ze źródłem „klasycznym”. Ze względu na niemożliwość do pozyskania liczbę źródeł dla jednego ośrodka, jak i ograniczenia finansowe, źródła były pozyskiwane z różnych miejsc. Na potrzeby rozprawy założono konkretne wartości procentowe, jednak metoda każdorazowo powinna być dostosowana do charakterystyki pozyskanych danych i badanych ośrodków miejskich. Gdy nie da się pozyskać współczynnika dostosowanego do badanego ośrodka, zalecane jest przyjęcie wartości podanej w niniejszej rozprawie doktorskiej. Poniżej przedstawiono sposób przyjmowania współczynników

przeliczeniowych dla każdego z wykorzystanych w rozprawie źródeł Big Data z konkretnymi przyjętymi wartościami liczbowymi.

Dane z sieci GSM występują w pierwszym członie wzoru 4.1 z poprzedniego podrozdziału. Zaletą tego źródła jest pokrycie nim obszaru niemal całego kraju w przypadku krajów rozwiniętych. W dodatku stosunkowo gęsta sieć stacji bazowych dokonuje podziału na rejony, który jest możliwy do odniesienia do rejonów transportowych. Uwzględniając dane pochodzące z sieci telefonii komórkowej uwzględniono po prostu odsetek danych od operatora, który w Polsce sięga zazwyczaj kilkunastu procent. W tym przypadku założono 16%, niemniej trzeba zaznaczyć, że mogą być to znacznie większe liczby, przekraczające 50%, jak wskazano w przeglądzie literaturowym. W literaturze znaleziono również podział na poszczególne grupy wiekowe, jednak nie wykorzystano podziału w metodzie. Podział na poszczególne grupy wiekowe może być przydatny gdy metodę rozszerzy się na pozostałe etapy modelu transportowego, odnosząc się tu do klasycznego modelu czterostadiowego. W zakresie niniejszej rozprawy doktorskiej nie rozszerzano modelu na inne etapy modelu.

Bluetooth mierzony jest skanerami. Daje to pełną dowolność w ich rozmieszczeniu, jednak będą to punkty pomiarowe, a nie pokrycie całego obszaru. Liczbę tych punktów ogranicza również cena skanera. W przypadku Bluetooth należy rozróżnić jaki typ urządzeń się wykrywa. Zgodnie z przeglądem literaturowym odmienny odsetek wykrywanych urządzeń będzie w przypadku, gdy wykrywamy zestawy głośnomówiące w samochodach, a inny gdy wykrywamy telefony komórkowe. W przypadku samochodów założono 62% wykrywalności. Zgodnie z najnowszymi badaniami z przeglądu literatury tą wartość procentową pomniejszono o zakładany podział modalny. W tym przypadku dla samochodów wynosi 40%, a współczynnik 24,8%. W przypadku telefonów 10% użytkowników korzysta z Bluetooth mając go cały czas włączonego w trybie widoczności. Odsetek pomniejszono o użytkowników systemu iOS, który ma zmienny adres MAC urządzeń, co szerzej opisano w przeglądzie literaturowym. W Polsce z iOS korzysta 13,8% użytkowników. Sumarycznie daje to 8,62% wykrywanych telefonów komórkowych. W rozprawie użyto macierzy z danymi pochodzącymi z pojazdów, więc użyto współczynnika odpowiedniego dla nich, czyli 24,8%.

Systemy zliczania pasażerów mają tę zaletę, że obecnie jest w nie wyposażona większość floty pojazdów publicznego transportu zbiorowego. Pozwala to na pomiar ciągły, możliwy do pozyskania od operatorów. Minusem tego źródła jest brak pomiaru w innych źródłach

transportu oraz nierozróżnianie podróży przesiadających się. Według systemu podróży, który przesiadł się z tramwaju na autobus, będzie liczony jako dwie różne podróże, co daje możliwość błędnej interpretacji liczby podróży publicznym transportem zbiorowym w podziale modalnym. Źródło danych z systemów zliczania pasażerów jest zależne od dwóch czynników - odsetka podróży wykonywanych transportem publicznym oraz odsetka taboru wyposażonego w systemy zliczania pasażerów. W przypadku rozprawy założono 30% podróży wykonywanych transportem publicznym oraz 90% taboru wyposażonego w systemy, czyli współczynnik dla tego źródła danych wynosi 27%. Przyjęto wysoki odsetek taboru wyposażonego w systemy zliczania pasażerów ze względu na możliwość montażu nawet w starszym taborze. Założono jednak, że nie każdy pojazd należy do floty miejskiego przewoźnika, a część połączeń mogą wykonywać podwykonawcy lub gminy ościennie, od których nie uzyska się danych.

Źródło danych pochodzących z UTO i rowerów miejskich pozwala na uwzględnienie w badaniach mikromobilności. Jednak w przypadku Big Data nie są możliwe do zmierzenia, jeśli są to urządzenia prywatne. Do badań założono flotę pojazdów wypożyczanych. Na bazie literatury zastosowano dość niestandardowy sposób wyliczenia. Wzorowano się na danych pozyskanych z Minneapolis. W literaturze znaleziono bazę danych z dwóch okresów, po 152 dni i 213 dni, co sumarycznie dało równo 365 dni pomiarów. Wszyscy podróżni w badanym okresie pokonywali średnio 2,37 miliarda mil dziennie, według statystyk pojedynczy podróżny pokonywał średnio 5567 mil, co dziennie dawało 425723 podróże w tym mieście. Uwzględniając pozyskane z literatury średnio 5266 podróże pojazdami z floty wypożyczanej, wyliczono odsetek w wysokości 1,237% podróży i taki współczynnik przyjęto. Należy też zwrócić uwagę, że operator może nie działać w trybie sezonowym, czyli w miesiącach zimowych flota nie jest wypożyczana.

Dane z GPS uwzględniono w metodzie podobny sposób. W tym przypadku wzorując się na badaniach z Glasgow [131], gdzie aglomeracja ma około 1,8 miliona mieszkańców, a z aplikacji z której odczytano dane z GPS korzysta 81 tysięcy użytkowników. Daje to 5,06% potencjalnie wykrytych podróży. Dokonano jednak szerszego przeglądu i na bazie dwóch badań wytypowano liczbę 37%, która pokrywała się na granicach dwóch zakresów procentowych wykrywalności w dwóch badaniach z przeglądu literatury, tj. w przypadku jednego badania wykrywano od 37% podróży, a w drugim do 37% podróży. Zaletą tego źródła danych jest możliwość pomiarów w czasie rzeczywistym ze względu na ciągłe odświeżanie pozycji

użytkownika. Wadą jednak są zakłócenia odbioru sygnału w miejskich kanionach (urban canyon), odcinkach pod ziemią (metro, tunele), estakadach, oraz spowodowane celowym zagłuszaniem, na przykład podczas konfliktu hybrydowego lub ćwiczeń wojskowych.

Systemy inteligentnych kart miejskich rozwiązują problem przesiadek w przypadku zastosowania systemu check-in/out. Wadą jest skonstruowanie systemu, gdzie karta jest używana tylko podczas kontroli biletowej biletów okresowych. Dla kart miejskich w literaturze znaleziono przykład, gdzie korzysta z nich aż 99% podróżnych przy podziale modalnym dla publicznego transportu zbiorowego na poziomie 30% w badanej aglomeracji, co dało współczynnik równy 29,7%. Jest to założenie optymistyczne i zależne od systemu biletowania w transporcie publicznym, jednak jest możliwe i przyjęto je w symulowanym ośrodku miejskim z niniejszej rozprawy.

Dane z mediów społecznościowych w literaturze określane są jako Location-based Social Media Data. Są to dane z mediów społecznościowych, które da się zlokalizować i dotyczą czasu i miejsca przebywania użytkowników mediów społecznościowych. Współczynnik dla mediów społecznościowych wzięto bezpośrednio z literatury i jest to 0,70%. Wadą tego typu danych może być tendencyjność ze względu na poszczególne grupy społeczne.

W przypadku taksówek założeniem jest rejestracja wszystkich podróży odbywanych tym środkiem transportu. Przyjęto znaleziony w literaturze podział modalny, wynoszący 3%. Należy jednak pamiętać, że operatorów może być wiele i dane pozyskane będą tylko od jednego z nich. Zaletą jest wzrost popularności zamawiania przewozów przez aplikacje mobilne, co pozwala na stworzenie bazy danych. Pozwala to również na stworzenie bazy danych opartych na GNSS. Wzrost popularności przewozów zamawianych przez aplikacje doprowadził do tego, że tradycyjne firmy taksówkowe również wprowadzają swoje dedykowane aplikacje, a kiedyś te przewozy zamawiano za pomocą połączenia telefonicznego dającego znacznie mniej danych do modelowania ruchu.

Systemy ANPR odczytują tablice rejestracyjne pojazdów w różnych punktach pomiarowych. Założono pełną wykrywalność zbliżoną do udziału samochodów w rejonie występowania kamer. Współczynnik będzie równy udziałowi modalnemu dla samochodów osobowych – 40%. Wadą tego źródła jest uwzględnianie tylko podróży indywidualnych, a zaletą automatyczna i wysoka skuteczność rejestracji tablic.

5. Metoda symulacji ośrodka miejskiego i macierze podróży z Big Data

5.1. Podstawy teoretyczne metody symulacji ośrodka miejskiego

Ze względu na brak dużej liczby źródeł danych z konkretnego ośrodka miejskiego, zdecydowano się na zasymulowanie realistycznego ośrodka miejskiego. Celem symulacji jest również uniknięcie konkretnego studium przypadku ośrodka miejskiego, czyli ze względu na afiliację autora – Aglomeracji Wrocławskiej na bazie Kompleksowych Badań Ruchu z 2018 roku. Tak przeprowadzone badania mają na celu stworzenie sposobu agregacji danych pochodzących z Big Data w sposób uniwersalny dla wykorzystania dla innych ośrodków miejskich. Umożliwiło to również zastosowanie dziewięciu źródeł big data, których użycie opisano w rozdziale 4.2. Niestety ma to też konsekwencje opisane w rozdziale 6.2.1, czyli rozbieżności w wynikach.

Zdecydowano się utworzyć ośrodek miejski uproszczony do 100 rejonów transportowych, co dało adekwatną macierz podróży 100x100. Założono 10 rejonów ścisłego centrum o największej generacji podróży - są to numery rejonów od 1 do 10. Następnie 40 rejonów obszaru centralnego aglomeracji - numery od 11 do 50 o średniej liczbie podróży. Rejony poza głównym miastem aglomeracji o najmniejszej liczbie podróży – 40 rejonów z numerami 51 do 90. Ostatnie 10 rejonów z numerami 91 do 100 to rejony reprezentujące drogi tranzytowe obejmujące podróże oddalonych rejonów aglomeracji oraz ruch pozaaglomeracyjny z wpływem na przepływy podróży w aglomeracji. W celu uzyskania zróżnicowanych liczb mieszkańców w rejonach transportowych zdecydowano się na sposób generowania liczb losowych podobny do stosowanego w metodzie Monte Carlo. Na potrzeby pracy dla rozróżnienia od klasycznej metody, nazwano to generowaniem pseudo Monte Carlo. Generatorem liczb pseudolosowych jest MS Excel. Losowanie wykonano za pomocą funkcji =LOS(), która zwraca liczbę pseudolosową w zakresie liczb rzeczywistych od 0 do 1. W celu uzyskania zakładanej liczby podróży funkcję przemnożono i dodano stałą. Wzory funkcji dla każdego typu rejonów zestawiono w tabeli 5.1. Wzór na liczbę grupy mieszkańców KX w rejonie i -tym (5.1) składa się z pierwszej liczby będącej stałą S plus liczbą zmienną Z mnożoną przez wygenerowaną liczbę pseudolosową L od 0 do 1:

$$KX_i = S + Z * L \quad (5.1)$$

Tabela 5.1 przedstawia jakie wartości S i Z założono dla konkretnego typu rejonu transportowego i dla konkretnej kategorii podróżnych. W tabeli K0 to dzieci do lat 7, K1 to osoby pracujące, K2 to osoby uczące się w szkołach ponadpodstawowych oraz studenci, K3 to osoby uczące się w szkołach podstawowych, a grupa K4 to emeryci.

Tab. 5.1 Zestawienie sposobu generacji liczby mieszkańców dla rejonów

Typ rejonu/Kategorie osób	K0	K1	K2	K3	K4
Centrum	600+600*L	2500+2500*L	1200+1200*L	1800+1800*L	1000+1000*L
Obszar miejski	250+250*L	1500+1500*L	500+500*L	750+750*L	400+400*L
Obszar podmiejski	100+100*L	700+700*L	200+200*L	300+300*L	150+150*L
Drogi zewnętrzne	600+600*L	2500+2500*L	1900+1900*L	300+300*L	150+150*L

Po wygenerowaniu liczby mieszkańców dla każdego rejonu transportowego założono współczynniki ruchliwości. W przypadku np. Kompleksowych Badań Ruchu bardziej wskazane byłoby określanie ruchliwości dla każdego rejonu z osobna, jednak ze względu na uproszczenie formuł w arkuszu kalkulacyjnym zdecydowano się uprościć ruchliwość przyjmując taką samą dla każdej z grup ze zróżnicowaniem na motywacje podróży. Współczynniki ruchliwości zestawiono w tab. 5.2.

Tab. 5.2 Zestawienie współczynników ruchliwości

Grupy\Motywacje	DP	DN	DI	PD	ND	ID	NZD
K0	0	0	0	0	0	0	0
K1	1	0	0,2	0,9	0	0,3	0,2
K2	0	0,9	0,2	0	0,7	0,3	0,3
K3	0	0	1,1	0	0	1,1	1
K4	0	0	1	0	0	1	0,9

Ze względu na symulacyjne warunki modelu zdecydowano określić atrakcyjność poszczególnych rejonów dla danych motywacji podróży w formie procentowej. Ze względu na nielosowanie liczb, a definiowanie określonych zakresów wartości maksymalne i minimalne zestawiono w tabelach 5.3 oraz 5.4. Liczby zakładano w sposób oddający charakter ośrodków miejskich. Przykładowo nie zakładano szkół w każdym rejonie transportowym, oraz ograniczono występowanie uczelni wyższych tylko do obszaru centralnego.

Tab. 5.3 Maksymalne wartości procentowe atrakcyjności rejonów

Typ rejonu\Max % Atrakcji	Praca	Nauka podst.	Nauka śr.	Nauka wyż.	Inne
Centrum	3,1%	4,5%	6,5%	25,4%	4,5%
Obszar miejski	1,7%	3,9%	6,1%	Brak	1,4%
Obszar podmiejski	1,1%	1,2%	Brak	Brak	0,7%
Drogi zewnętrzne	3,2%	Brak	Brak	Brak	2,0%

Tab. 5.4 Minimalne wartości procentowe atrakcyjności rejonów

Typ rejonu\Min % Atrakcji	Praca	Nauka pods.	Nauka śr.	Nauka wyż.	Inne
Centrum	1,5%	2,5%	3,2%	10,8%	2,5%
Obszar miejski	1,3%	0,2%	2,2%	Brak	0,5%
Obszar podmiejski	0,1%	0,2%	Brak	Brak	0,2%
Drogi zewnętrzne	1,4%	Brak	Brak	Brak	0,9%

Używając powyższych danych przystąpiono do wygenerowania macierzy podróży. Ruchotwórczość rejonów Y wyliczono za pomocą wzoru 5.3, gdzie R oznacza współczynnik ruchliwości z indeksem dolnym pokazującym skrótem motywację podróży, np. ID jako Inne-Dom, $\%A$ oznacza atrakcyjność konkretnego typu, np. $\%AP_j$ to $\%$ atrakcyjności rejonu j -tego dla motywacji praca, $\%ANPod$ to procent atrakcyjności dla motywacji nauka w szkołach podstawowych, analogicznie dopisek Śr i Wyż oznacza szkoły średnie i wyższe. $\%AI_j$ oznacza atrakcyjność rejonu j -tego dla podróży innych, czyli niezwiązanych z pracą i domem.

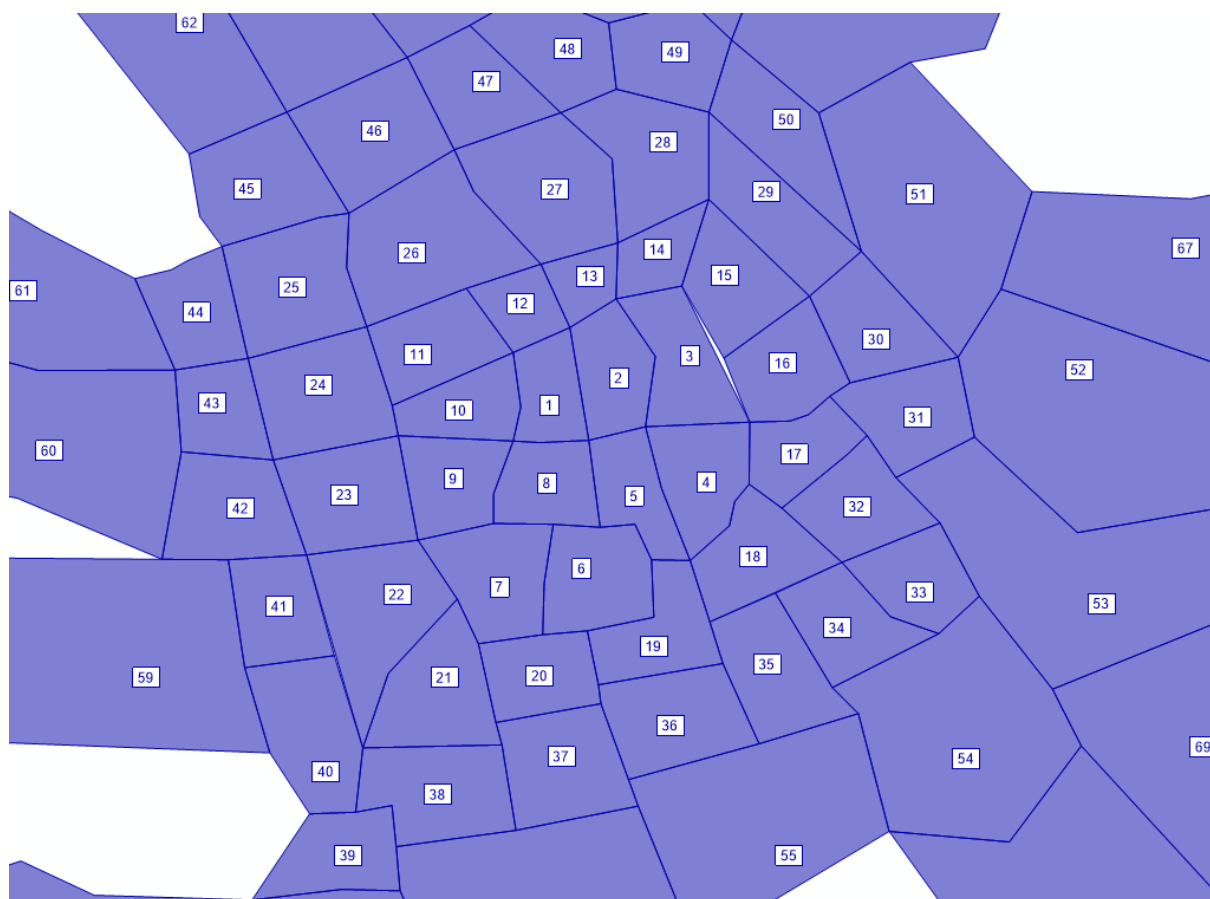
$$\begin{aligned}
 Y_{ij} = & R_{DP1} * K1_i * \%AP_j + R_{DN2} * K2_i * \left(\frac{\%A\acute{S}r_j}{\sum AN\acute{S}r_j} + \frac{\%AWy\acute{z}_j}{\sum ANWy\acute{z}_j} \right) + \\
 & + R_{DN3} * K3_i * \%ANPod_j + R_{DI1} * K1_i * \%AI_j * 0,8 + R_{DI2} * K2_i * \%AI_j * 0,8 + \\
 & + R_{DI3} * K3_i * \%AI_j * 0,8 + R_{DI4} * K4_i * \%AI_j * 0,8 + R_{NZD1} * K1_i * \%AI_j * 0,8 + \\
 & + R_{NZD2} * K2_i * \%AI_j * 0,8 + R_{NZD3} * K3_i * \%AI_j * 0,8 + R_{NZD4} * K4_i * \%AI_j * 0,8
 \end{aligned} \quad (5.3)$$

W ten sposób obliczono gotową macierz podróży dla symulowanego ośrodka miejskiego. Ze względu na rozmiary macierzy – 100x100, macierze dołączono wyłącznie w formie elektronicznej w załącznikach 3-12. W celu zaprezentowania potencjalnych przepływów podróży w rozdziale 5.3. stworzono poglądowy model aglomeracji.

5.2. Model poglądowy symulowanego ośrodka miejskiego

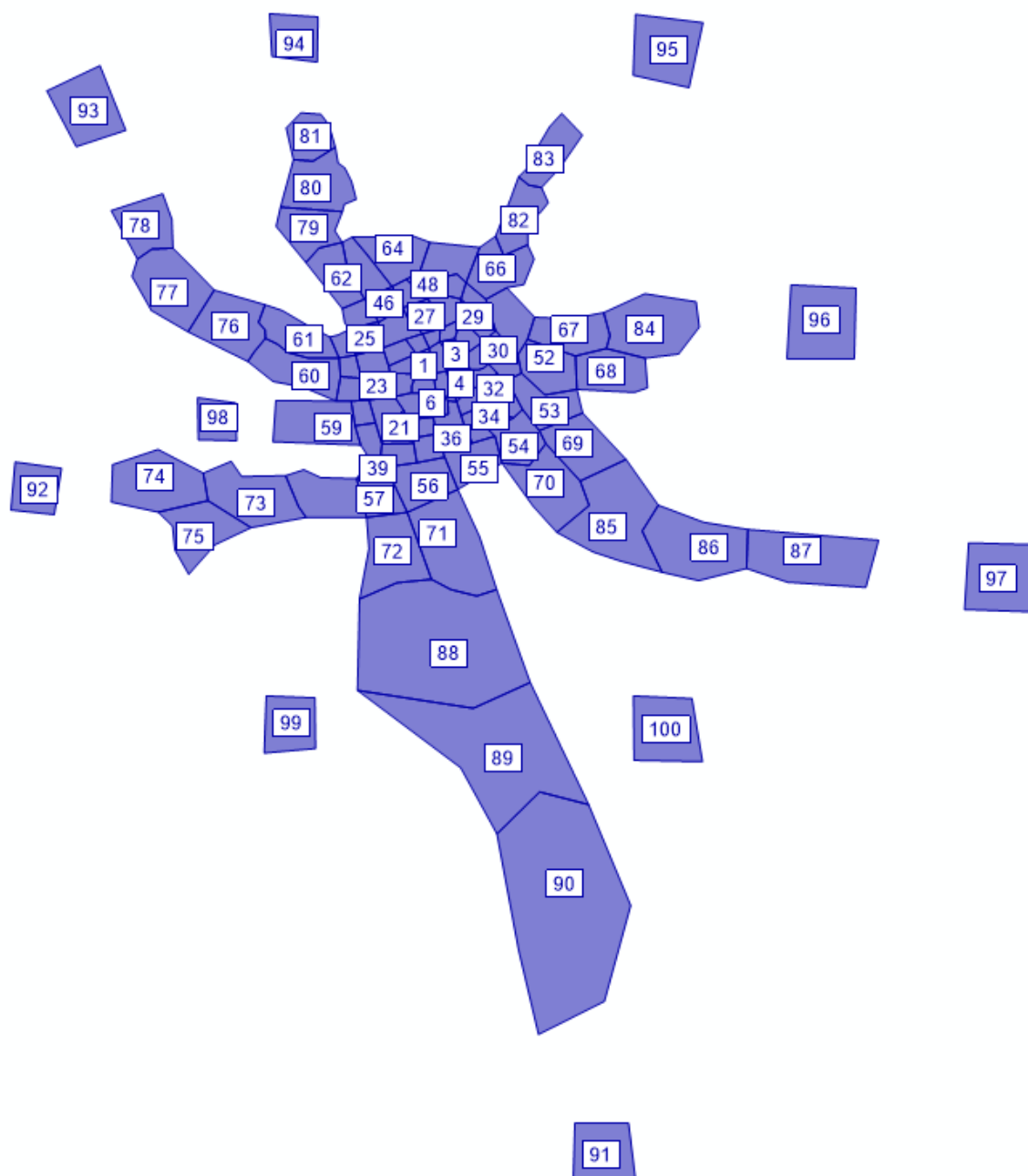
Symulacja ośrodka miejskiego ma na celu pokazanie przepływów podróży w analizowanych wariantach zestawienia źródeł danych, jak i wariacie sprawdzającym całą metodę, opisanych w rozdziale 6.2. Ze względu na to, że metoda dotyczy tylko macierzy podróży, a nie rozkładu na sieć transportową, sieć jest znacząco uproszczona. Przykładowo nie uwzględniano podziału na środek lokomocji czy nie zmieniano większości podstawowych parametrów programu PTV Visum.

Tworzenie modelu rozpoczęto od rozmieszczenia siatki rejonów transportowych. Jest ich 100 na około półmilionowy, zasymulowany ośrodek miejski. W celu stworzenia dokładnego modelu transportowego należałoby stworzyć większą liczbę rejonów, jednak wykracza to poza zakres potrzebny do niniejszej rozprawy, więc zdecydowano się nie dzielić zasymulowanych rejonów na mniejsze. Układ rozpoczęto od obszaru ścisłego centrum – rejonów od 1 do 10. Założeniem jest obszar gęstej zabudowy o wysokiej gęstości zaludnienia. Obszar ten okrążają rejony od 11 do 50 - będącej obszarem miejskiej zabudowy o mniejszej liczbie mieszkańców. Obydwa typy rejonów pokazano na rzucie ekranu z programu PTV Visum na rys. 5.1.



Rys. 5.1 Układ rejonów transportowych w centrum aglomeracji

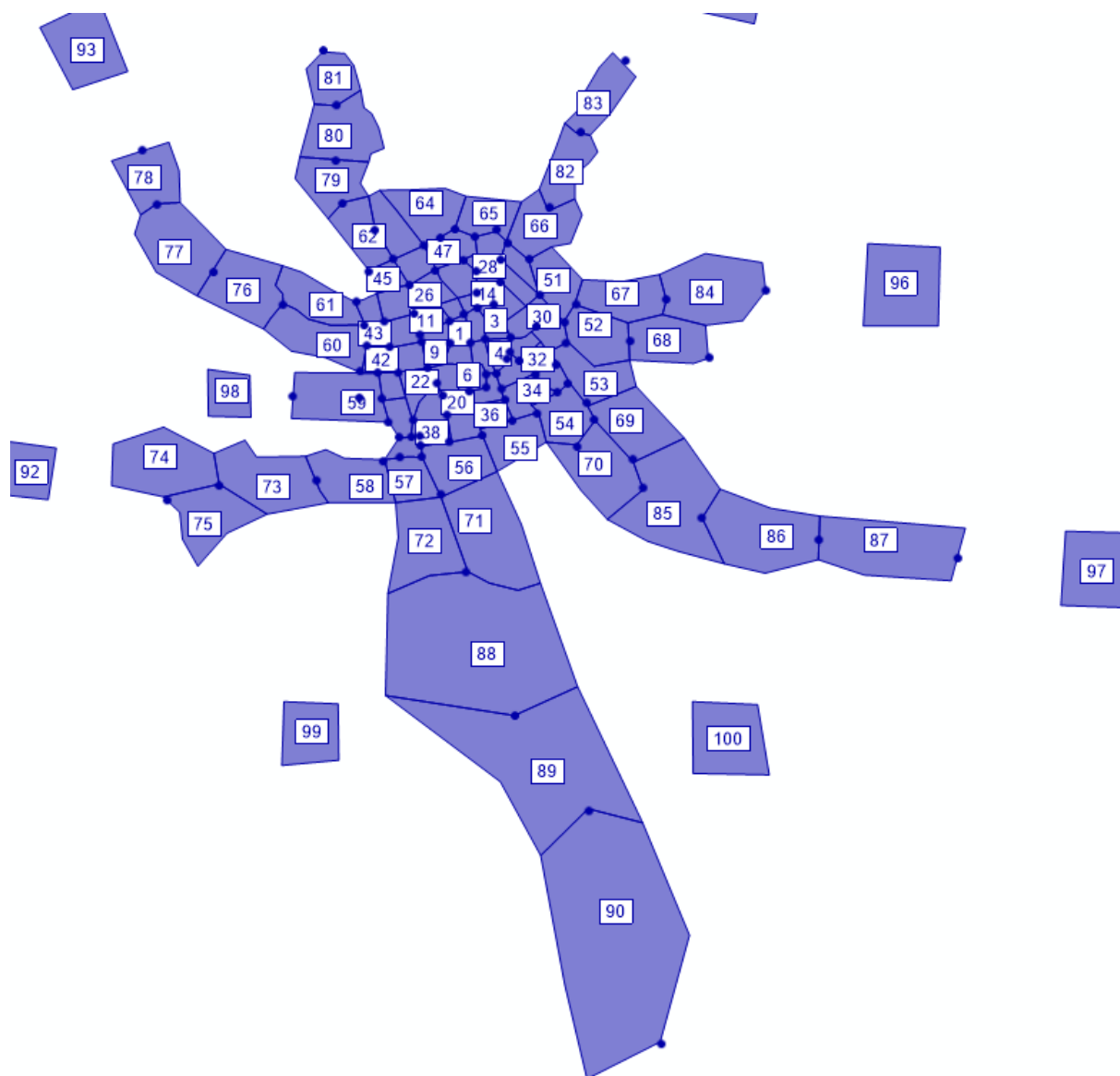
Stwierdzono, że układ sieci miejskiej ma nie być idealną siatką kwadratów, pomimo tego, że wymiar macierzy podróży 100x100 na to pozwalał. Zdecydowano również, że nie będzie to miasto o układzie typowo amerykańskim z równoleżnikową siatką połączeń. Aglomeracja ma kształt typowy dla miast europejskich i rozchodzi się ona odnogami. Przybrała kształt „ośmiornicy” z „mackami” w postaci odnóg; rejony od 51 do 90 prezentują kształt takiego układu. Część to tzw. wsie przyległe do aglomeracji, ciasno oplatające siatkę rejonów 1-50. Reszta reprezentuje odnogi aglomeracji będące miejscowościami skupionymi wokół dróg dojazdowych. Rejony 91-100 prezentują wpływ odległych miejscowości w stosunku do aglomeracji. Zdecydowano się na to uproszczenie ze względu na czytelność niniejszej rozprawy. Bardzo duży model uniemożliwiłby pokazanie większości wartości liczbowych na połączeniach. Rejony te są symbolicznie pokazane jako czworokątne rejony odległe od reszty. Układ pozostałych rejonów i tym samym całego układu rejonów transportowych w symulowanej aglomeracji pokazano na rys. 5.2.



Rys. 5.2 Układ rejonów w całej aglomeracji

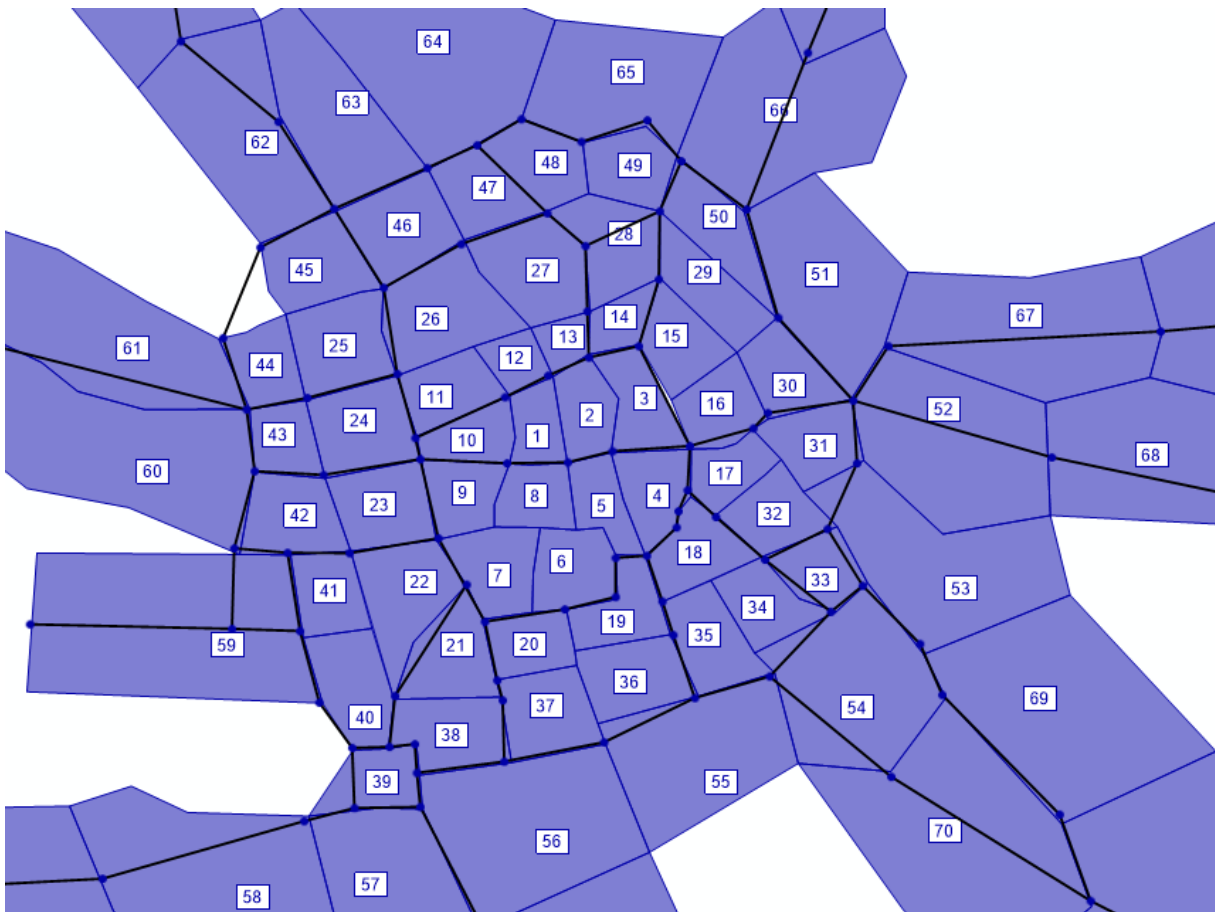
Mając gotowy układ rejonów transportowych zaplanowano sieć głównych połączeń w aglomeracji. Ze względu na nieporuszanie w rozprawie innych elementów modelu transportowego niż macierze podróży, zrezygnowano z wykonywania tego w pełnej dokładności i ograniczono się do stworzenia siatki głównych arterii miejskich bez dróg osiedlowych. W rozprawie nie ma również podziału na środki transportu, więc stworzono ogólną sieć opartą w oprogramowaniu o sieć drogową o wysokiej przepustowości, by

zwizualizować tylko przepływy podróżnych między rejonami na złagodzonych warunkach związanych z przepustowością. Na rysunku 5.3. pokazano układ węzłów w całej aglomeracji.



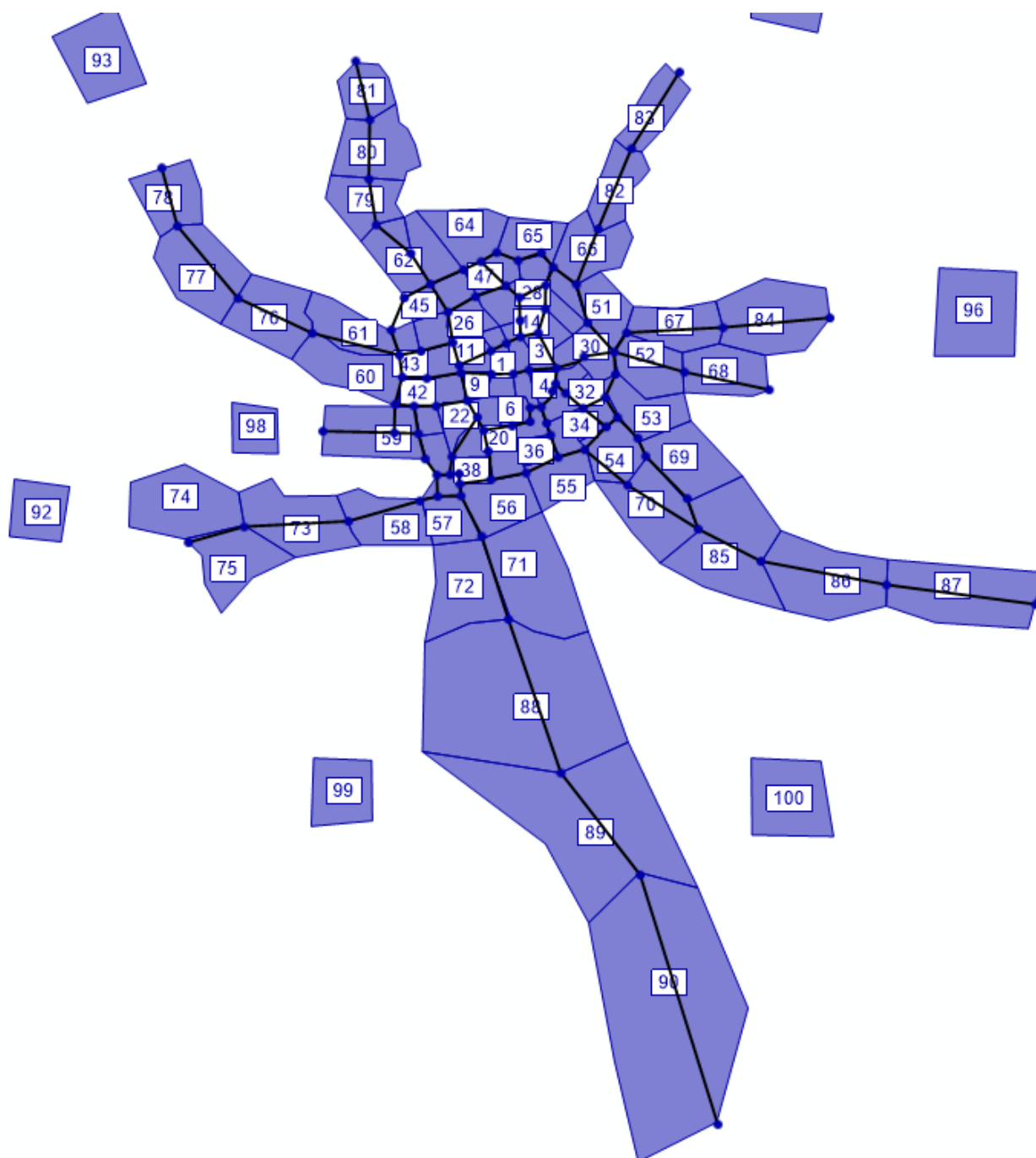
Rys. 5.3 Układ węzłów w całej aglomeracji

Na bazie węzłów z poprzedniego rysunku stworzono sieć drogową. Na rysunku 5.4 pokazano układ połączeń w centrum aglomeracji, a na rysunku 5.5. pokazano widok oddalony całego modelu aglomeracji. Rysunek 5.6. pokazuje sieć bez włączonej warstwy rejonów transportowych.



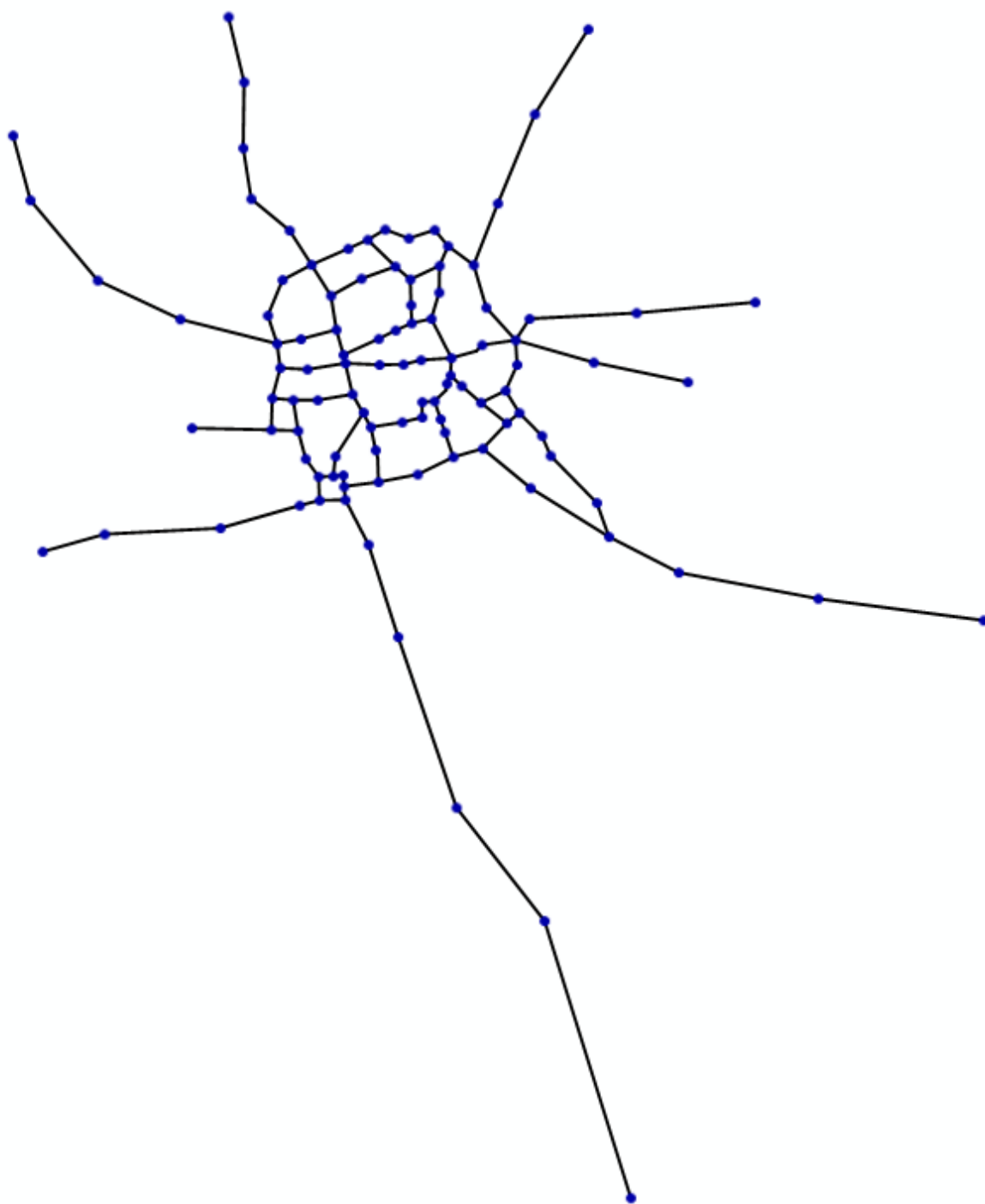
Rys. 5.4 Układ połączeń w obszarze centralnym aglomeracji

W następnych rozdziałach używano określeń w celu ułatwienia czytelnikowi zrozumienia, o które połączenia chodzi. Wyszczególniono tylko niektóre połączenia, które okazały się mieć charakterystyczne zmiany w liczbie podróżnych na tych połączeniach dla poszczególnych wariantów. Nawiązują one do klasycznych nazw połączeń kojarzonych z aglomeracjami. Znalazła się zatem tam np. obwodnica śródmiejska. Jest to „ring” wokół obszaru centralnego aglomeracji, czyli rejonów od 1 do 50. Oprócz tego połączenie równoleżnikowe przechodzące przez obszar ścisłego centrum (rejonów od 1 do 10) na odcinku pomiędzy obwodnicą śródmiejską nazwano trasą W-Z. Reszta połączeń nie wymagała szczegółowej nazwy w celach opisowych.

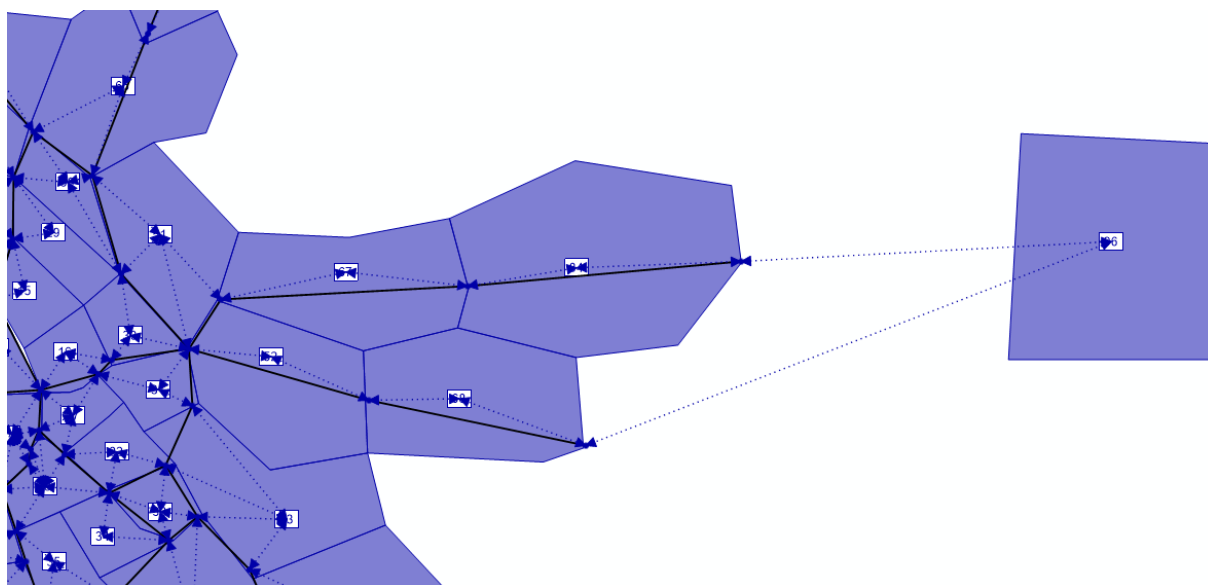


Rys. 5.5 Układ połączeń w całej aglomeracji

W celu pokazania przepływów bez ograniczeń infrastrukturalnych każde z połączeń ma te same parametry. Są to pozostawione bazowe parametry zasugerowane przez PTV Visum. Przepustowość – CapPrT to 99999, a V0PrT to 50 km/h. Nie dokonywano również podziału modalnego na środki transportu, więc na model nie wpływa brak parametrów dla transportu publicznego.

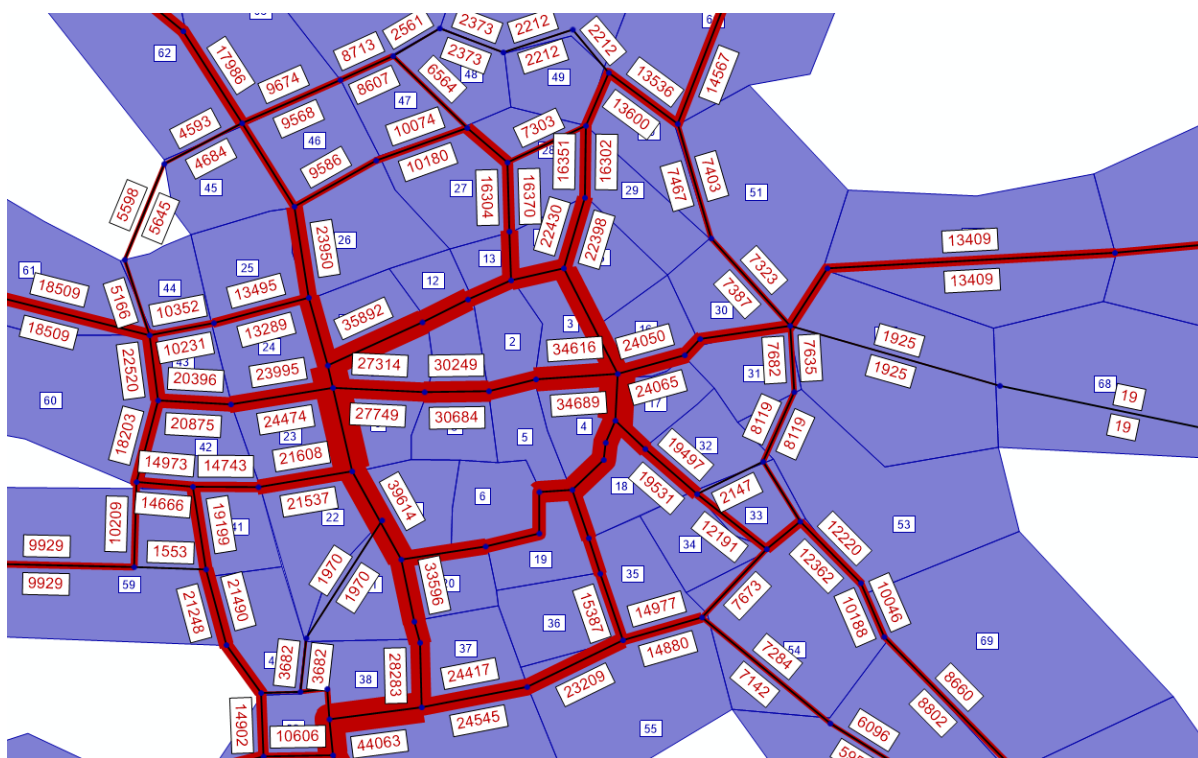


Rys. 5.6 Układ połączeń bez siatki rejonów transportowych

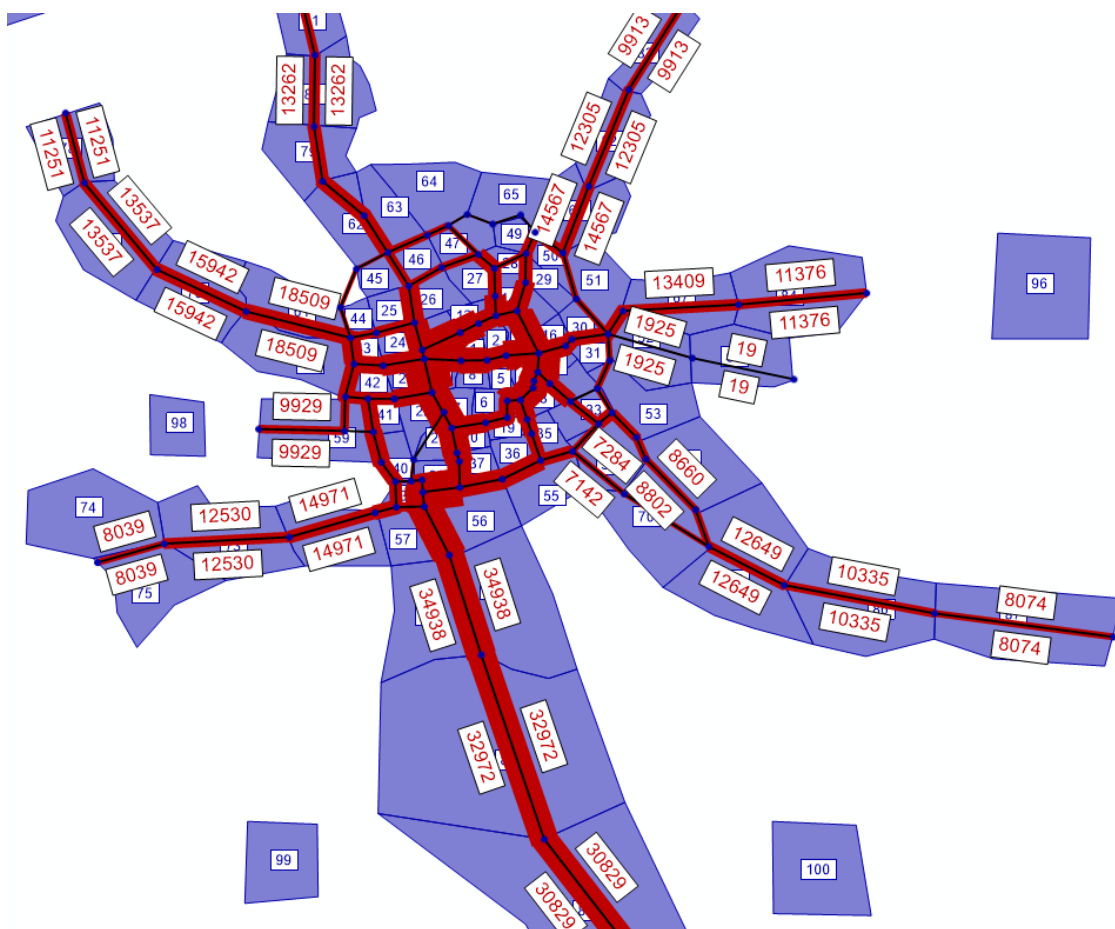


Rys. 5.7 Wycinek modelu pokazujący sposób przyłączania rejonów do sieci

Na rysunku 5.7. pokazano sposób przyłączania rejonów do sieci transportowej. Nie wprowadzano dróg w głąb rejonów ze względu na czytelność modelu. Nie było to konieczne na poziomie dokładności samych przepływów na bazie macierzy podróży. W tym przypadku alternatywną opcją zaprezentowania rezultatów mogłyby być więźby ruchu, jednak dla celów opisowych zdecydowano się pokazać rezultat na uproszczonym modelu. Po wprowadzeniu zasymulowanej macierzy podróży przeliczono model. Rezultat w postaci natężeń ruchu na połączeniach zestawiono na rysunkach 5.8 i 5.9. Na rysunku 5.8 pokazano obszar centralny aglomeracji. Na rysunku 5.9 widok całego obszaru aglomeracji. W rozdziale 6 te wyniki są w opisie modelem bazowym. W przypadku każdego wariantu zestawienia źródeł Big Data za pomocą metody opisanej w rozdziale 4 różnice pokazane na scenariuszach modelu w rozdziale 6 są różnicą w wartościach od poniższych rysunków. Ze względu na brak obwodnicy zewnętrznej aglomeracji, przez centrum przebiegają duże potoki podróżnych, które są odzwierciedleniem podróży pomiędzy drogami tranzytowymi. Podobnie, gromadzący również dużą liczbę ludności obszar ścisłego centrum (rejonów od 1 do 10) „ściąga” duże potoki ruchu. Wyjątkowa odnoga aglomeracji to odnoga południowa, w której zbiegają się trzy rejony odpowiadające za drogi zewnętrzne, co powoduje znacząco wyższą liczbę podróży. Na rysunku 5.10 pokazano samą macierz podróży dla lepszej czytelności w postaci mapy ciepła. Macierz podróży z wartościami liczbowymi znajduje się w Załączniku 3 do rozprawy.



Rys. 5.8 Obszar centralny zasymulowanej aglomeracji



Rys. 5.9 Widok na obszar całej aglomeracji



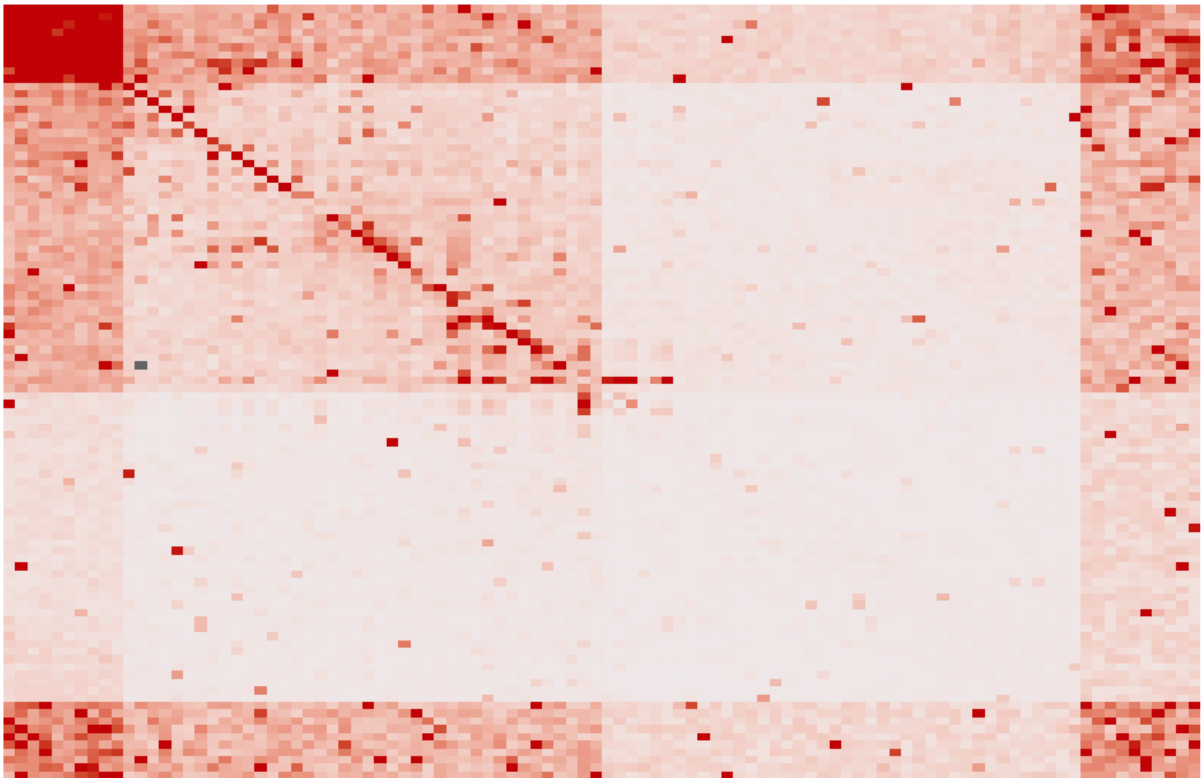
Rys. 5.10 Mapa ciepła macierzy podróży modelu bazowego

5.3. Macierze podróży ze źródeł Big Data

W celu urealistycznienia wyników i problemów związanych z ich pozyskaniem, w maksymalnym stopniu wykorzystano dane dostępne w literaturze. W pierwszej kolejności użyto danych z wrocławskiego KBR 2018 [167], gdzie były dostępne godzinne macierze podróży z sieci GSM i dane GPS z sondowania pojazdów. Pierwszym wyzwaniem było dopasowanie macierzy podróży. W rozprawie model bazowy to macierz 100x100, a macierz wrocławska ma 475x475 rejonów. W celu wykorzystania macierzy z Wrocławia zredukowano je do rozmiarów modelu bazowego.

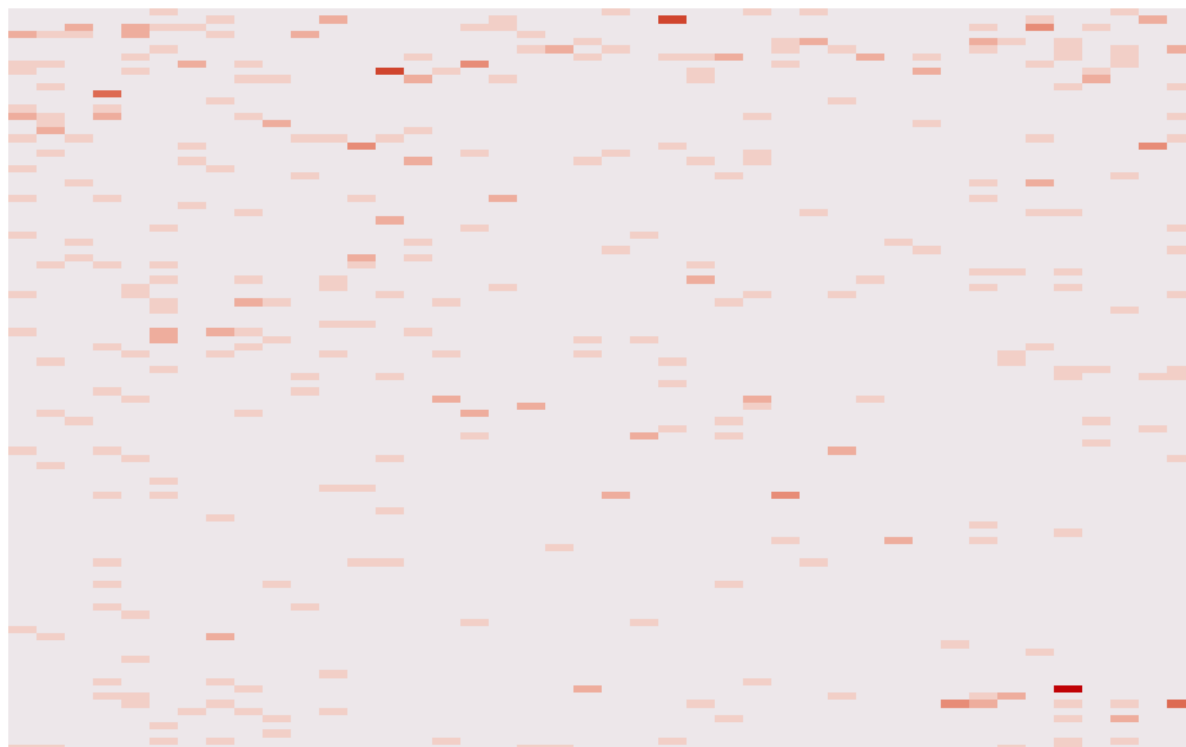
Macierz modelu bazowego dzieli się na zakresy: 10% rejonów - obszar ścisłego centrum aglomeracji, 40% - obszar miejski i 40% - obszar podmiejski oraz 10% rejonów prezentujących dalsze rejony aglomeracji wraz z drogami tranzytowymi. Zdecydowano się posegregować rejony od największego do najmniejszego i wydzielono trzy obszary: 20% najbardziej obciążonych podróżami rejonów, następne 40% i 40% najmniej obciążonych. Z zakresu 20% najbardziej obciążonych rejonów, zaczynając od tego najbliższego medianie, wzięto 10 najbliższych rejonów ponad medianę i 10 rejonów poniżej. 10 rejonów powyżej mediany

założono na obszar centralny aglomeracji. Rejony poniżej mediany to będą drogi tranzytowe, które mają stosunkowo zbliżone wartości podróży w modelu bazowym do tych z rejonu ścisłej zabudowy centrum aglomeracji. Podobnie postąpiono przy następnych 40% rejonów z wrocławskiej macierzy, jednak w tym przypadku wzięto 40 rejonów najbliższych medianie po obu stronach na obszar miejski. Analogicznie postąpiono tworząc obszar podmiejski z pozostałych 40% rejonów. Wybieranie odległych od siebie zakresów pozwoliło na uzyskanie przeskoków wartości pomiędzy numerami rejonów tak jak w modelu bazowym, co jest widoczne na mapie ciepła na rys. 5.11.



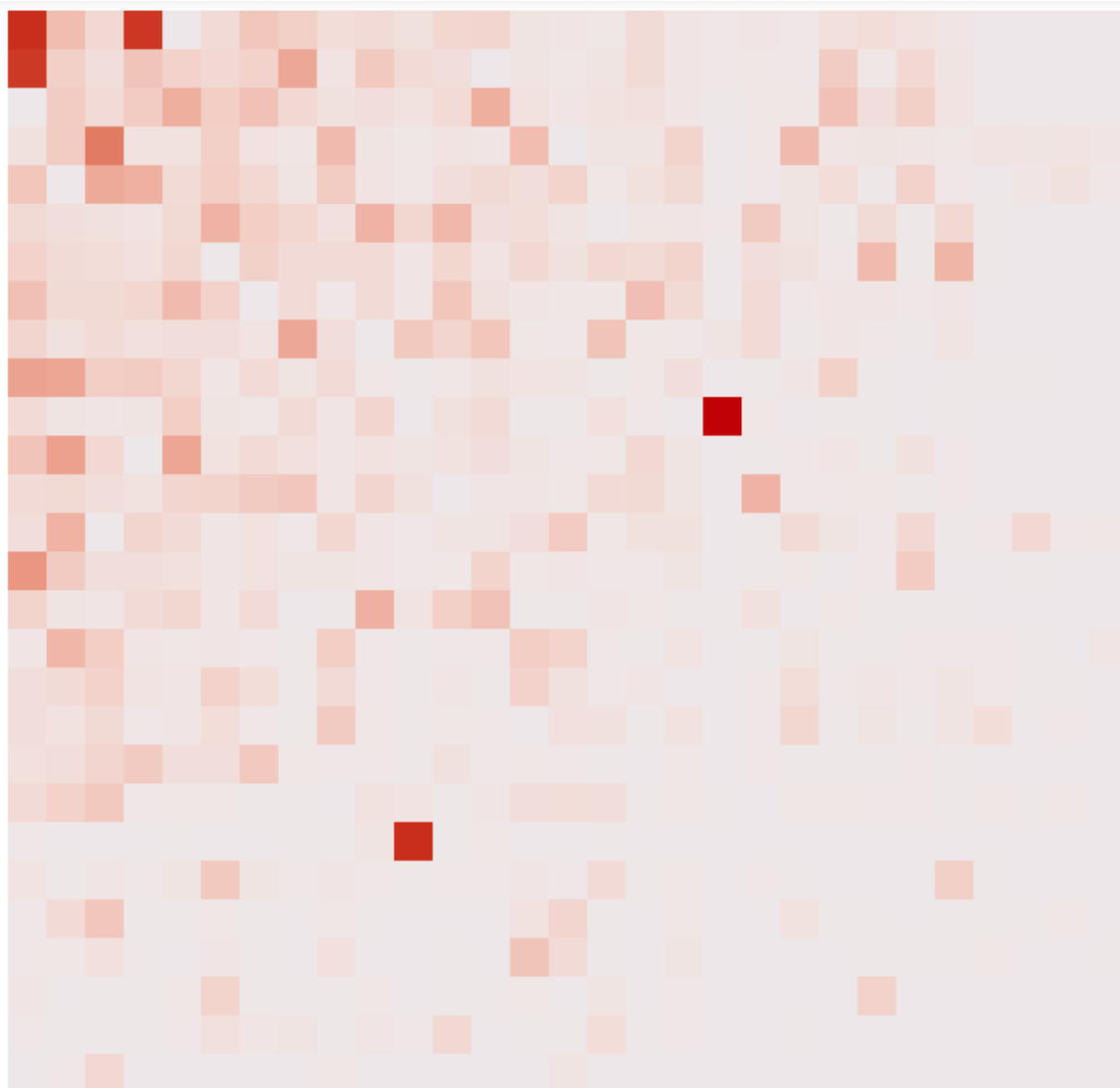
Rys. 5.11 Macierz podróży z sieci telefonii komórkowej – mapa ciepła

Analogicznie postąpiono z danymi z GPS, które również pochodzą z wrocławskiego KBR 2018. W tym przypadku jednak 9438 na 10000 relacji międzyrejonowych prezentowało zerowe wartości podróży. Mimo tego, zdecydowano się uwzględnić to źródło danych, jako źródło wybrakowane. Na rys. 5.12 pokazano macierz podróży, na której widać brakujące wartości. Warto również wspomnieć, że są to wartości niskie - maksymalnie w jednej relacji zarejestrowano zaledwie 6 (!) podróży.



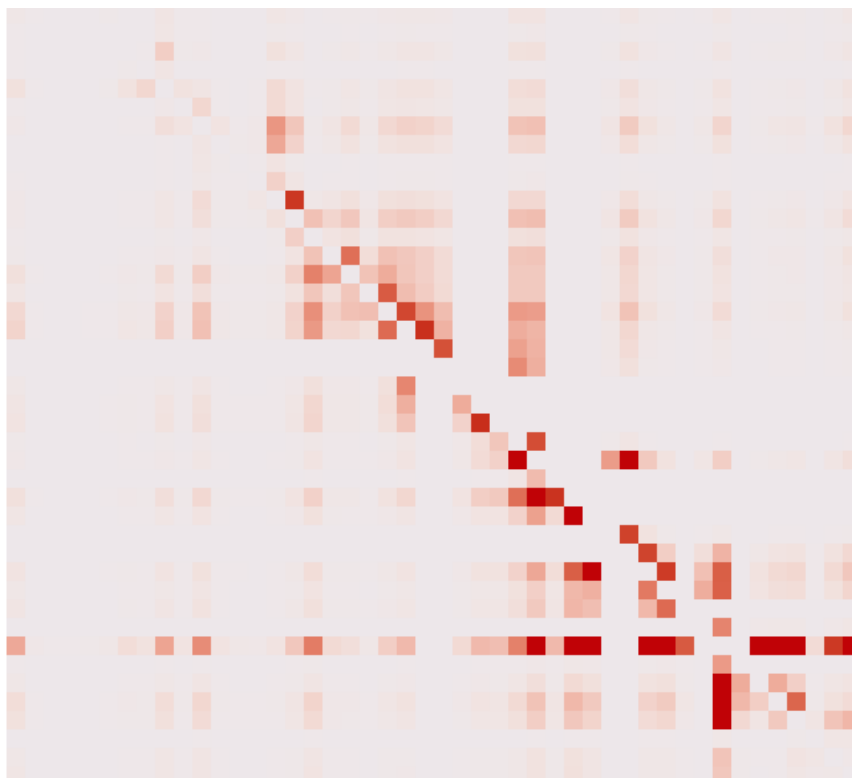
Rys. 5.12 Macierz podróży z GPS – mapa ciepła

Następne źródła pozyskano z literatury. W artykule [76] na rys. 6. pokazano macierz podróży z dokładnymi wartościami z każdego skanera Bluetooth. Zaznaczono tam skanery dające niewiarygodne wartości, więc w niniejszej rozprawie ich nie uwzględniono. W celu przepisania tabeli z artykułu użyto narzędzi korzystających ze sztucznej inteligencji dostępnych w internecie oraz ręcznie korygowano błędnie zinterpretowane przez narzędzie cyfry. Badanie zsumowało w macierzy wykrycia podróży ze 150 dni, więc zredukowano macierz do uśrednionych wartości dobowych. Następnie dokonano analogicznej segregacji rejonów jak w przypadku poprzednich źródeł. Tym razem macierz 29x29 nie zapewniała danych w pełnym wymiarze 100x100, więc nie wybierano zakresów, a wykorzystano przesegregowaną macierz w całości. Ze względu na uwzględnianie w badaniu obszaru centrum miasta, stwierdzono, że będzie to pierwsze 29 rejonów w macierzy, a pozostałe dadzą dane zerowe. Rezultat w postaci mapy ciepła pokazano na rys. 5.13. Nie pokazano całej macierzy 100x100, ponieważ w większości byłyby pusta. Pokazano część macierzy bez rejonów niezawierających dane, więc na rysunku jest macierz będąca tylko wycinkiem całej sieci transportowej. Pokrywa obszar tylko 29x29 rejonów pokazany na mapie ciepła. W tym przypadku jest to macierz powstała z danych z pojazdów, a nie telefonów, co uwzględniono odpowiednim współczynnikiem w metodzie.



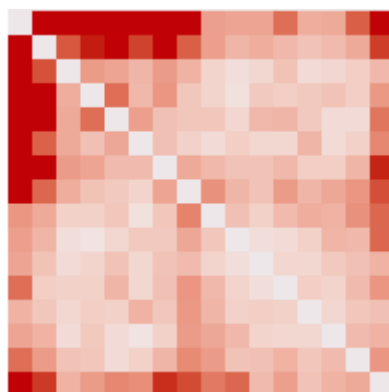
Rys. 5.13 Macierz podróży z Bluetooth – mapa ciepła 29x29 rejonów z danymi

Kolejną macierzą, którą udało się pozyskać z literatury jest macierz z automatycznego zliczania pasażerów. Dzięki uprzejmości autora pracy [114] otrzymano pliki robocze w programie Excel. Wykorzystano macierze z dni pracujących, ale dla każdej z czterech tras pokrywały się maksymalnie cztery przystanki. Z tego powodu wykorzystano dane tylko z jednej linii - nr 1. Innym problemem był charakter danych, ponieważ dane zawierały jedynie pojedynczą godzinę pomiarów. Przeskalowano je zakładając 17 godzin kursowania autobusów. Ze względu na rozmiar macierzy 46x46 użyto jej dla rejonów obszaru miejskiego (bez obszaru ścisłej zabudowy centrum) oraz 6 rejonów podmiejskich. Podobnie jak w poprzednim przypadku chcąc zwizualizować rezultat na rys. 5.14 przedstawiono tylko obszar macierzy, który zawierał dane.



Rys. 5.14 Macierz podróży z APC – mapa ciepła

W literaturze udało się również znaleźć macierz podróży z systemów inteligentnych kart miejskich. Macierze załączone były do raportu [43] W związku z tym, że dane opierały się na pojedynczych liniach autobusowych, zastosowano analogiczne podejście do poprzedniego źródła danych. Wykorzystano fakt, że linie znacząco różnią się liczbą wykrywanych podróży. „Trasę D” użyto jako macierz dla rejonów obszaru miejskiego, tak samo jak w poprzednim przypadku, jednak dla rejonów odnog aglomeracji użyto „Trasy B”. Rezultat w postaci macierzy podróży dla rejonów obszaru miejskiego pokazano na rys. 5.15, gdzie macierz ma wymiar 16x16 i na rys. 5.16, gdzie macierz ma wymiar 35x35 dla samego obszaru odnog aglomeracji.

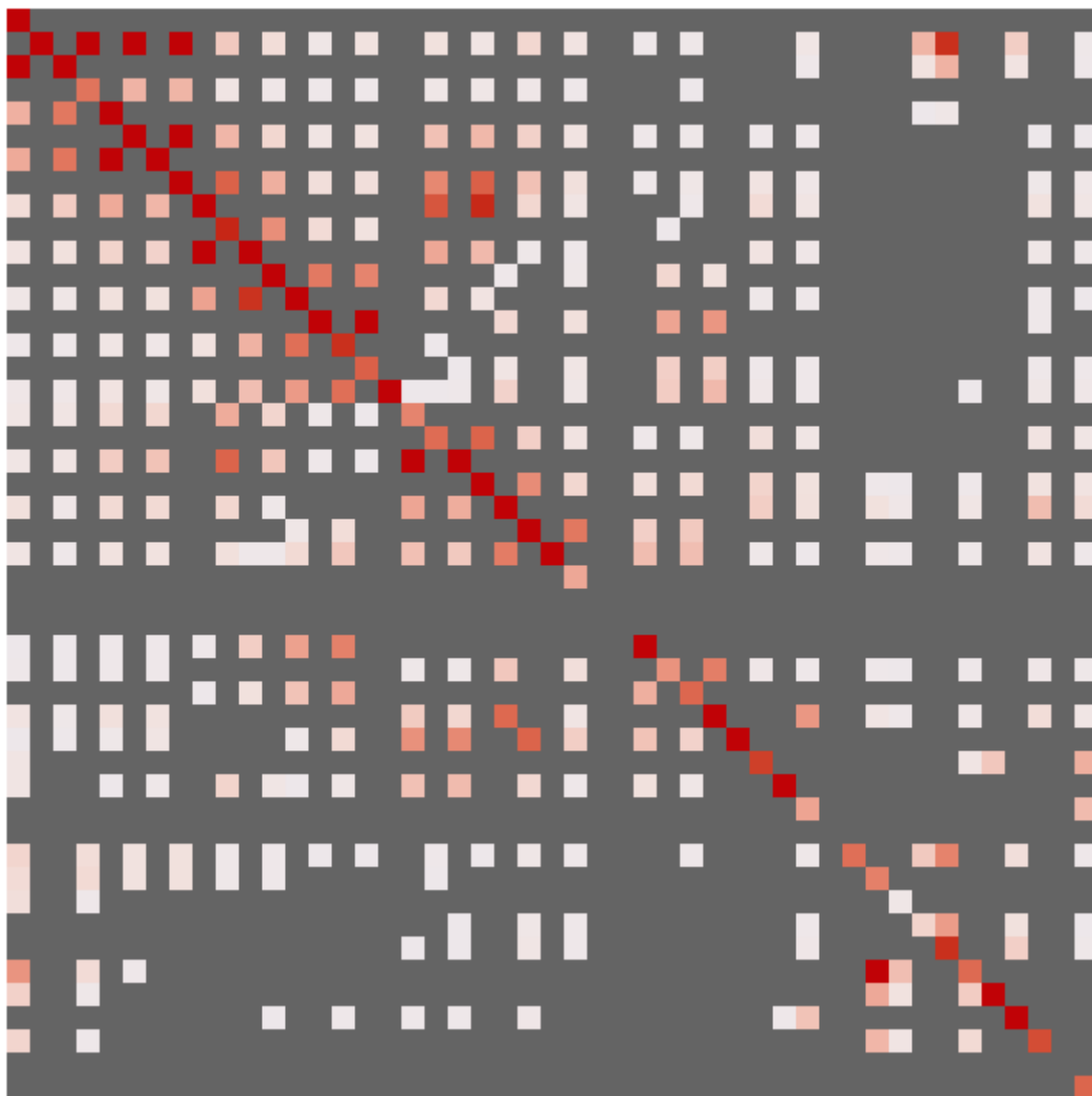


Rys. 5.15 Macierz podróży z kart miejskich obszar miejski – mapa ciepła



Rys. 5.16 Macierz podróży z kart miejskich obszar podmiejski – mapa ciepła

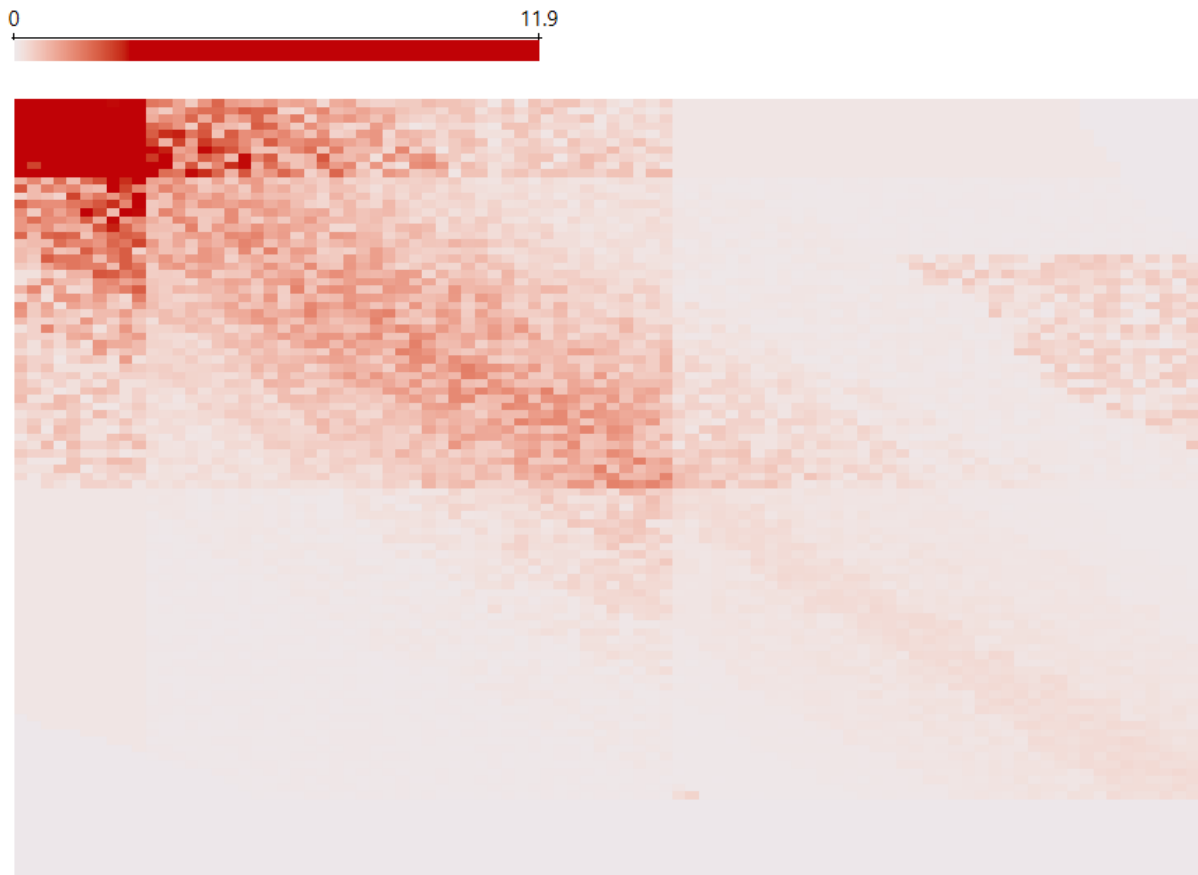
Ostatnią z macierzy, którą udało się pozyskać z literatury w postaci liczbowej, są dane z ANPR – automatycznego rozpoznawania numerów tablic rejestracyjnych. Dane pochodzą z pracy [124]. W tym przypadku ze względu na dużą liczbę podróży, w rejonach w lewym górnym rogu macierzy, macierz 47x47 w niezmienionej formie wykorzystano w rejonach od 1 do 47. Rezultat przedstawiono na rys. 5.17, gdzie szare miejsca nie zawierają żadnych podróży.



Rys. 5.17 Macierz podróży z ANPR – mapa ciepła

Pozostałych trzech źródeł Big Data nie udało się uzyskać w postaci liczbowej. Zdecydowano się na odtworzenie ich na podstawie opisów i map ciepła dostępnych w literaturze. W pierwszej kolejności stworzono macierz z bazy danych z UTO i rowerów miejskich. Ze względu na układ ośrodka miejskiego, założeniem była przewaga podróży na krótkim dystansie, więc większe liczby podróży są skupione bliżej przekątnej macierzy. Sposób generowania liczb jest analogiczny jak w przypadku symulowania ośrodka miejskiego za pomocą wzoru 5.1. W macierzy nie ma rejonów odpowiadających za drogi zewnętrzne. Ze względu na około 30 różniących się zakresem liczbowym wzorów, podano przykład największego i najmniejszego. Relacje z największą liczbą podróży generowano wzorem przy stałej równej 4 i liczbie losowanej z zakresu od 0 do 8. Najmniejsze liczby to 0,05 i zmienna od 0 do 0,02, w celu

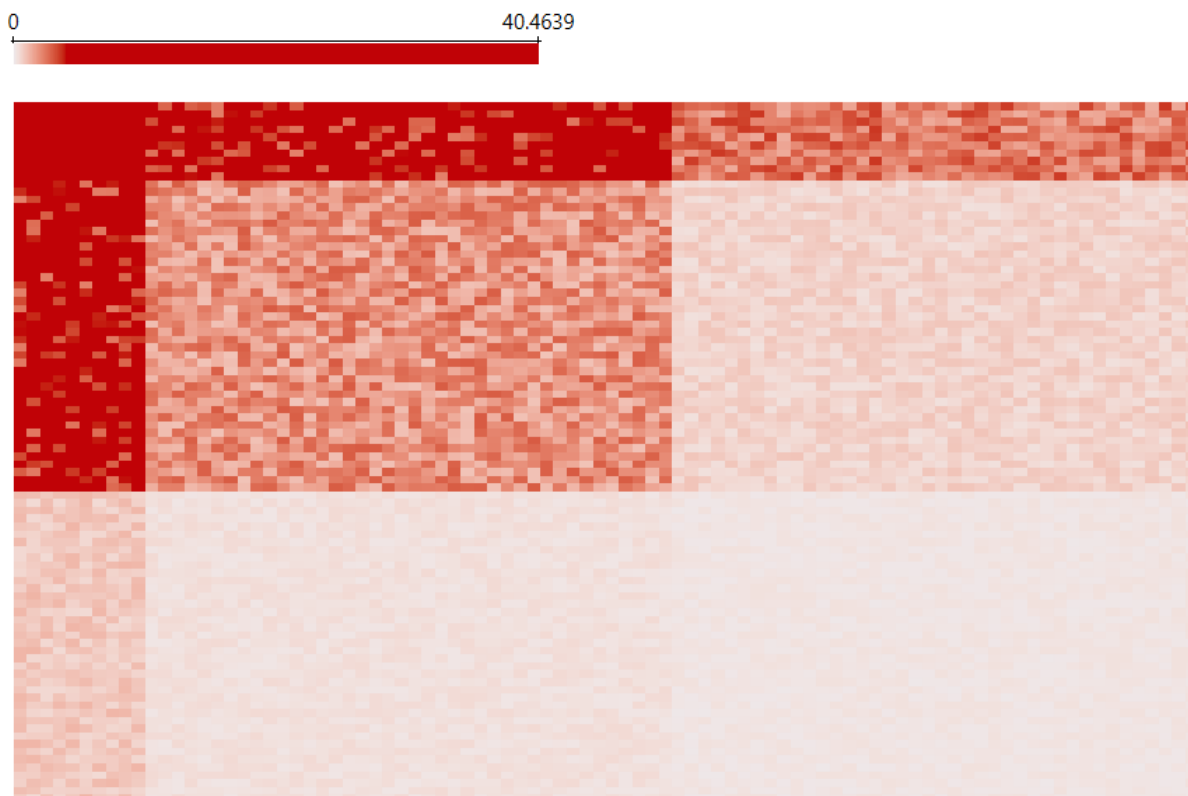
pokazania zaniżonych wyników na dalekich dystansach w aglomeracji. Efekt przedstawiono na rysunku 5.18.



Rys. 5.18 Macierz podróży z UTO i rowerów miejskich – mapa ciepła

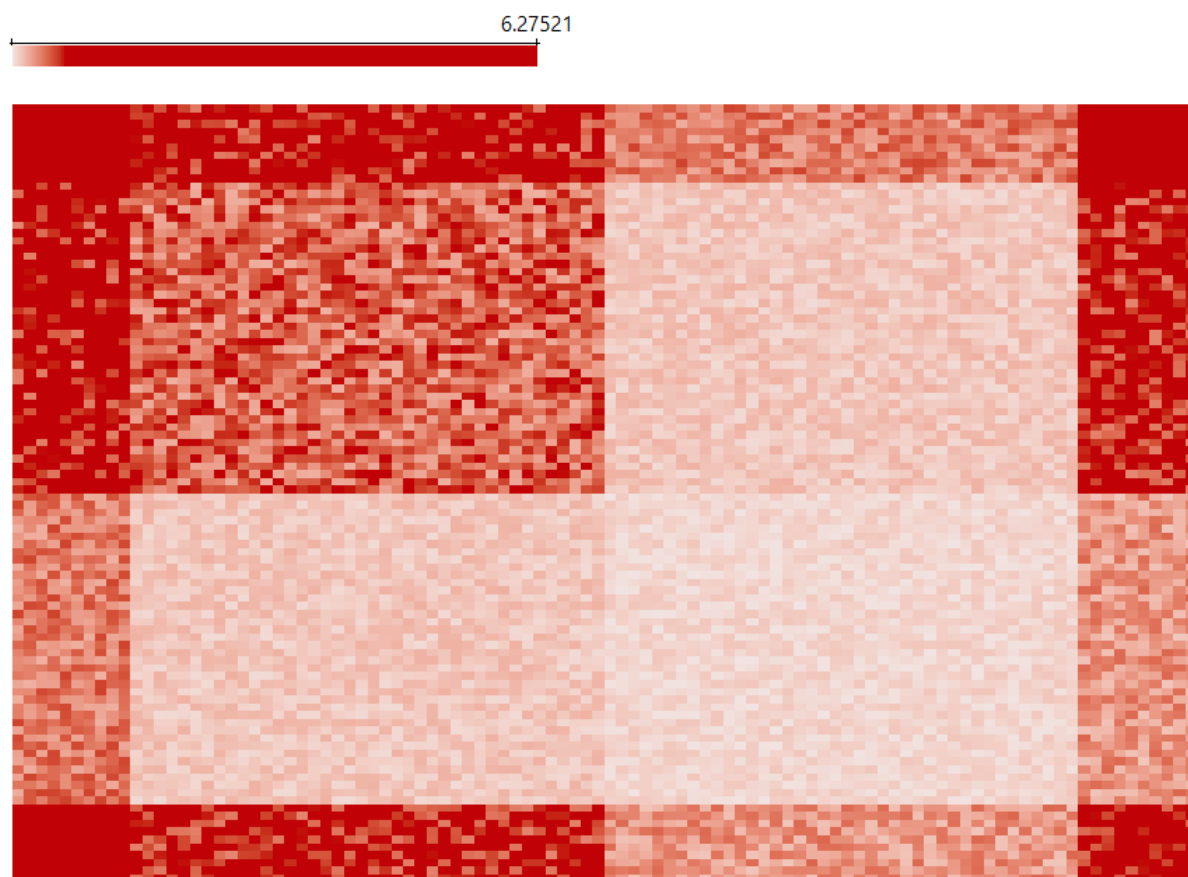
Na bazie map ciepła z literatury udało się odtworzyć także macierz z Taxi. Wykorzystano głównie mapę ciepła Nowego Yorku z artykułu [121], gdzie widać większość podróży w centrum miasta oraz asymetryczność. Zdecydowanie więcej podróży miało swój cel niż źródło w obszarze odpowiadającym w rozprawie obszarowi miejskiemu w rejonach 11-50. W macierzy widocznej na rys. 5.19 postarano się odtworzyć powyższe zależności. Motywem przewodnim była największa możliwa liczba podróży w obrębie obszaru ścisłego centrum oraz asymetryczność całej macierzy. W macierzy nie uwzględniono rejonów odpowiadających za drogi zewnętrzne. Liczby podobnie jak w przypadku danych z UTO i rowerów miejskich generowano wzorem 5.1, gdzie najbardziej obciążone relacje miały stałą 13,5, a liczbę losową w zakresie od 0 do 27. Najmniejsze liczby we wzorze to stała 0,075 i losowa od 0 do 0,225. Podobnie jak w przypadku UTO i rowerów miejskich część losowa była większa niż stała, by

maksymalnie zróżnicować liczby, pomimo świadomości, że tylko niektóre rejony w modelu bazowym mają znacząco wyższą liczbę podróży od rejonów tego samego typu.



Rys. 5.19 Macierz podróży z Taxi – mapa ciepła

Ostatnią macierzą, trzecią symulowaną i dziewiątą ogólnie, jest macierz z danych lokalizacyjnych z sieci mediów społecznościowych. W tym przypadku opisane w literaturze „anomalie” dotyczyły danych socjodemograficznych, które nie są poruszane w rozprawie. Tym samym stworzono macierz uśrednioną. Jest to również macierz sprawdzająca, czy symulowane macierze nie zawyżą wyników metody, jako te potencjalnie najdokładniejsze, tworzone z myślą o odwzorowaniu ośrodka miejskiego. Tezę tę sprawdzono w ostatnim z wariantów walidowanych w następnym rozdziale niniejszej rozprawy. Relacje pomiędzy rejonami obszaru ścisłego centrum aglomeracji losowano według wzoru 5.1. podając stałą 2,1 i zmienną od 0 do 4,2. Dla relacji o najmniejszej liczbie podróży, pomiędzy odnogami aglomeracji – stałą 0,14 i zmienną od 0 do 0,315.



Rys. 5.20 Macierz podróży z mediów społecznościowych – mapa ciepła

6. Walidacja metody

6.1. Prezentacja wyników

Po zastosowaniu metody w arkuszu kalkulacyjnym programu Excel dokonano wizualizacji macierzy podróży tworząc scenariusze porównawcze w oprogramowaniu PTV Visum. Ze względu na duże wymiary macierzy podróży - 100x100 - będzie ona nieczytelna jako rysunek w niniejszej rozprawie. Scenariusze porównawcze w programie Visum pozwalają na zobrazowanie dokładności uzupełniania danych w skali całego symulowanego ośrodka miejskiego. Przedstawiono też mapy ciepła macierzy. Oprócz tego w niniejszym rozdziale pokazano liczbową analizę porównawczą różnicy w całkowitej liczbie podróży pomiędzy wariantem „rzeczywistym”, a wariantami symulowanymi. W celu sprawdzenia tezy badawczej o korelacji pomiędzy większą liczbą źródeł Big Data, a dokładnością wyników stworzono kilka scenariuszy łączących różne źródła. Zamiast samego porównania symulowanego ośrodka miejskiego, jako modelu bazowego z danymi uzupełnionymi każdym z wykorzystanych źródeł.

Wybrano dziewięć wariantów zestawienia źródeł Big Data, które zdaniem autora dają przypadki z grup dających najbardziej zróżnicowane wyniki oraz z grup danych występujących w badaniach wspólnie. Należy podkreślić, że nie chodzi tu o dziewięć wykorzystanych źródeł danych. Zdecydowano się nie pokazywać przypadków spodziewanie najbardziej podobnych do siebie. Tym samym pierwszy wariant to macierz stworzona ze wszystkich źródeł danych z niniejszej rozprawy. Drugi to trzy bazy danych pozyskane z symulowania wyników na bazie charakterystyki źródła danych i dostępnych macierzy podróży w postaci „map ciepła” w literaturze. Trzeci z wariantów to dane pozyskane z wrocławskiego KBR. Czwarty to lepsze z dwóch źródeł dostępnych we wrocławskim KBR, czyli danych z sieci telefonii komórkowej. Piąty wariant to dane ze skanerów Bluetooth, systemów zliczania pasażerów, smart card data oraz ANPR. Następne trzy przypadki to pojedyncze źródła danych; kolejno: dane z mediów społecznościowych, taksówek i jako wariant ósmy dane z rowerów miejskich i UTO. Dziewiątym wariantem jest połączenie danych z sieci telefonii komórkowej, mediów społecznościowych i taksówek.

Różnice od danych symulowanych na każdym z poniżej prezentowanych wariantów pokazano na czterech zrzutach ekranu. Wartości są prezentowane niezależnie od sugerowania czy są „dobre” lub „złe”. Kolor czerwony symbolizuje więcej podróży niż w przykładzie modelowym, kolor zielony mniejszą liczbę podróży. Należy pamiętać, że jest to interpretacja macierzy

podróży na modelu o podstawowych parametrach oprogramowania PTV Visum, a przedmiotem rozprawy nie jest podział modalny lub podział na sieć. Jest to prezentacja wizualna, a nie rzeczywista. Pierwszy z zestawu czterech zrzutów ekranu, dla każdego z przypadków, pokazuje różnice liczbowe w obszarze centralnym. Drugi w oddaleniu pokazuje wartości liczbowe na obszarze aglomeracji i dróg dojazdowych do miasta. Następne zrzuty ekranu są pokazaniem analogicznych obszarów, jednak w wartościach procentowych różnic. Jest to szczególnie istotnie, ponieważ część połączeń ma znacząco mniejsze wartości liczbowe podróży, więc zmiany mające kilka tysięcy podróży dla najbardziej obciążonych rejonów mogą mieć stosunkowo mało istotną różnicę. Analogicznie na sieci istnieją połączenia mające zaledwie kilkadziesiąt podróży i tam zmiana może mieć znacząco większą różnicę procentową sięgającą kilkudziesięciu procent, która ma marginalny wpływ na całą sieć transportową ze względu na niewielką liczbę podróży.

6.2. Wykorzystanie metody i analiza rezultatów

W celu jasnego przedstawienia sposobu działania metody i stworzenia wariantu 1, w niniejszym podrozdziale na początku zawarto przykład liczbowy. Przykład pokazuje w jaki sposób obliczono wartość dla jednej wybranej komórki. Wybrano konkretną relację w macierzy podróży. Są to podróże z rejonu 1 do rejonu 7. Metodę zaimplementowano w MS Excel. W arkuszu kalkulacyjnym jest to komórka H2. W modelu bazowym symulowanego ośrodka miejskiego, opisanego w rozdziale 5, z rejonu 1 do rejonu 7 dobowo odbywa się 478 podróży. Kolejno w macierzach ze źródeł Big Data z danych GSM wykryto dobowo średnio 9 podróży. Z danych pochodzących z Bluetooth wykryto 79 podróży na relacji. Z systemów automatycznego zliczania pasażerów nie ma danych. Rejony 1 i 7 nie posiadały przystanków autobusowych, z których pozyskano dane. Obszar ścisłego centrum w założeniu jest obsługiwany transportem szynowym, a dane dotyczyły linii autobusowych. Podobnie nie ma danych z systemów „smart card data”, które w tym przypadku były kartami miejskimi. W przypadku mocno wybrakowanej macierzy z danych z GPS wykryto jedną podróż. Macierz podróży z ANPR w zakresie tych rejonów posiada dane, jednak akurat na tej relacji nie wykryto podróży. W danych geolokalizacyjnych z mediów społecznościowych wykrywano na tej parze rejonów 6,018 podróży. Z Taxi jest w tej komórce średnio 33,869 podróży dobowo. Ostatnie źródło to dane z rowerów miejskich oraz UTO, gdzie korzystając z tej floty odbywa się średnio 5,1 podróży dobowo na stacjach z podróżą początkową w rejonie 1 i końcową w rejonie 7.

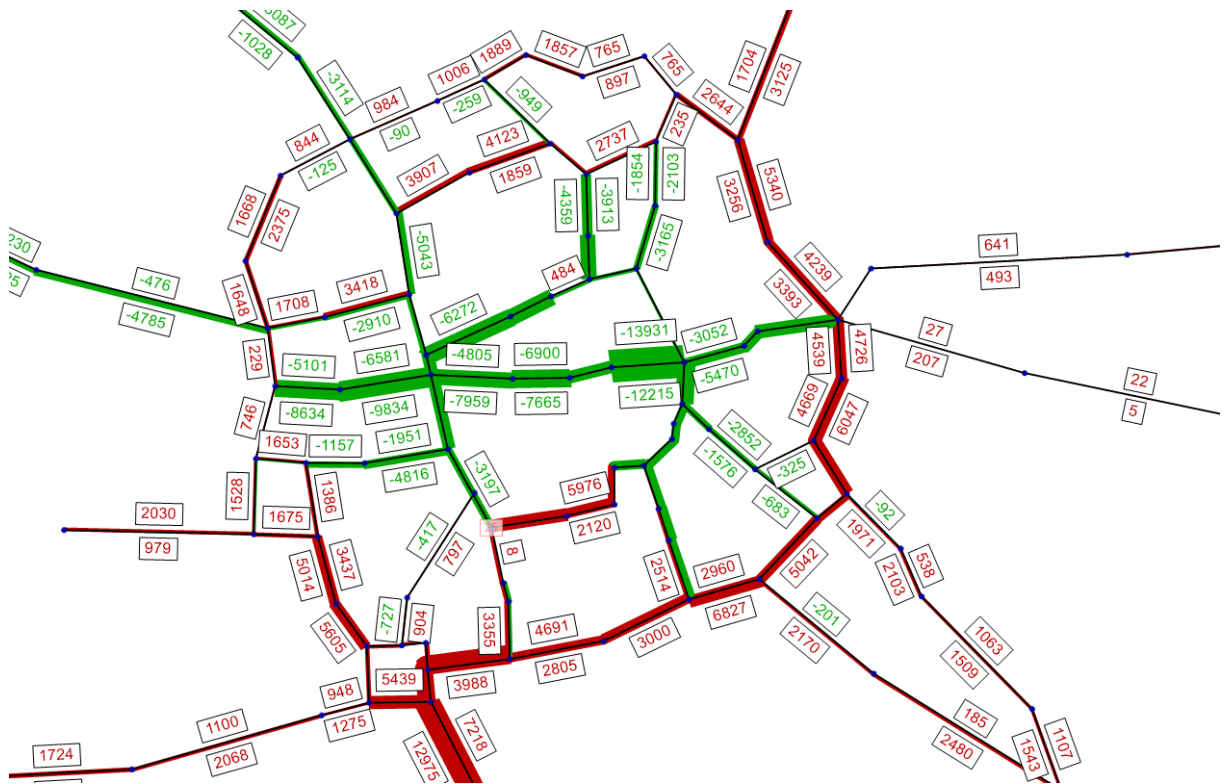
Algorytm opisany w podrozdziale 4.1. pominął w obliczeniach trzy źródła, które nie wykazały danych na tej relacji, resztę pomnożył przez współczynniki przeliczeniowe opisane w podrozdziale 4.2. Kolejno będzie to 9 podróży z GSM pomnożone przez odwrotność 16%, co daje 56,25. Z Bluetooth to 79 pomnożone przez odwrotność 24,8%, co daje 318,55. Z GPS to jedna podróż pomnożona przez $1/0,37$, co daje 2,7. Następnie ze źródeł symulowanych, opisanych w rozdziale 5, wzięto 6,018 podróży i podzielono przez odwrotność współczynnika przeliczeniowego wynoszącego 0,7%, co daje 859,71. Podróże z Taxi to 33,869 pomnożone przez odwrotność 3% dające wynik 1128,97. Na końcu dane z UTO i floty rowerów miejskich wynoszące 5,1 pomnożone przez odwrotność 1,237% i dające 412,29.

Sumarycznie średnia z wyników dała wartość zbliżoną do modelu bazowego: 471 podróży w stosunku do 478 z modelu bazowego. Ze względu na dobry rezultat końcowy opisany szczegółowo w podrozdziale 6.2.1. nie zdecydowano się na wprowadzenie dodatkowych wag. Wymagałoby to pogłębionych analiz rezultatów cząstkowych i stwierdzono, że wykracza to poza zakres rozprawy i jej symulacyjny charakter. Ze względu na mocno zaniżone wartości z GPS, na pewno byłoby to źródło z najniższą z sugerowanych wag. Badania takie wymagałyby analizy wielu studiów przypadków i wytypowania, czy w każdym z przypadków zastosowania danego źródła dane pochodzące z niego są niewiarygodne. W niniejszej rozprawie nadając wagi na podstawie dostępnych danych dokonano by tylko analizy na potrzeby studium tego przypadku, więc wyniki byłyby nieadekwatne w zastosowaniu do innych przypadków. Przeprowadzenie tego typu analizy jest również niemożliwe ze względu na małą liczbę pozyskanych przez autora macierzy podróży pochodzących z tego samego źródła. Mimo tego wyraźnego braku, wyniki opisane w niniejszym rozdziale są obiecujące. Ma na to wpływ fakt, że w przypadku danych z GPS, które były najmniej wiarygodne, tylko 627 komórek macierzy podróży z 10000 zawierało dane, więc wpłynęło to na niewielki odsetek komórek w macierzy podróży (6,27%) .

6.2.1. Uzupełnienie wszystkimi dostępnymi źródłami

Pierwszy wariant sprawdza całą metodę. To wykorzystanie dostępnych źródeł, nawet tych wybrakowanych, by uzyskać najbardziej zbliżone do rzeczywistości wyniki. Na rysunku 6.1. pokazano sieć obszaru centralnego miasta. Rysunek 6.2. pokazuje obszar oddalony, obszar centralny jest bez opisu wartości, jednak widać na nim proporcje grubości pasów

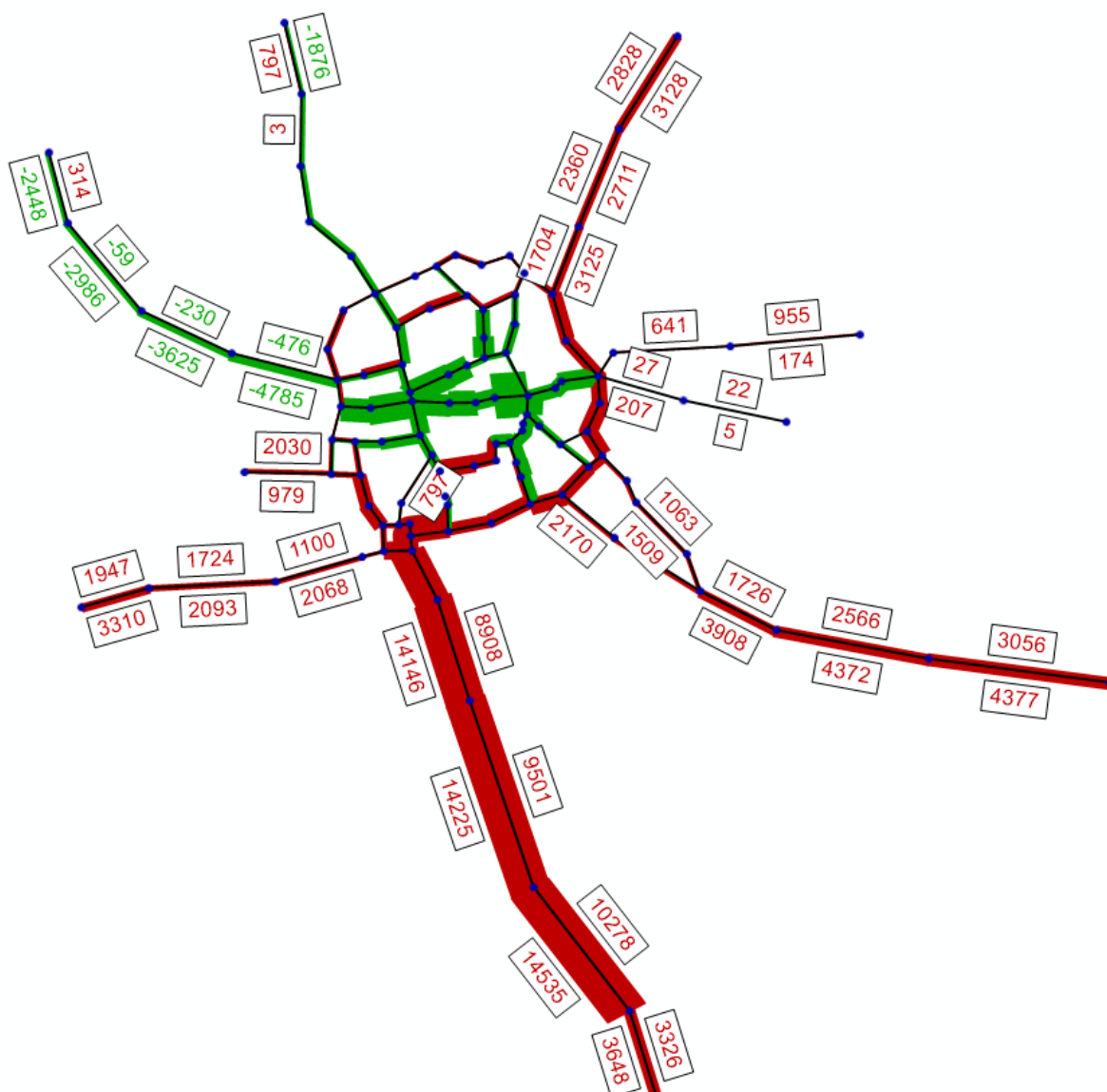
reprezentujących różnice. Analogiczną kolejność mają rysunki w każdym z następnych przykładów poniżej.



Rys. 6.1 Wariant 1, widok na obszar centrum aglomeracji

Na rysunku 6.1. widać dużą zgodność przepływów podróży z modelem bazowym. Rozbieżności mogły wynikać z nieporuszania w metodzie rozkładu przestrzennego podróży, który został wykonany automatycznie przez program. Tym samym parametry podstawowe programu PTV Visum mogły wpłynąć na zwiększenie różnic widocznych na modelu. Pomimo tego wynik zdaniem autora jest satysfakcjonujący, biorąc pod uwagę wykorzystanie danych rzeczywistych z różnych miejsc, dostosowanych samą skalą do ośrodka symulowanego, a nie danych rzeczywistych do ośrodka istniejącego. Wyniki wariantu są mocno zbliżone z wyjątkiem kilku połączeń. Szczególnie przeszacowana jest „odnoga” południowa aglomeracji widoczna w całości na rysunku 6.2. Jest to spowodowane założeniem wychodzących z niej i łączących się tam trzech dróg tranzytowych, co spotęgowało niedokładności, szczególnie, że spora część (sześć na dziewięć) źródeł nie zawierała tego obszaru. Dlatego liczba użytych źródeł w tych połączeniach jest znacząco niższa od obszaru centralnego aglomeracji. Należy zwrócić uwagę, że ze względu na brak obwodnicy pozamiejskiej ośrodka, te wartości wpływają na niedokładność reszty połączeń w obszarze centralnym. Ten sam obszar pokazano na

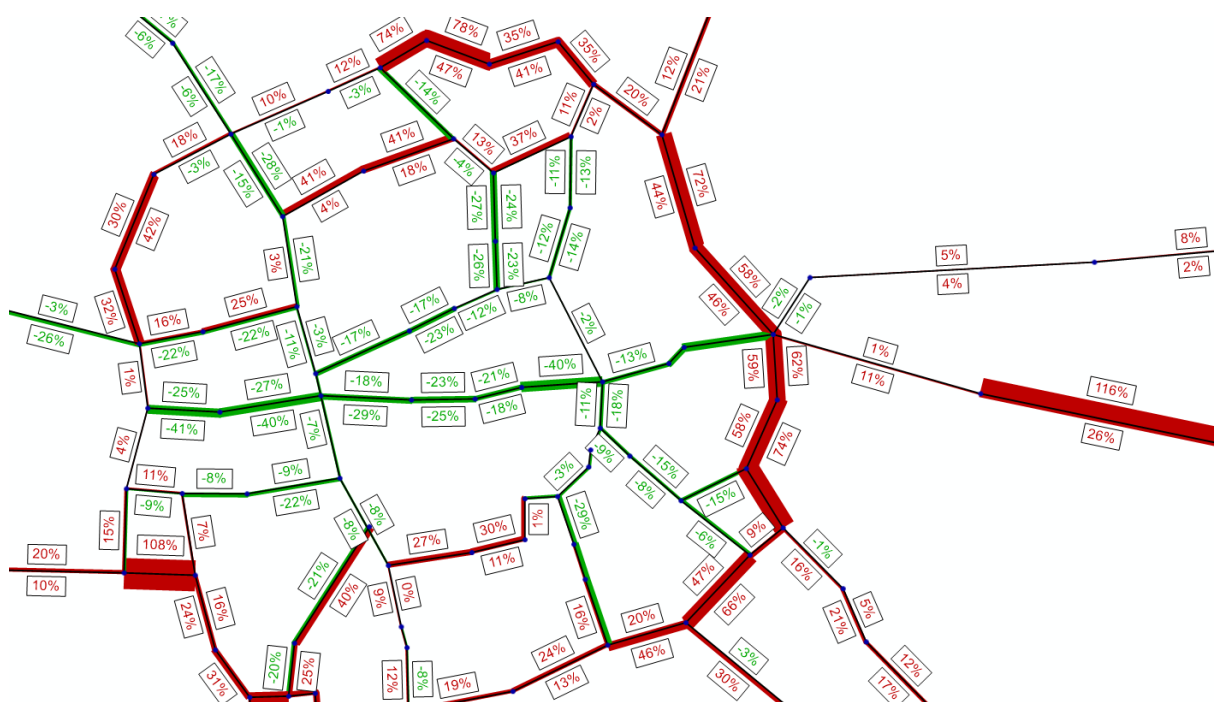
wartościach różnic procentowych na rysunku 6.3. Ogólnie nadmiary i niedobory podróży równoważą się na sieci. Mimo tego niezadowalające jest znacząco większe natężenie ruchu na obwodnicy śródmiejskiej i nierealistycznie odciążony odcinek „trasy W-Z”. Różnice też mogą wynikać z podejścia do macierzy bazowej, która jest symetryczna, a dane pochodzące z big data nie zostały symetryzowane. W związku z tym, że dane z macierzy pochodzących z big data, pochodziły z różnych okresów i zostały ekstrapolowane do wartości dobowych, nie są one w pełni symetryczne.



Rys. 6.2 Wariant 1, widok na obszar całej aglomeracji

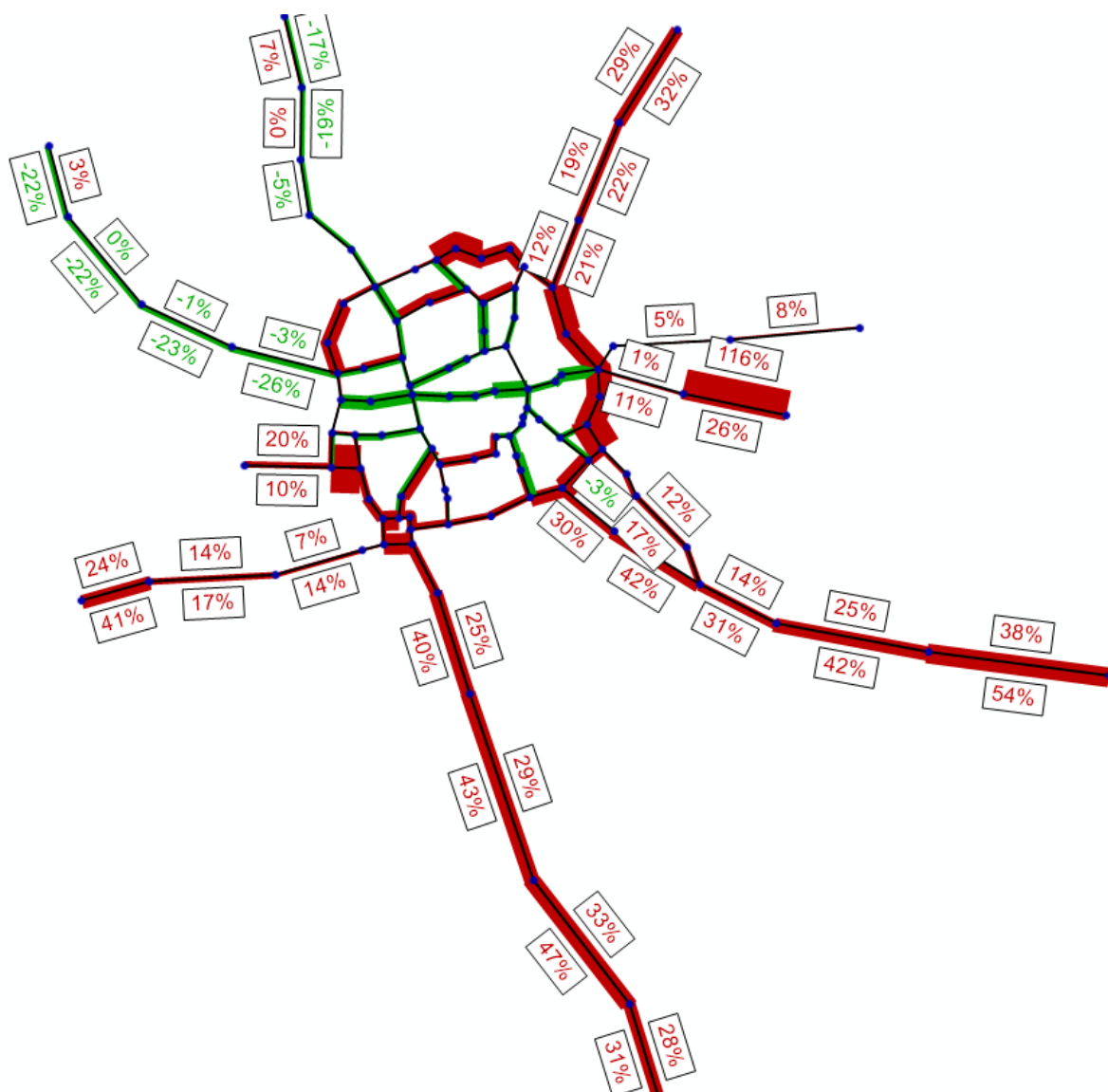
Rysunek 6.2. pokazuje szczególnie wartości na odnogach aglomeracji, na które mają wpływ drogi zewnętrzne. Największe zmiany w liczbie podróży można zaobserwować na odnodze południowej. Jednak pomimo tego, że wpływają na nią trzy drogi zewnętrzne, różnice w

wartościach procentowych nie przekraczają 50%. Na rysunku 6.2. można też zaobserwować źródła zmian w całej sieci transportowej do modelu bazowego. Największe przyrosty podróży z połączeń zewnętrznych mają odnogi zlokalizowane na południu oraz ta północnowschodnia. Jednocześnie odnogi na północnym zachodzie mają niedoszacowane liczby podróży. Wywołało to odbiegające od modelu bazowego obciążenie na osi wewnętrznej obwodnicy śródmiejskiej na całym jej odcinku, pomiędzy odnogami aglomeracji, na których doszło do przeszacowania liczby podróży. Tłumaczy to też odciążenie osi trasy W-Z przechodzącej przez środek obszaru miasta. Połączenia zasilające ją od zachodu i północy mają niedoszacowaną liczbę podróży i obciążenie sieci przeniosło się na oś południowy zachód – północny wschód.



Rys. 6.3 Wariant 1, widok na obszar centrum aglomeracji – wartości procentowe

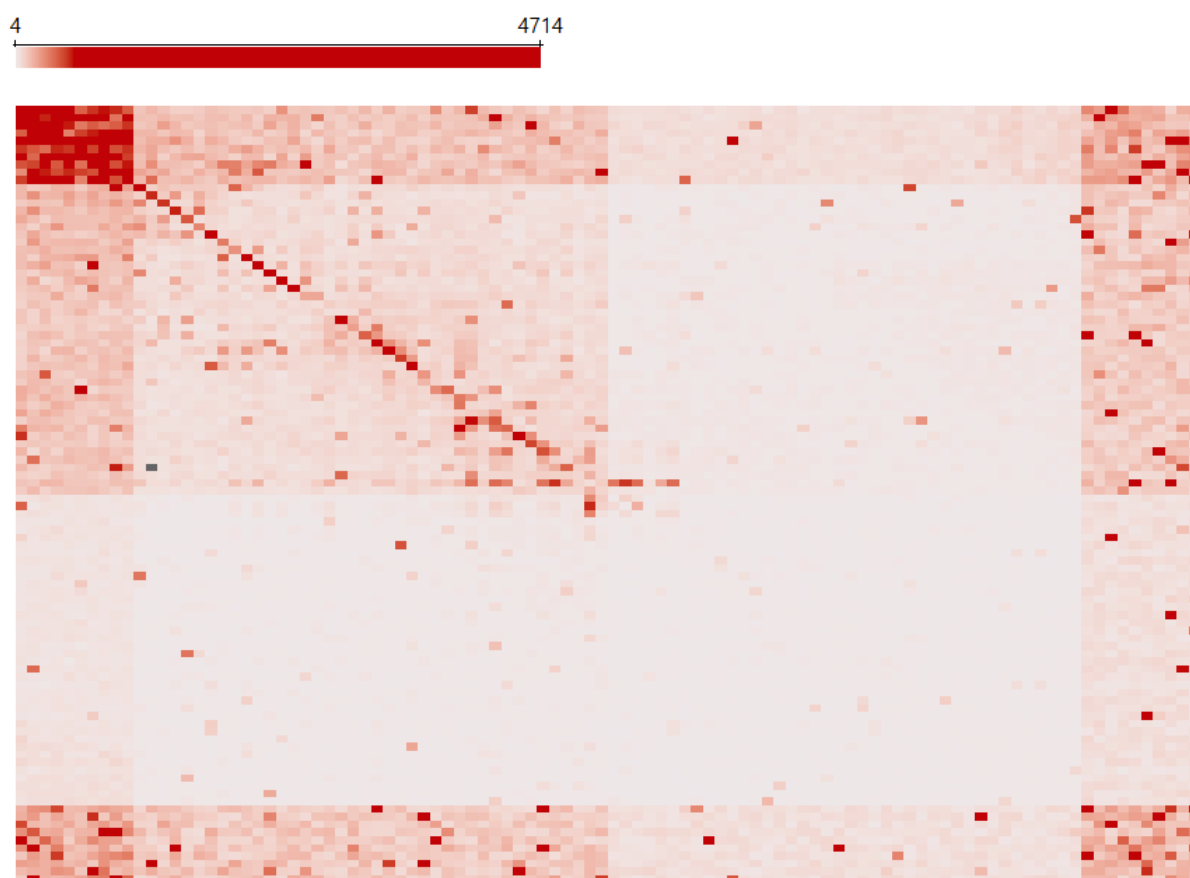
Na rysunku 6.3. widać zmiany wartości procentowych, a porównanie ich z rysunkiem 6.1. koryguje wcześniejsze wnioski. Na dwóch połączeniach największe zmiany w wartościach procentowych przekraczają nawet 100%. W przypadku połączenia na wschodzie ośrodka aglomeracyjnego (jest to połączenie z niewielką liczbą podróży w modelu bazowym) jest zaledwie 19 (!) podróży, tym samym zmiana globalnie ma nieistotny wpływ na całą sieć. Bardziej znaczącą zmianą jest połączenie widoczne na południowym zachodzie. Może być to związane ze zwiększeniem podróży opisanym wyżej w ramach zwiększenia natężenia ruchu na odnogach południowo-zachodnich i północno-wschodniej.



Rys. 6.4 Wariant 1, widok na obszar całej aglomeracji – wartości procentowe

Na rysunku 6.4. widać zmiany na połączeniach z drogami zewnętrznymi i arteriach wewnątrz „macek” aglomeracji. Ciekawy przypadek to odnoga na południowym wschodzie, gdzie im bliżej środka aglomeracji tym bardziej nadmiarowe podróże „giną”, dochodząc na jednym kierunku do pożądanej wartości -3% przy połączeniu się z obwodnicą śródmiejską. Może być to spowodowane większą liczbą rejonów „po drodze” i tym samym dokładność danych się zwiększa. Bardzo duży przeskok jest na połączeniu wschodnim środkowym, jednak jest to kwestia bardzo małej liczby podróży przez ten rejon, sięgających zaledwie kilkanaście/kilkadziesiąt. Znacząca część połączeń ma wartości bliskie (sięgające zaledwie kilku procent) w stosunku do modelu bazowego na połączeniach przyległych do obszaru centralnego aglomeracji (pomimo braku danych dotyczących dróg zewnętrznych w sześciu na dziewięciu źródłach big data). Tym samym rezultat jest oceniony pozytywnie i wskazuje to na

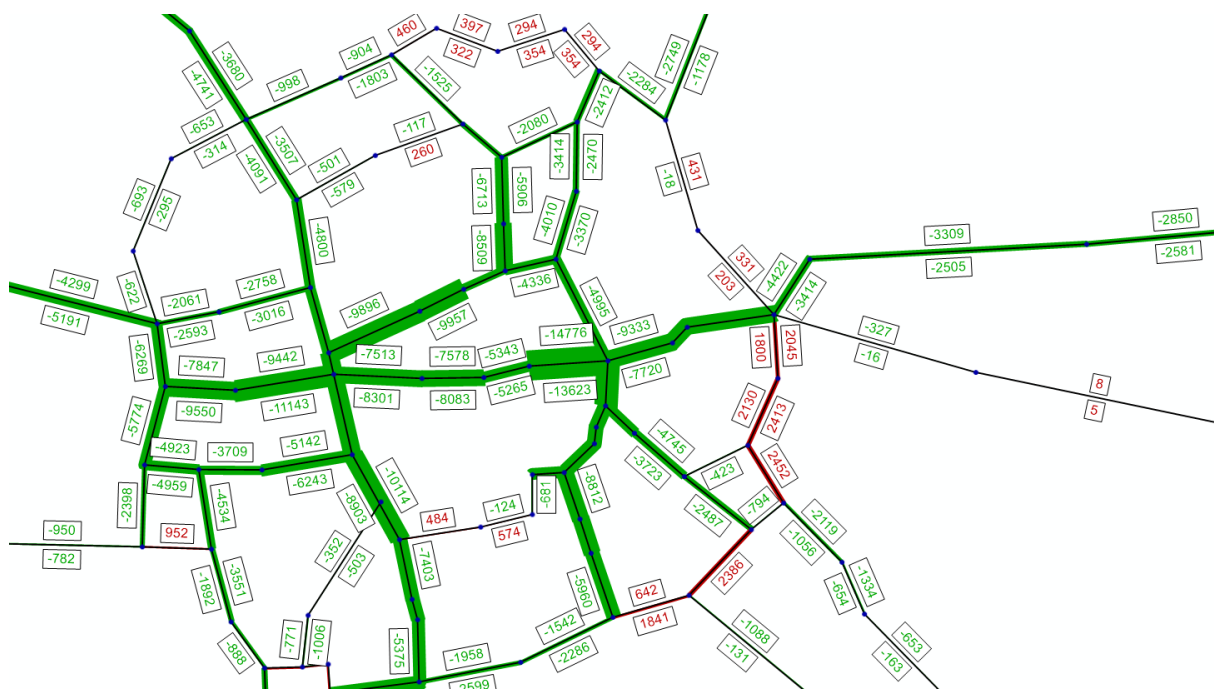
potrzebę korzystania z większej liczby źródeł przy tego typu połączeniach, by zbliżyć wszystkie wyniki do rzeczywistości i wpłynąć na mniejsze niedokładności w obszarze centralnym aglomeracji. Na rysunku 6.5. pokazano mapę ciepła, którą można porównać z mapą ciepła modelu bazowego na rysunku 5.10. Widać na niej zachowane proporcje spodziewane po poszczególnych relacjach między typami rejonów, choć występują lokalne anomalie spowodowane ekstremalnymi wartościami z niektórych macierzy, szczególnie widoczne np. pochodzące z danych z ANPR w obszarze górnej połowy przekątnej macierzy. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 4.



Rys. 6.5 Wariant 1, mapa ciepła macierzy podróży

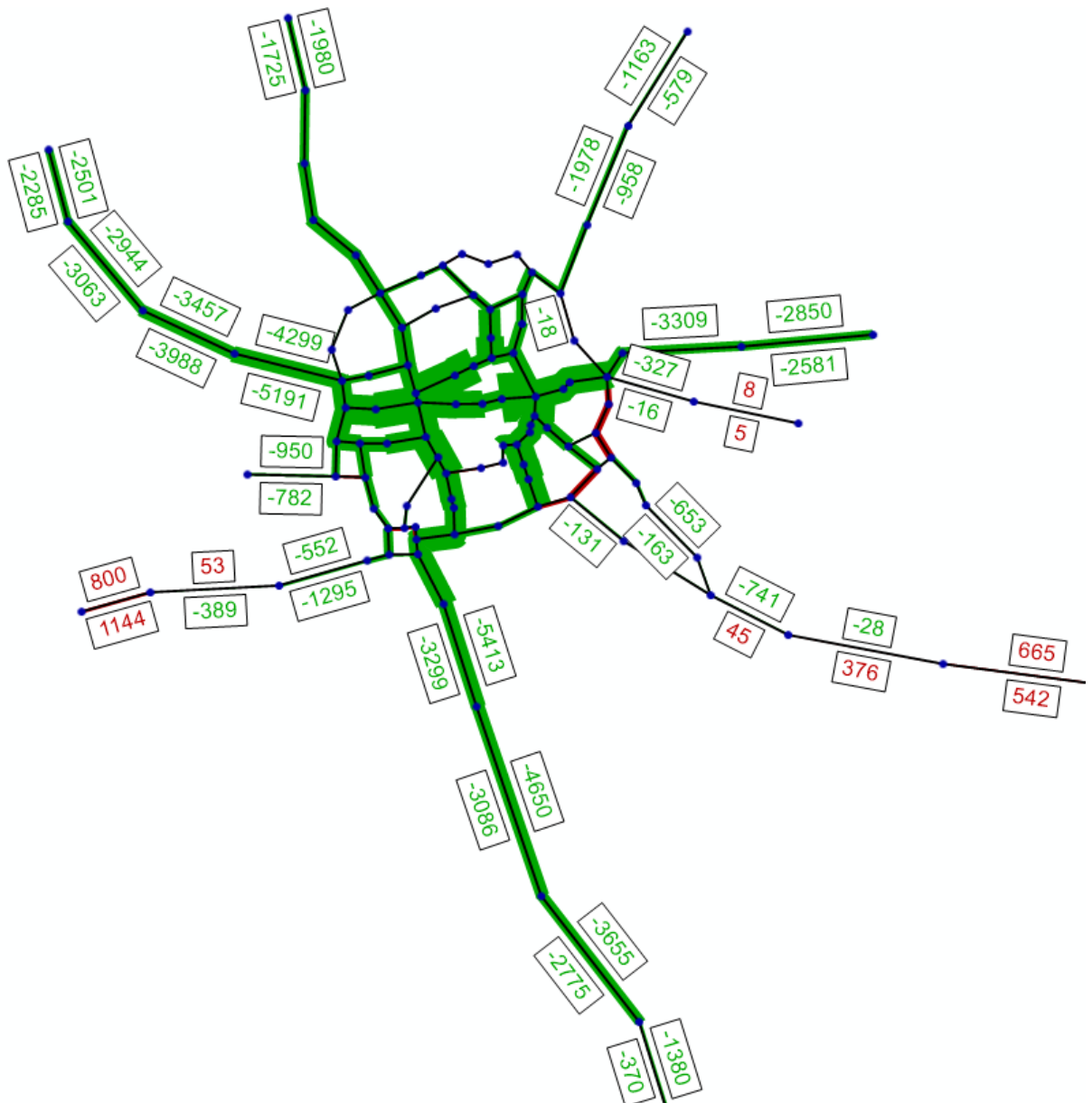
6.2.2. Wariant 2 - Uzupełnienie źródłami symulowanymi na bazie map ciepła i charakterystyki źródła danych

Drugi wariant miał na celu sprawdzenie czy dane symulowane nie będą prezentować bliższych wartości w stosunku do modelu bazowego niż wariant ze wszystkimi danymi, ponieważ zostały dopasowane do charakteru symulowanego ośrodka miejskiego od podstaw. Symulowano dane z mediów społecznościowych, Taxi oraz UTO + rowerów.



Rys. 6.6 Wariant 2, widok na obszar centrum aglomeracji

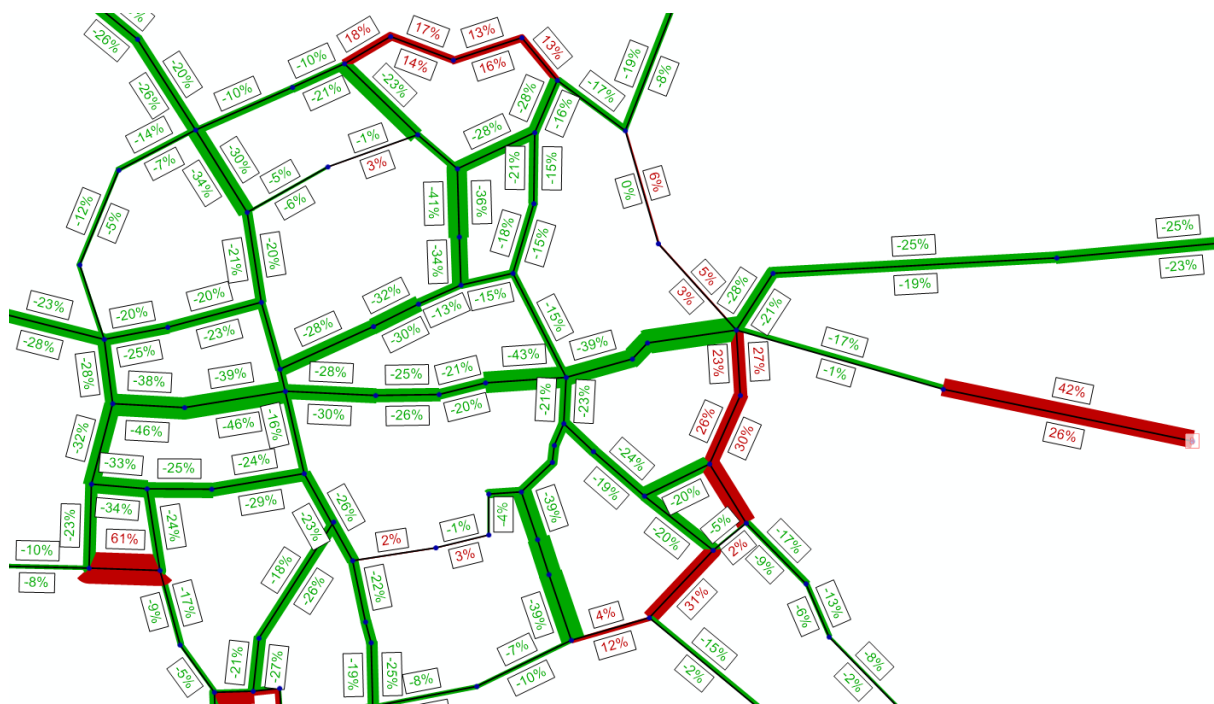
Rysunek 6.6. wskazuje na znacznie większą niedokładność niż model wykorzystujący wszystkie dostępne źródła danych. Zdecydowanie większa liczba połączeń ma mniejszą liczbę podróży niż model bazowy. Liczby na połączeniach są również większe. Największy widoczny niedobór liczby podróży wystąpił podobnie na wschodnim odcinku trasy W-Z wewnątrz obszaru ścisłego centrum. Liczby są nieznacznie większe o około 1000 podróży w każdym kierunku. Różnice są również mniejsze na obwodnicy śródmiejskiej, gdzie sytuacja uległa poprawie w stosunku do wariantu uwzględniającego wszystkie źródła danych. Biorąc pod uwagę całość obszaru centralnego aglomeracji, sytuacja jest daleka od ideału, obszar centralny ma większe niedobory podróży, a co za tym idzie, w przypadku hipotetycznego planowania transportu publicznego na bazie niniejszych danych, założenia mogą nie spełniać potrzeb transportowych podróżnych. Rezultat w porównaniu do samej liczby podróży jest bardziej optymistyczny niż mogłoby się wydawać. Spora część połączeń jest bliższa modelowi bazowemu. Niekorzystne jest jednak, że są to głównie niedobory podróży i cała sieć nie „równoważy” się jak w wariantcie pierwszym, gdzie nie ma wyraźnej przewagi niedoborów lub nadmiarów podróży.



Rys. 6.7 Wariant 2, widok na obszar całej aglomeracji

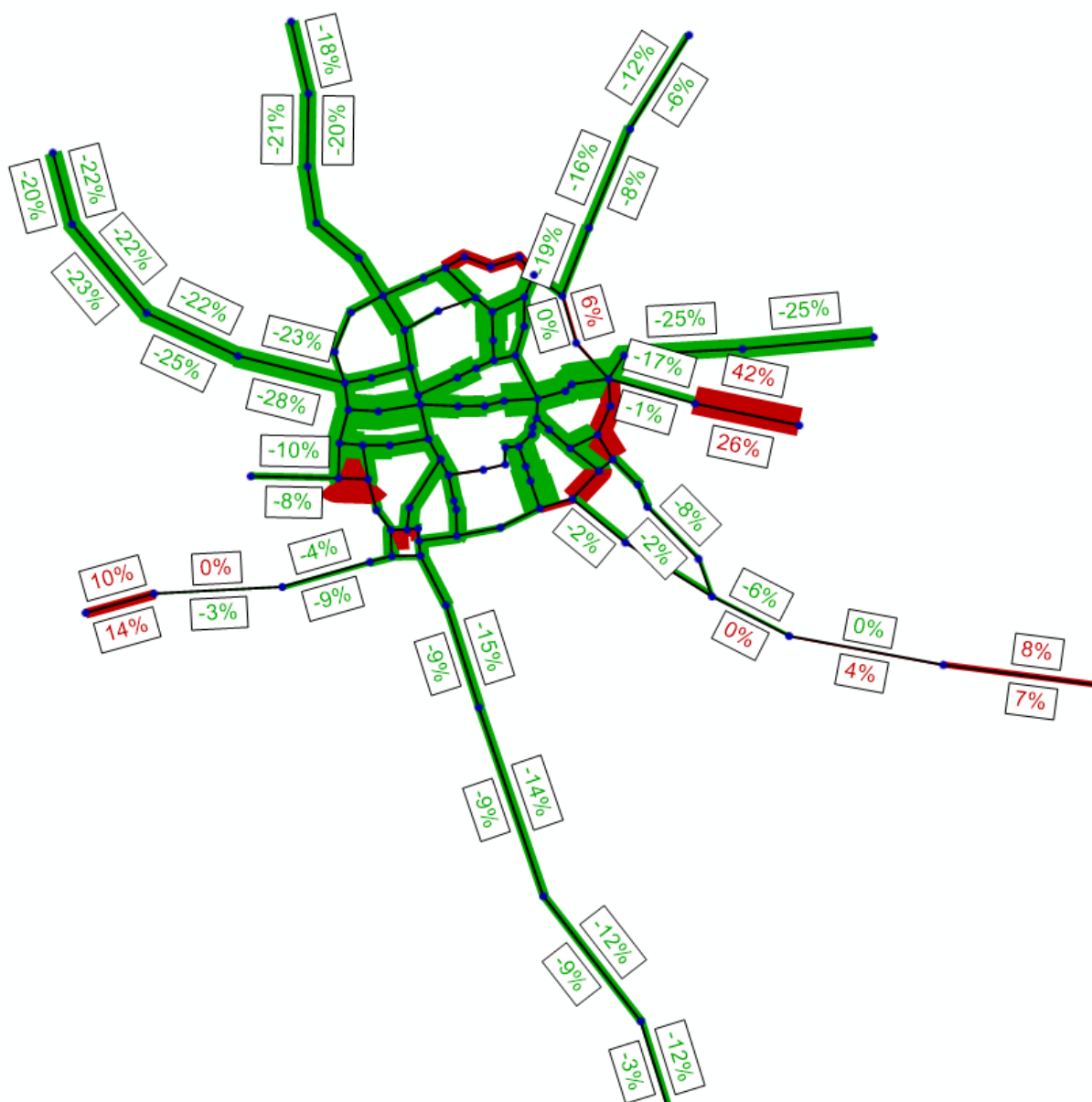
Rysunek 6.7 natomiast wskazuje na dużą dokładność połączeń zewnętrznych, na rysunku 6.6. widać też jak wpłynęło to na obwodnicę śródmiejską - niedokładności są znacznie mniejsze. Biorąc pod uwagę odnogi aglomeracji, oraz obwodnicę śródmiejską, wyniki prezentują większą zgodność z modelem bazowym, ale liczba podróży nie równoważy się, większość połączeń rejestruje ubytki w liczbie podróży. Powodować może to skuteczne zamodelowanie macierzy podróży z mediów społecznościowych. Jest to baza danych, którą trudno było zróżnicować na bazie literatury w stosunku do np. długości podróży jak w przypadku danych z rowerów miejskich i UTO, tak samo jak danych z Taxi. Te bazy danych nie były uwzględniane

w rejonach symbolizujących drogi zewnętrzne, więc w odnogach aglomeracji za większość przepływów podróży odpowiada tylko jedno ze źródeł danych.



Rys. 6.8 Wariant 2, widok na obszar centrum aglomeracji – wartości procentowe

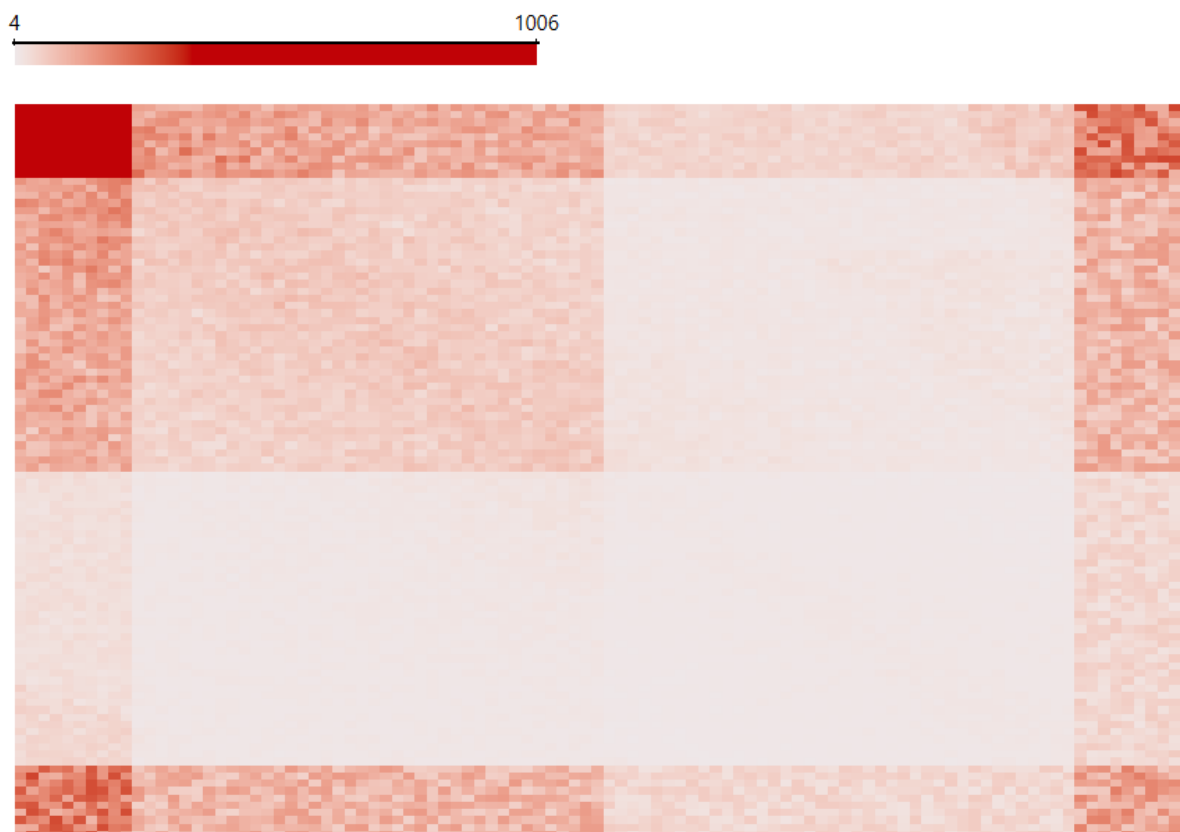
Rysunek 6.8 w porównaniu do analogicznego rysunku 6.3 ma mniejsze wartości ekstremalne, nie występują na nim rozbieżności powyżej 100%. Śledząc kluczowe połączenia jest zdecydowanie więcej połączeń mających niedobory ruchu i są to większe niedobory. Na większości połączeń oscylują w okolicach wartości 20-40%, gdzie na analogicznych połączeniach w wariantcie 1 to wartości bliższe 10-25%. Przez ich ogólną przewagę nad nadmiarami podróży rozbieżności są większe w porównaniu do użycia wszystkich źródeł big data z poprzedniego wariantu, stosującego „pełną metodę” wykorzystania wszystkich dostępnych źródeł. Poprzez ogólny niedobór podróży połączenia na których są nadwyżki ruchu prezentują mniejsze wartości procentowe rozbieżności niż w wariantcie 1, co może wydawać się lepszym rezultatem, jednak odnosząc się do samej macierzy podróży, rozbieżność jest znacząca, ponieważ takich połączeń jest znacząco mniej w wariantcie 2.



Rys. 6.9 Wariant 2, widok na obszar całej aglomeracji – wartości procentowe

Rysunek 6.9. pokazuje odnogi aglomeracji w wartościach procentowych. W porównaniu do rysunku 6.5. w wartościach procentowych jest bliższy modelowi bazowemu niż wariant symulowany ze źródeł big data. Może to być spowodowane ogólnie zaniżoną liczbą podróży w całej macierzy z wariantu 2, a w wariantcie 1 uwzględniającym wszystkie źródła - tych dotyczących dróg zewnętrznych jest znacząco mniej, tylko trzy na dziewięć. Zdaniem autora jest to spowodowane dobrym doszacowaniem pojedynczego źródła danych – danych z mediów społecznościowych ponieważ było to źródło symulowane. Pozostałe dwa źródła uwzględniane w wariantcie 1 wpływają zatem negatywnie na wyniki. Pokazuje to potrzebę synchronizowania większej liczby źródeł danych lub korzystania z tych lepszych jakościowo. Mapę ciepła dla

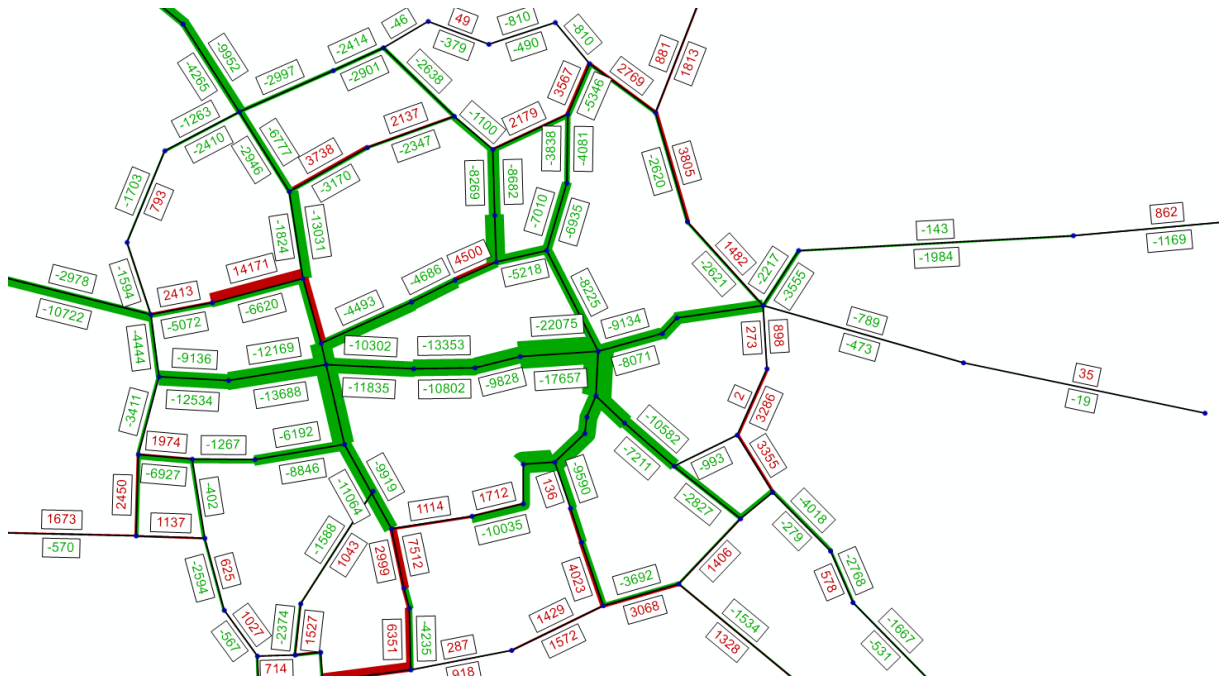
tego wariantu pokazano na rysunku 6.10. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 5.



Rys. 6.10 Wariant 2, mapa ciepła macierzy podróży

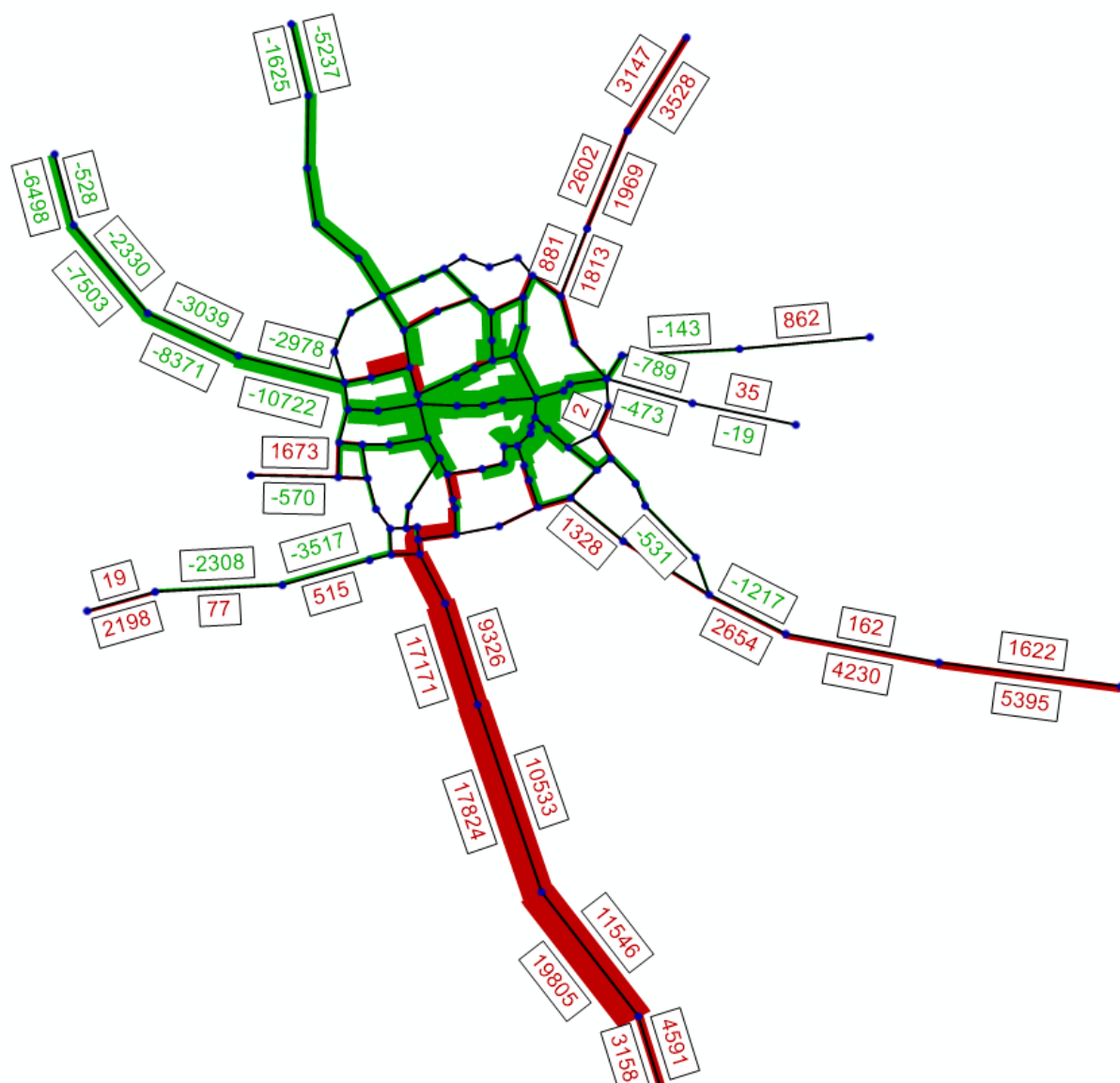
6.2.3. Wariant 3 – uzupełnienie danymi z wrocławskiego KBR z 2018

Wariant 3 jest kolejną propozycją, która w założeniu powinna dostarczać poprawne wyniki. Są to jakościowe dane pochodzące z gotowych macierzy podróży z KBR. We wrocławskim KBR w 2018 roku wykorzystano dane z big data - były to dane z kart SIM od jednego operatora sieci komórkowej oraz dane z sondowania pojazdów służbowych od jednego producenta. Niestety, w KBR 2024 w dużym zakresie nie wykorzystano danych z Big Data, użyto tylko danych od operatora sieci telefonii komórkowej w celu szacowania liczby osób pracujących w danych rejonach. Dane z KBR 2018 zostały przeskalowane do symulowanego ośrodka miejskiego. Obejmują wartości z dwóch macierzy obejmujących pełen obszar, łącznie z odnogami aglomeracji i drogami zewnętrznymi.



Rys. 6.11 Wariant 3, widok na obszar centrum aglomeracji

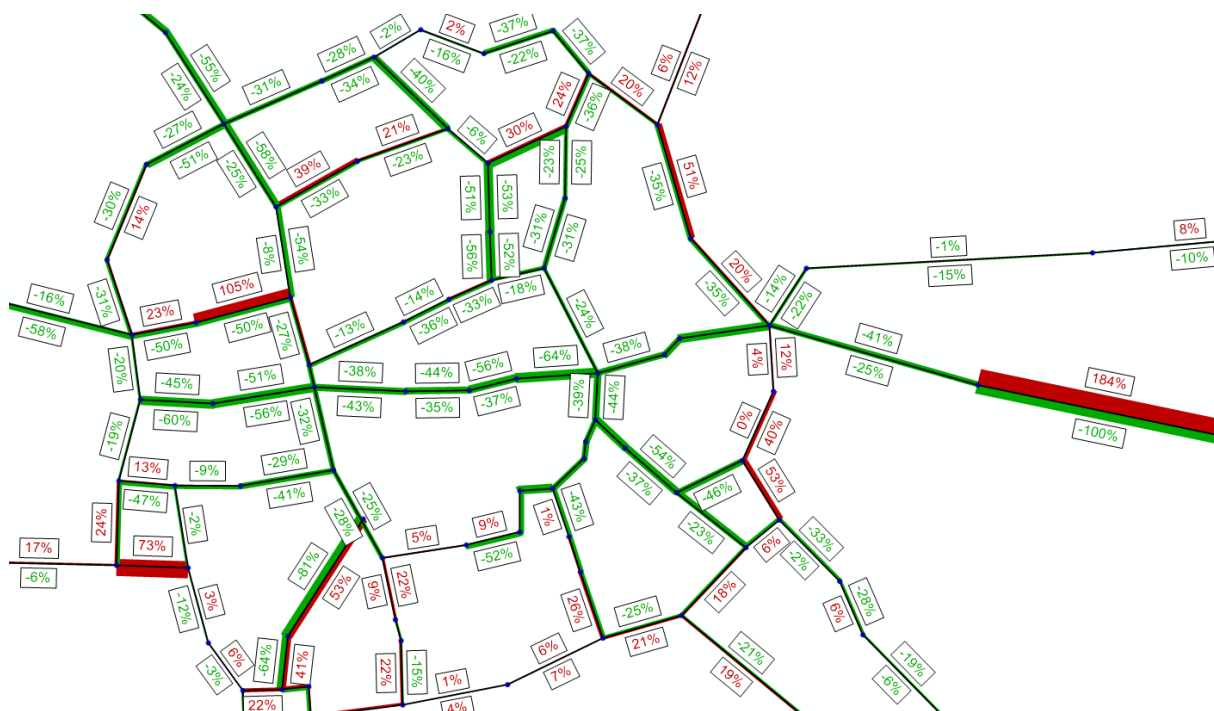
Rysunek 6.11 wskazuje na dużą zgodność danych w obszarze obwodnicy śródmiejskiej. Miejscami prezentuje wartości bliższe modelowi bazowemu niż w wariantach 1 i 2, niemniej w wartościach liczbowych jest zbliżony. Mocno niedokładne wyniki, przekraczające 10000 brakujących podróży na pojedynczych połączeniach, daje sieć połączeń w obszarze wewnątrz obwodnicy. Wartości liczbowe niedoborów ruchu są znacznie większe niż w wariantcie 1. Na trasie W-Z są większe także od wariantu 2. Jest to powód jeszcze większej rozbieżności liczby podróży od modelu bazowego niż w poprzednich wariantach.



Rys. 6.12 Wariant 3, widok na obszar całej aglomeracji

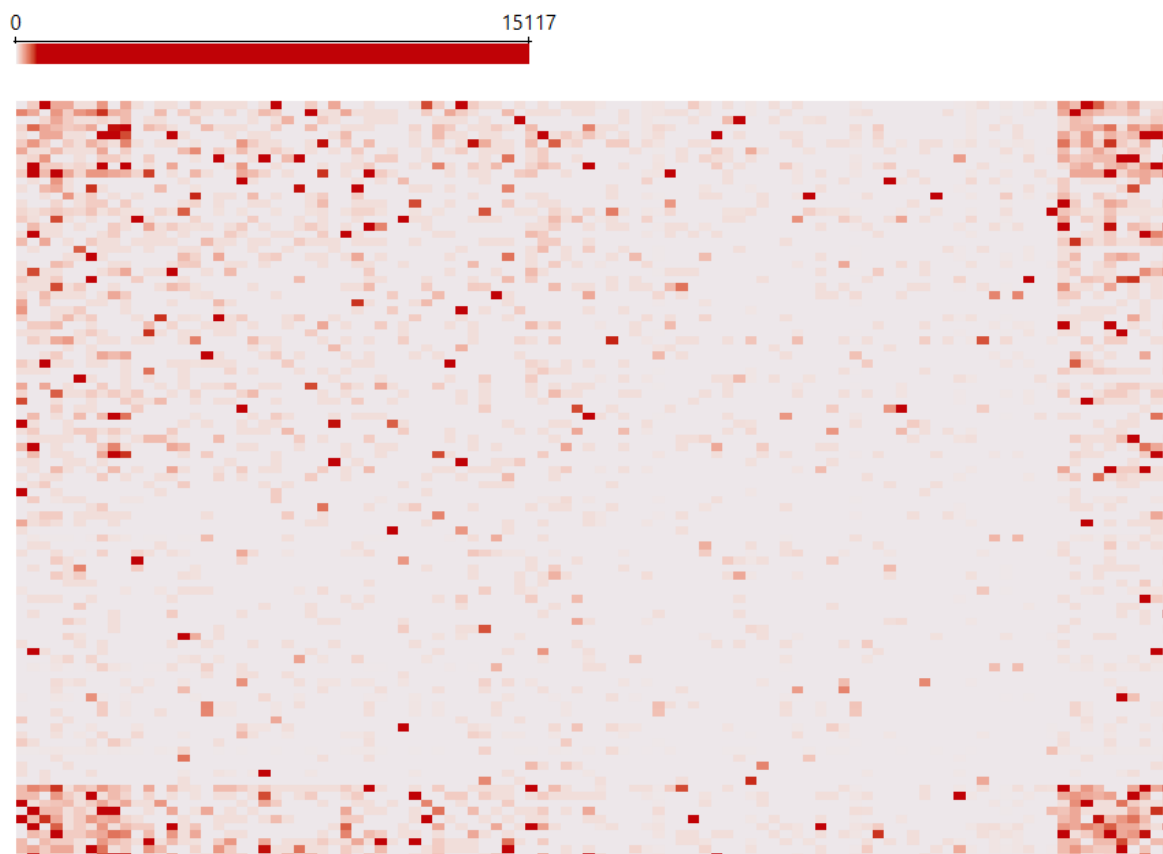
Rysunek 6.12 pokazuje tendencje podobne do rysunku 6.2 pokazującego ten sam widok dla wariantu 1. Odnogi północnozachodnie mają większe niedobory ruchu, a odnoga południowa większe nadmiary ruchu, nawet o kilka tysięcy podróży. Reszta odnóg miała wartości bliskie modelowi bazowemu w wariantie 1, tutaj są one zbliżone, a nawet nieznacznie bliższe w niektórych przypadkach. Ze względu na większe rozbieżności w kluczowych połączeniach przykład jest oceniony negatywnie i dane zsynchronizowane z większą liczbą baz danych dają lepsze rezultaty. Większe rozbieżności mogą być powodem, niesatysfakcjonujących wyników na odnogach aglomeracji wariantie 1. W tym przypadku obydwie bazy danych obejmują wszystkie typy rejonów. Zatem są to „brakujące” dane w porównaniu do wariantu 2, który zawierał ostatnią z trzech baz danych – obejmującą dane o podróżach z rejonów zewnętrznych.

Analizując dane z rysunku 6.12 można dojść do wniosku, że dane z wrocławskiego KBR były znacząco słabsze od bazy z mediów społecznościowych i wpłynęły na zaburzenie wartości na przykładzie z wariantu 1 uwzględniających całą metodę.



Rys. 6.13 Wariant 3, widok na obszar centrum aglomeracji – wartości procentowe

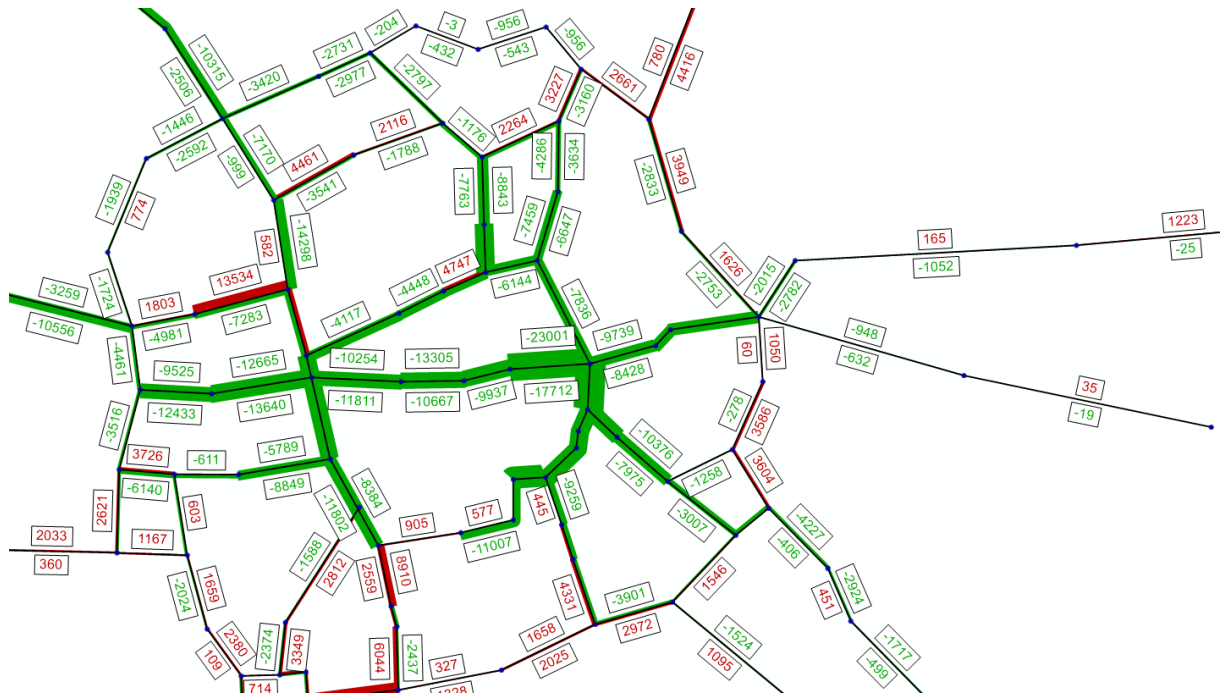
Wartości procentowe na rysunku 6.13. potwierdzają opis rysunku 6.11. Są w niektórych połączeniach znacznie większe od wartości z wariantu 1 przedstawionych na rysunku 6.3. Na trasie W-Z przekraczają ponad 60% i wykazują około 2x większy niedobór podróży niż w wariantcie 1. Podobnie jak w przypadku rysunku 6.3. dolna ze wschodnich odnóg aglomeracji ma duże wartości procentowe, co jest spowodowane bardzo małą liczbą podróży na tym połączeniu (widocznych na rysunku 6.11). Odnogi aglomeracji widać w całości na rysunku 6.14. Na rysunku 6.15. pokazano macierz podróży w postaci mapy ciepła. Jest zdecydowanie bardziej odbiegająca od modelu bazowego i poprzednich wariantów. Macierz podróży w formie liczbowej załączono do pracy w Zał. 6.



Rys. 6.15 Wariant 3, mapa ciepła macierzy podróży

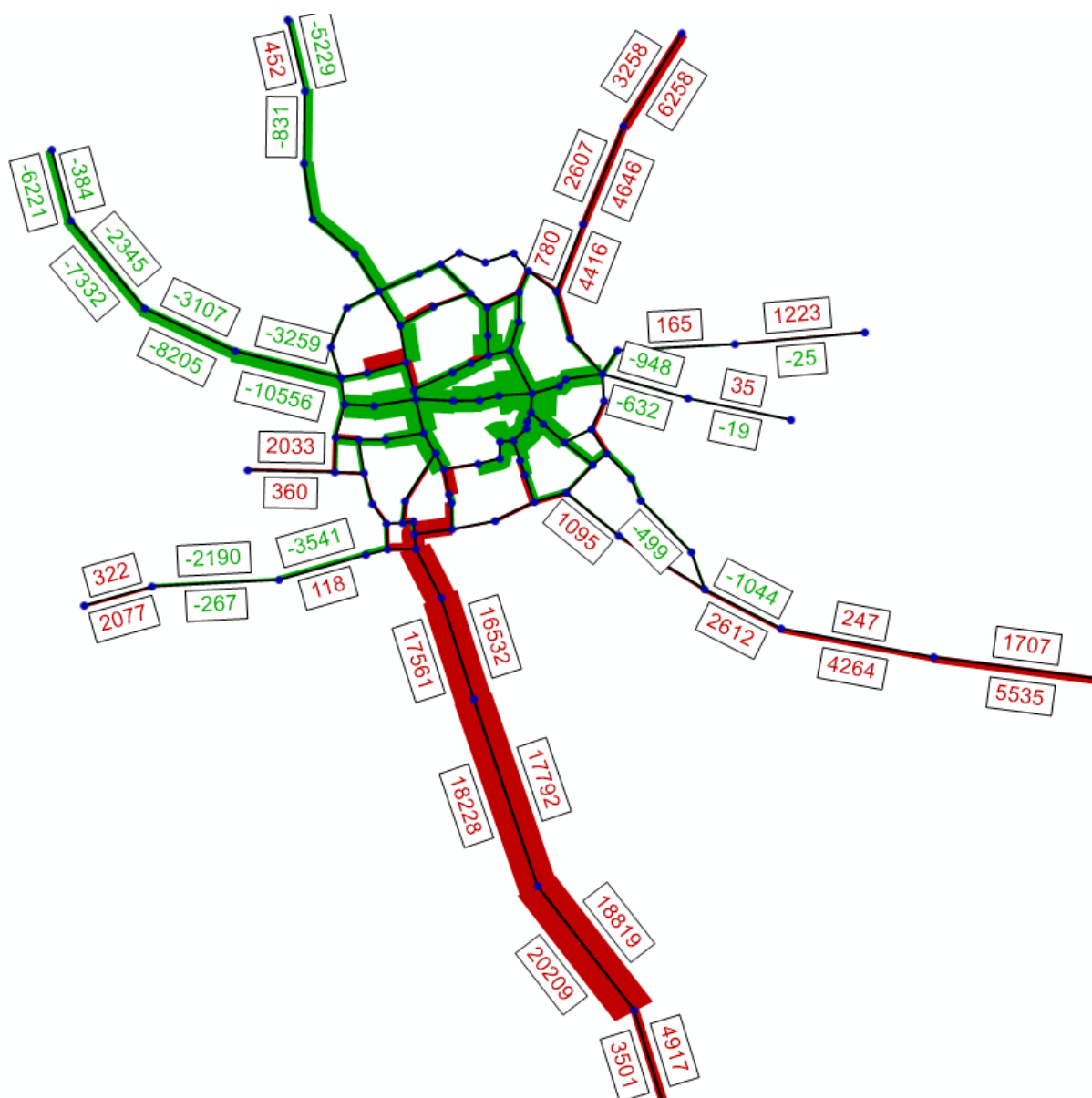
6.2.4. Wariant 4 – sam GSM

Na bazie wariantu poprzedniego – trzeciego stwierdzono, że warto przebadąć jakie rezultaty da tylko jedno ze źródeł w porównaniu do modelu bazowego. Wariant czwarty w zasadzie nie jest wariantem samym w sobie, a porównaniem modelu bazowego do pojedynczej bazy danych pochodzącej z GSM. Stwierdzono, że takie porównanie jest kluczowe by zbadać jaki wpływ na metodę ma zestawienie dobrego i słabego źródła danych. W tym przypadku są to źródła danych z tego samego miejsca – wrocławskiego KBR, gdzie użyto danych z GPS – z sondowania pojazdów oraz dane z sieci GSM. Dane z sondowania pojazdów posiadały dużo braków w podróżach pomiędzy rejonami – na 10000 liczb w macierzy, aż 9348 to zera, a sumaryczna liczba podróży to zaledwie 627. W związku z tym, wariant 4 zawiera wyłącznie GSM, przy założeniu, że wyniki dla samego sondowania pojazdów – GPSa dane będą znacząco gorsze.



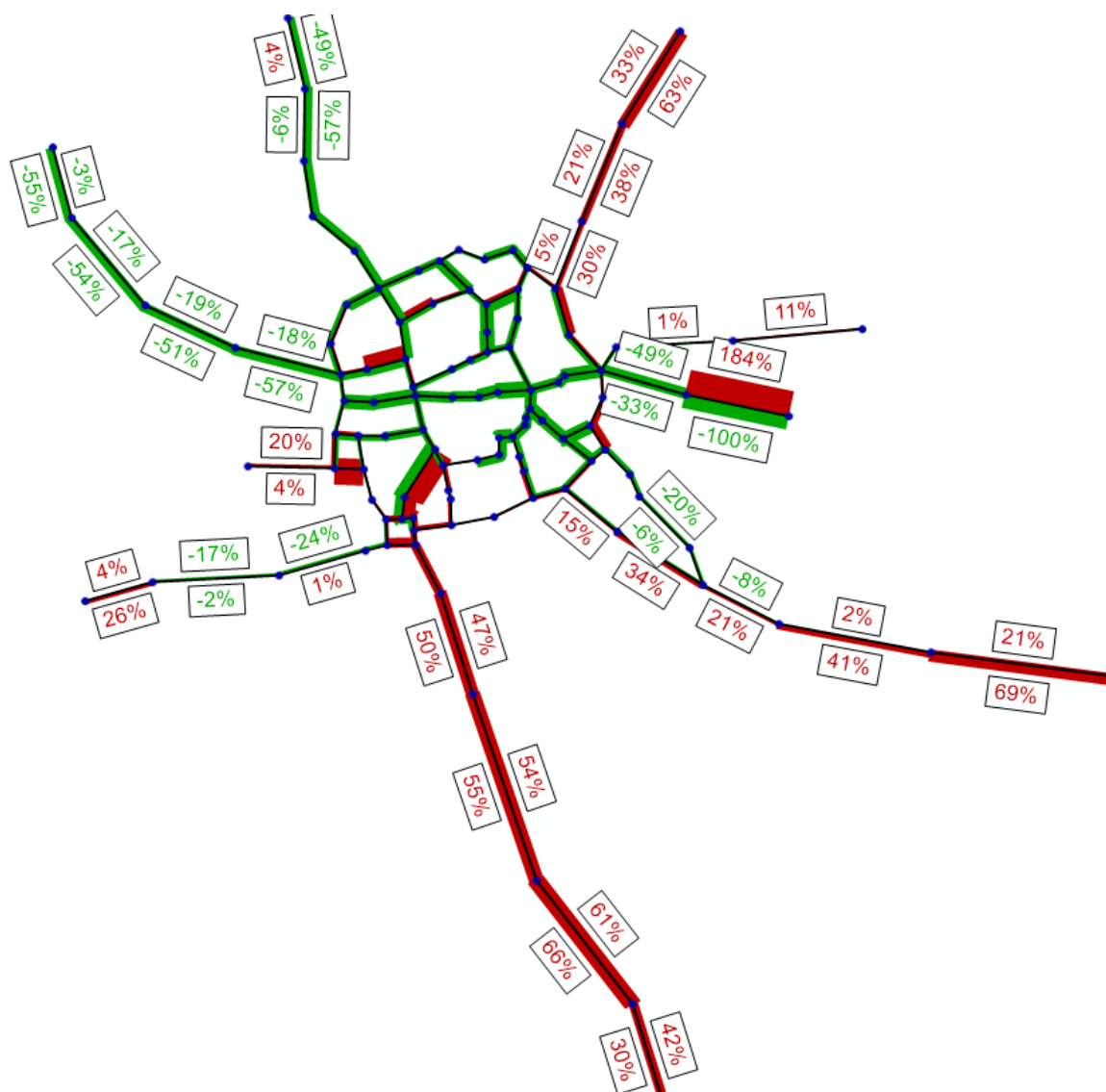
Rys. 6.16 Wariant 4, widok na obszar centrum aglomeracji

Rys. 6.16. bez przyjrzenia się szczegółom jest bardzo podobny do poprzedniego wariantu i odpowiadającemu mu rys. 6.11. Nie jest to przypadek, ponieważ to dwa najbliższe sobie warianty. Przez znikomy wpływ danych z sondowania pojazdów na części połączeń występują te same liczby, a te różniące się są do siebie zbliżone rzędem. Na rysunku widać różnice rzędu kilkuset podróży. Tylko w niektórych miejscach, liczby różnią się o ponad 1000.



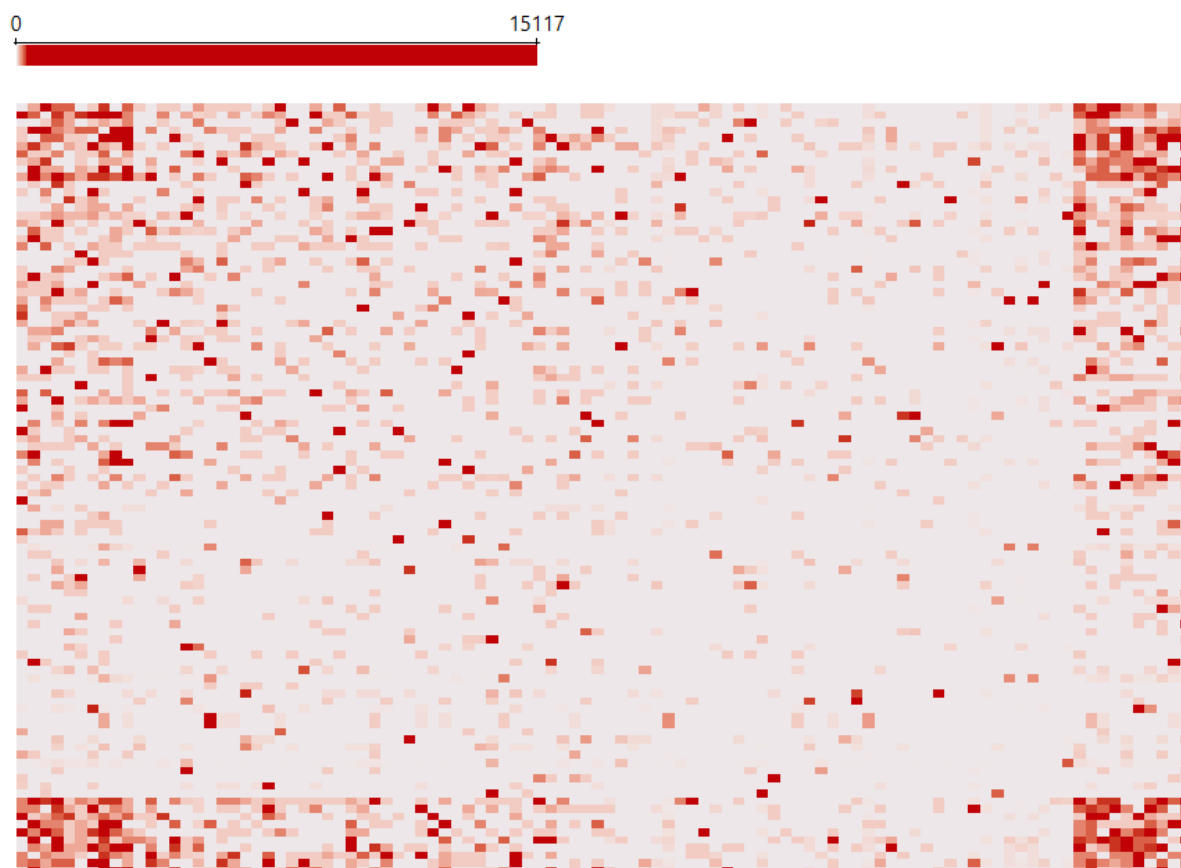
Rys. 6.17 Wariant 4, widok na obszar całej aglomeracji

Rysunek 6.17 pokazuje ciekawszą zmianę, którą w niewielkiej skali widać również na rys. 6.16. Większość odnóg aglomeracji wykazuje różnice rzędu kilkuset podróży, jednak odnoga południowa ma kilkutysięczne rozbieżności. Większe niezgodności w stosunku do wariantu 3 widać też na odnodze północnej - tej po lewej stronie. Ma to również odzwierciedlenie na rys. 6.16, można na nim zaobserwować największe różnice z wariantem 3 na połączeniu o obszarze centralnym, który łączy te dwie odnogi.



Rys. 6.19 Wariant 4, widok na obszar całej aglomeracji – wartości procentowe

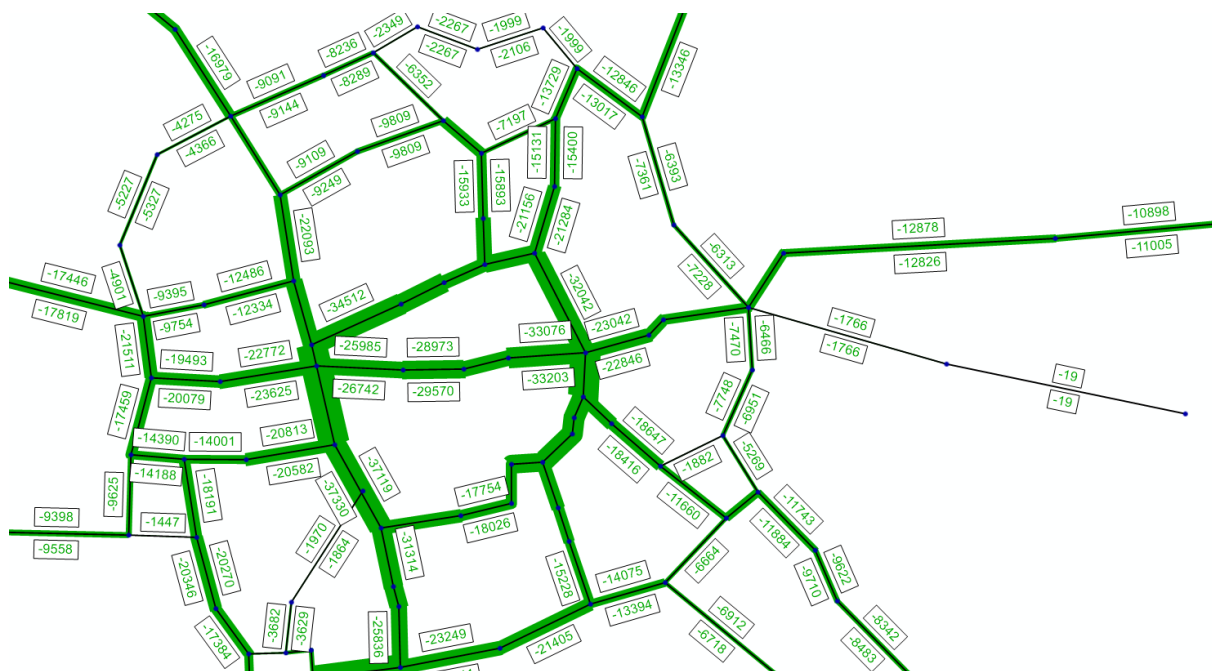
Ostatni rysunek przedstawiający wariant 4 pokazuje duże różnice na odcinku południowej, jednak tylko na relacji północnej. Relacja w odwrotnym kierunku na tych połączeniach nie osiąga dwucyfrowych różnic, gdy porówna się ją z rysunkiem 6.14. Podobnie jest na odcinku północnej lewej. Sumarycznie wszystkie różnice wypadły korzystnie, patrząc na ogólną liczbę podróży w tym wariantcie. Wynik jest nieznacznie bliższy wartości w stosunku do modelu bazowego niż w przypadku połączenia z danymi z GPS. Na rysunku 6.20. pokazano macierz podróży w formie mapy ciepła. Jest ona podobna do tej z wariantu 4. Macierz podróży w formie liczbowym załączono do pracy w Zał. 7.



Rys. 6.20 Wariant 4, mapa ciepła macierzy podróży

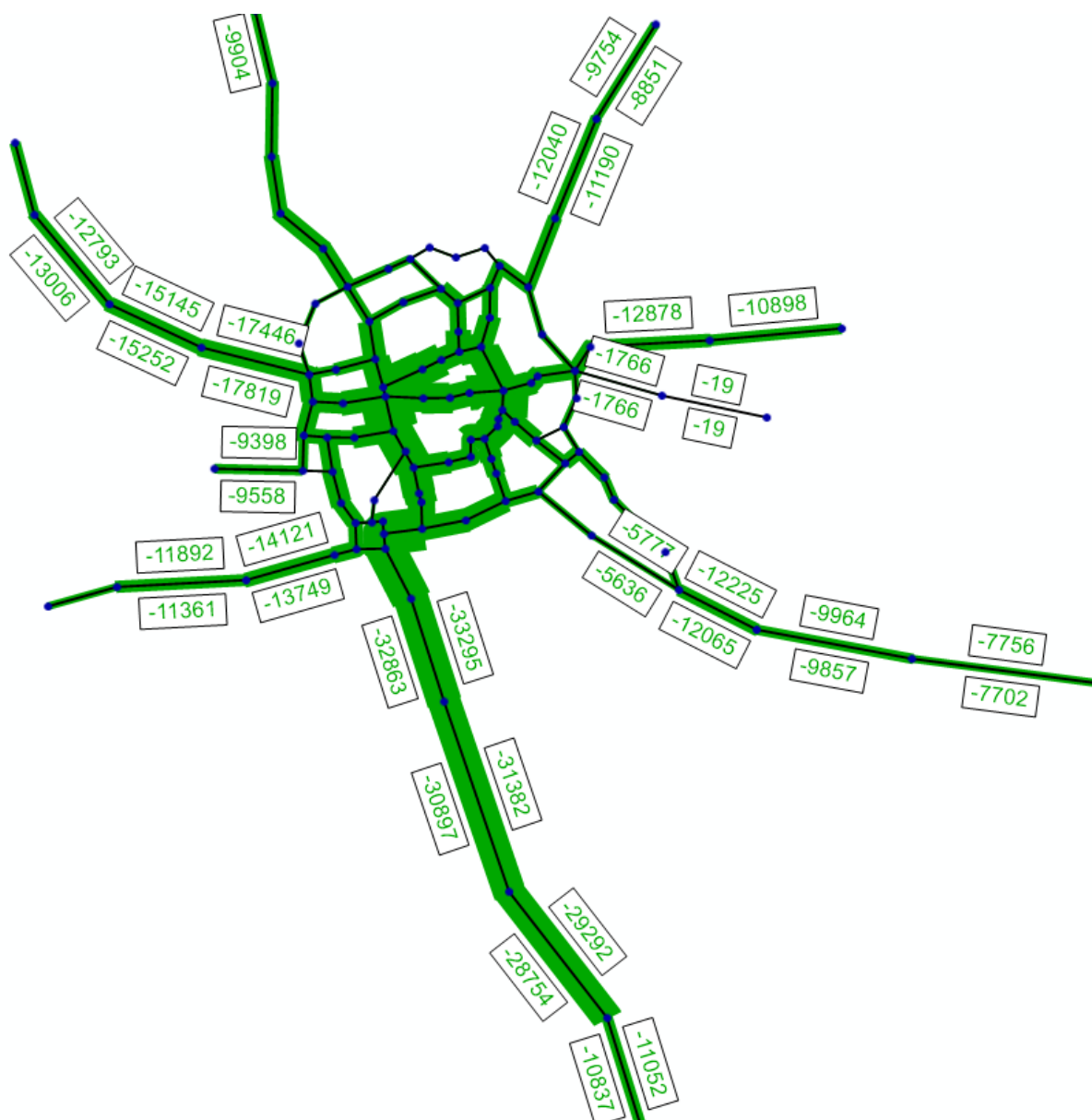
6.2.5. Wariant 5 – Bluetooth + APC + SmartCard Data + ANPR

Wariant piąty jest połączeniem wszystkich źródeł które dają tylko wycinek aglomeracji. Są to źródła cząstkowe, więc jest to pierwszy wariant który nie miał na celu znalezienia „dobrego” wariantu. Celem było sprawdzenie jaki rezultat dało połączenie źródeł, które dawały ograniczone wyniki. Ze względu na dużą liczbę rejonów, z których nie ma danych, spodziewany jest skrajnie duży niedomiar podróży. Są to dane, które należałoby wykorzystać jako pomocnicze przy źródłach mających większe pokrycie relacji pomiędzy rejonami transportowymi.



Rys. 6.21 Wariant 5, widok na obszar centrum aglomeracji

Rysunek 6.21 posiada największe jak dotąd niedobory podróży w porównaniu z pozostałymi wariantami. Taki rezultat nie pozwoli na poprawne oszacowanie liczby podróży i pokazuje, że połączenie nawet czterech źródeł, jednak mocno wybrakowanych da niewystarczający rezultat. Widać to jeszcze mocniej na rysunku 6.23 na wartościach procentowych. Na żadnym z połączeń nie ma wartości nadmiarowych podróży. Pomimo barwy zielonej, która mogłaby sugerować „dobre” wyniki w niniejszych wariantach, wartościami pożądanymi są jak najmniejsze wartości bezwzględne połączeń. W tym przypadku wykorzystanie takiego modelu mogłoby doprowadzić do niedoszacowania infrastruktury. Wykorzystanie tych czterech źródeł danych ma jednak wartość dodaną do docelowej metody, ponieważ relacje z zerowymi danymi, które doprowadziły do zaniżenia wyników, nie są uwzględniane. Relacje, z których źródło ma dane, mogą doprowadzić do uszczegółowienia tych pojedynczych relacji w metodzie docelowej.



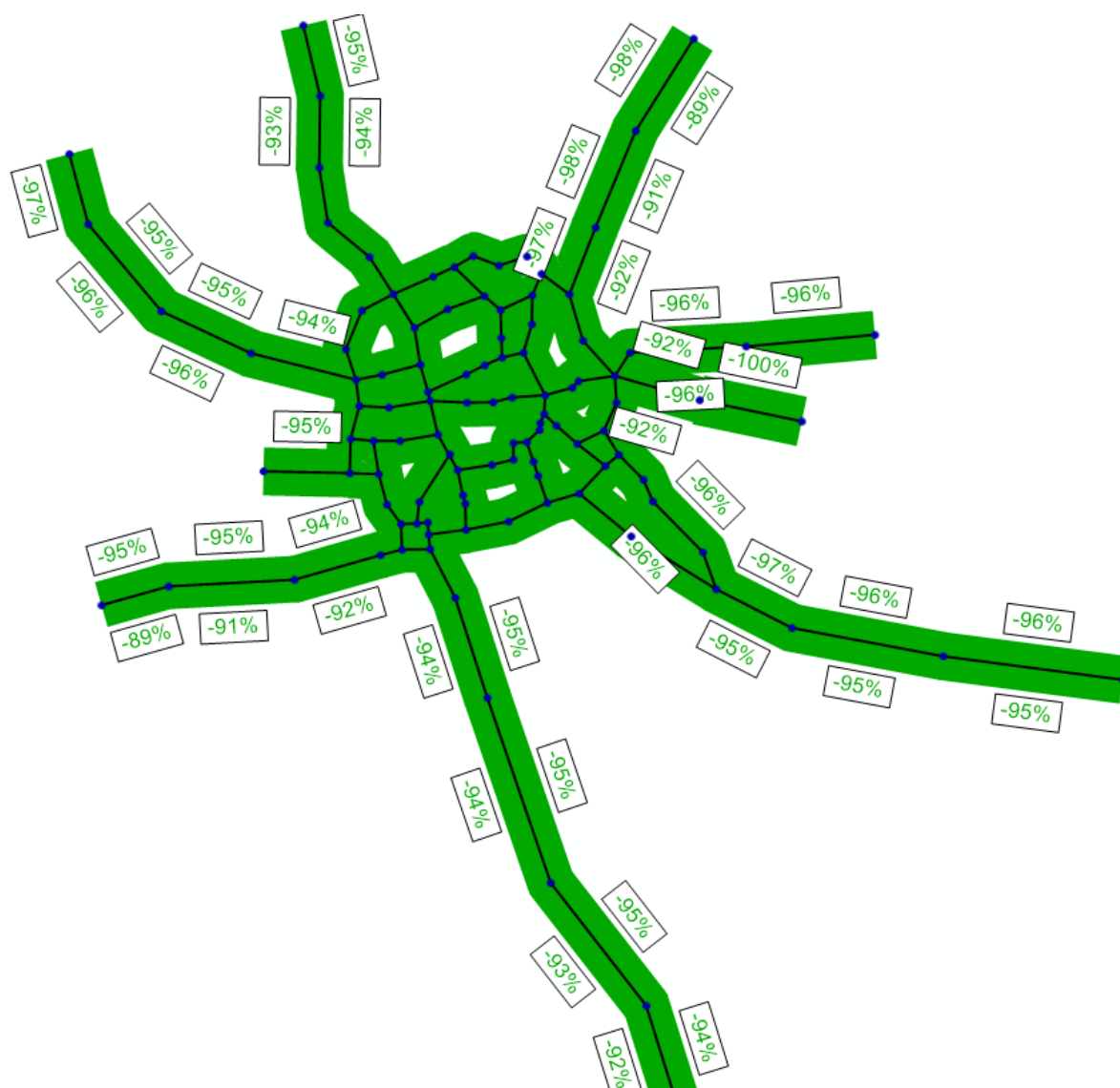
Rys. 6.22 Wariant 5, widok na obszar całej aglomeracji

Rysunek 6.22. pokazuje odnogi aglomeracji, na których wyniki też są mocno odbiegające od modelu bazowego. Nie są to wartości w 100% niezgodne, ponieważ dane szcątkowe dotyczyły rejonów odnóg aglomeracji, pomimo braku danych z dróg zewnętrznych. Na jednym z połączeń na wschodniej części aglomeracji widać niewielkie wartości (-19 podróży), w modelu bazowym jest tam 19 podróży w obydwie strony, czyli rozbieżność w tym przypadku jest skrajna, nie ma tam w ogóle danych o podróżach.



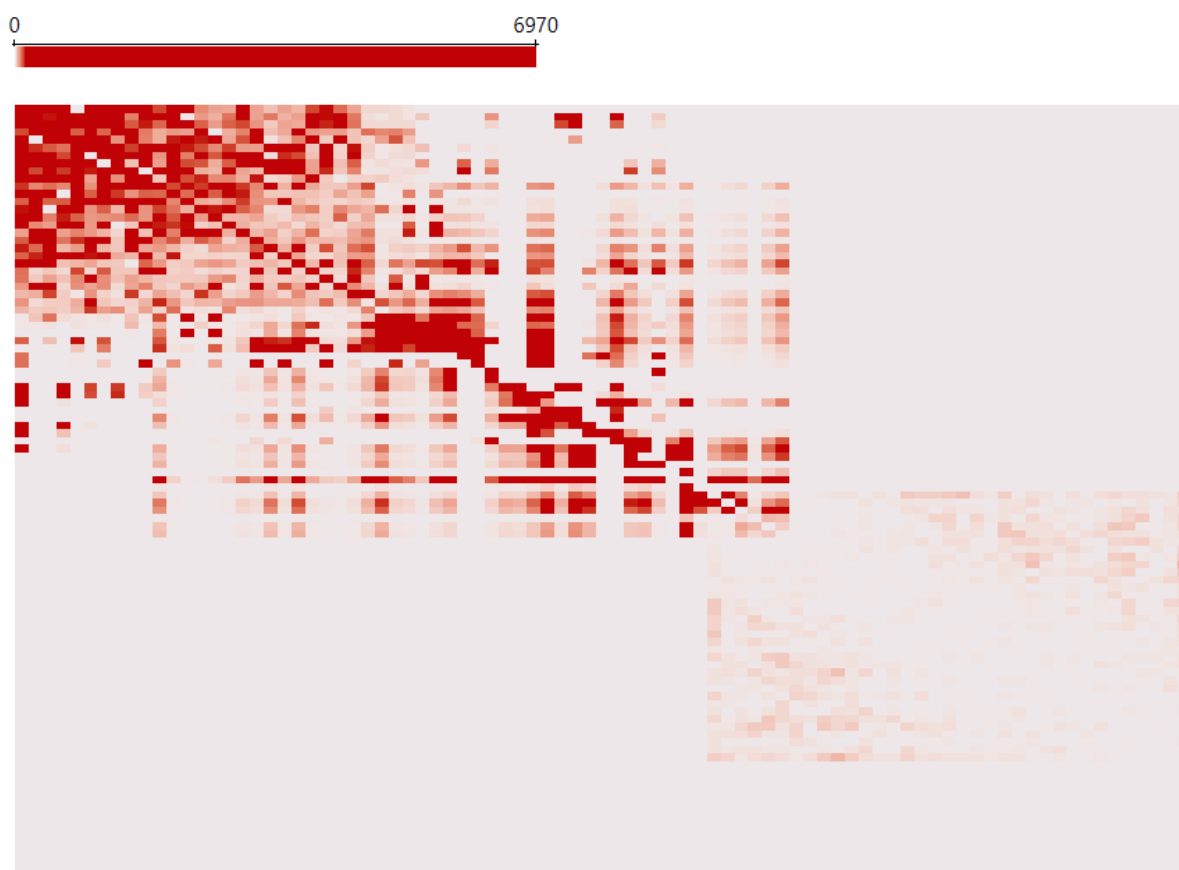
Rys. 6.23 Wariant 5, widok na obszar centrum aglomeracji – wartości procentowe

Rysunek 6.23 pokazuje wartości procentowe, które przekraczają 80-90% niedoboru podróży, co jeszcze lepiej oddaje niedokładność wariantu w porównaniu do rysunków 6.21 i 6.22. Na kilku połączeniach widać nawet brak podróży – jest 100% niedoboru. Pokazuje to, że dane cząstkowe same nie dadzą satysfakcjonującego rezultatu, jednak nie można wykluczać ich ograniczonej przydatności na poziomie macierzowym. Mogą również dać rezultaty gdy zaobserwuje się wśród nich tendencje charakterystyczne dla rejonów, które uwzględniły w większym zakresie, lub przy analizie ad hoc ukierunkowanej na charakterystykę konkretnego ośrodka miejskiego.



Rys. 6.24 Wariant 5, widok na obszar całej aglomeracji – wartości procentowe

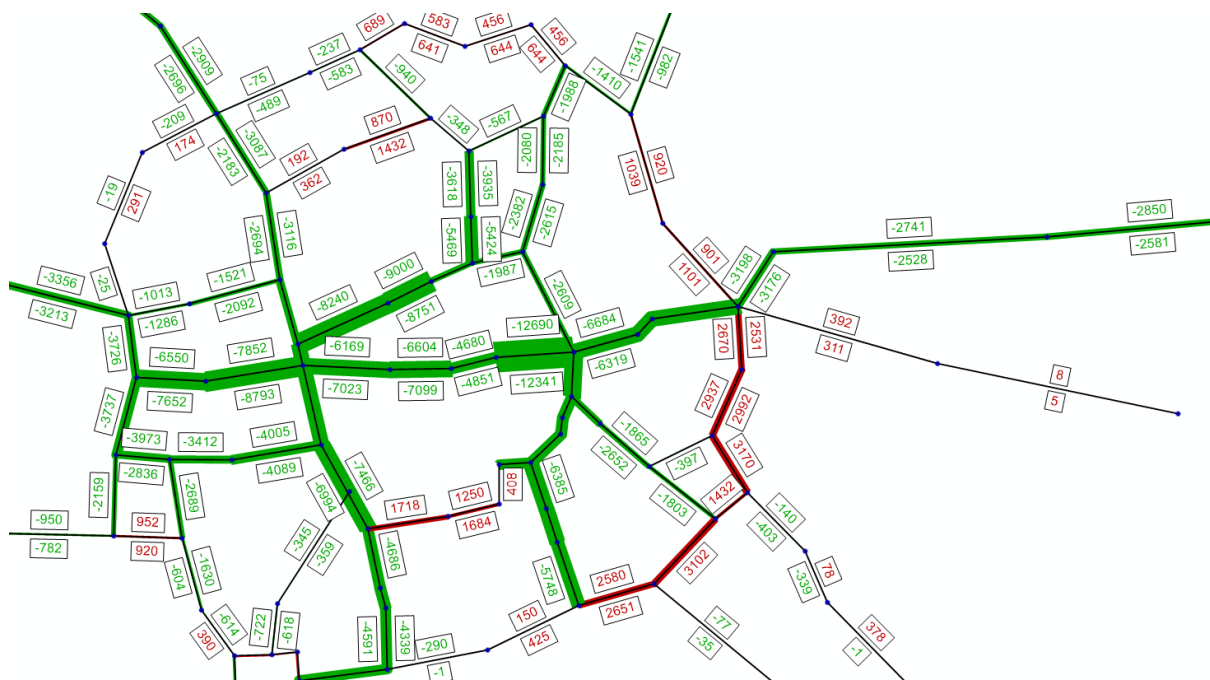
Rysunek 6.24 pokazuje sytuację którą można opisać podobnie jak rysunek 6.23. Jak wspomniano przy opisie rysunku 6.22, jedna ze wschodnich odnóg aglomeracji ma 100% niedoboru podróży. Jest to odnoga o najmniejszym natężeniu ruchu, a pozostałe odnogi procentowo przekraczają w większości przypadków 90% niedoborów. Wariant 5 dał najgorsze rezultaty. Na rysunku 6.25 pokazano macierz podróży w formie mapy ciepła, na której widać źródło niedoskonałości. Są to tylko obszary macierzy w których są podróże pochodzące z wybranych w tym wariantcie źródeł. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 8.



Rys. 6.25 Wariant 5, mapa ciepła macierzy podróży

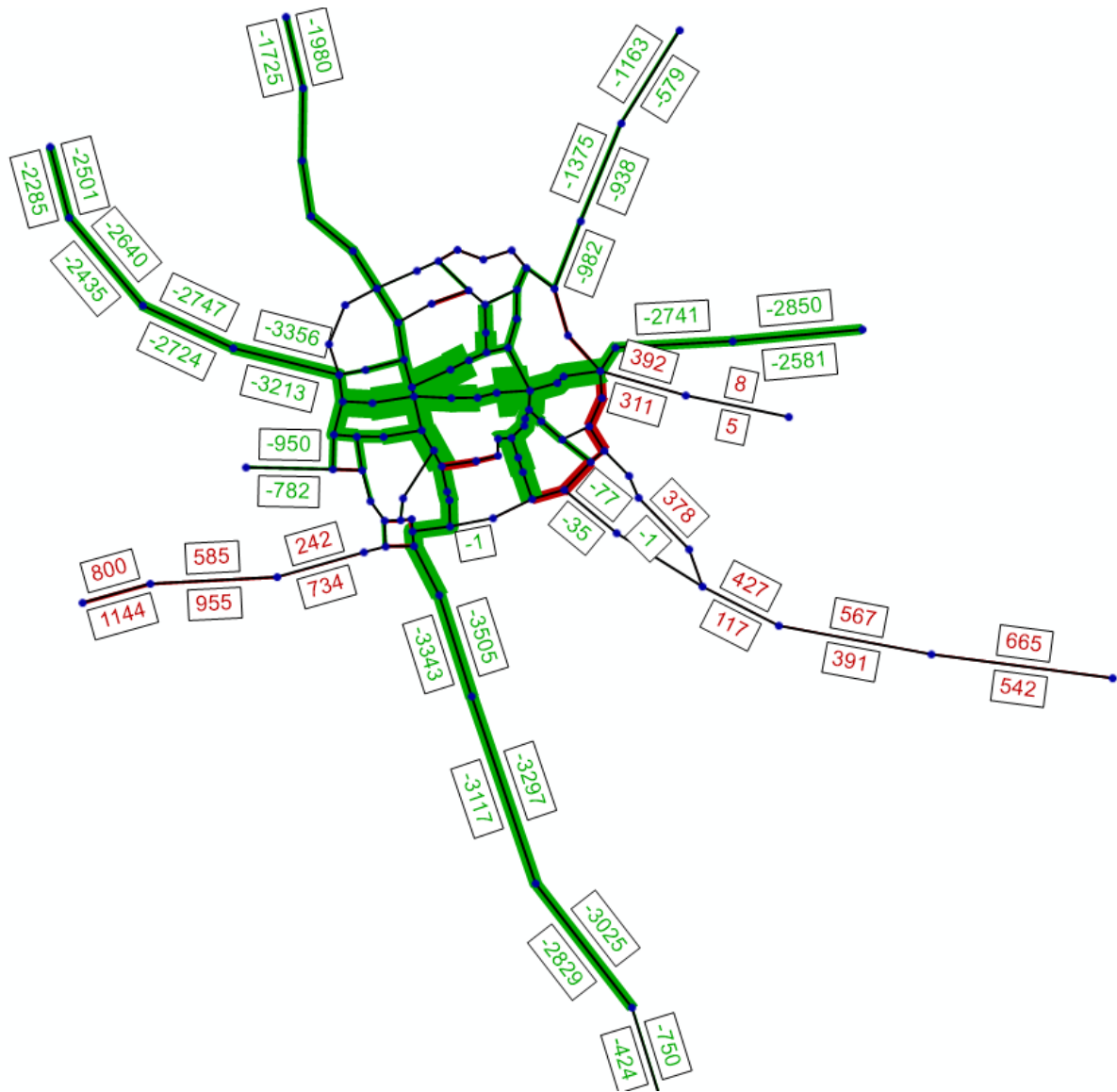
6.2.6. Wariant 6 – uzupełnienie danymi z mediów społecznościowych

Warianty 6, 7 i 8 sprawdzają które z symulowanych źródeł danych dają najlepsze rezultaty, które pozwolą wytypować najlepsze trzy źródła i zestawić je w wariancie 9. Jej cel to sprawdzenie czy teza o większej liczbie źródeł ma sens i czy selekcja najlepszych źródeł pozwoli uzyskać lepsze rezultaty. Wariant 4 wykorzystywał sam GSM, który jest jednym ze źródeł dającym dobre rezultaty. Niniejszy wariant uwzględnia jedynie dane z mediów społecznościowych. Ze względu, że jest to macierz pełna, posiadająca również rejony zewnętrzne, spodziewane wyniki będą najlepsze ze źródeł symulowanych na bazie map ciepła i charakterystyki danych pozyskanych z literatury.



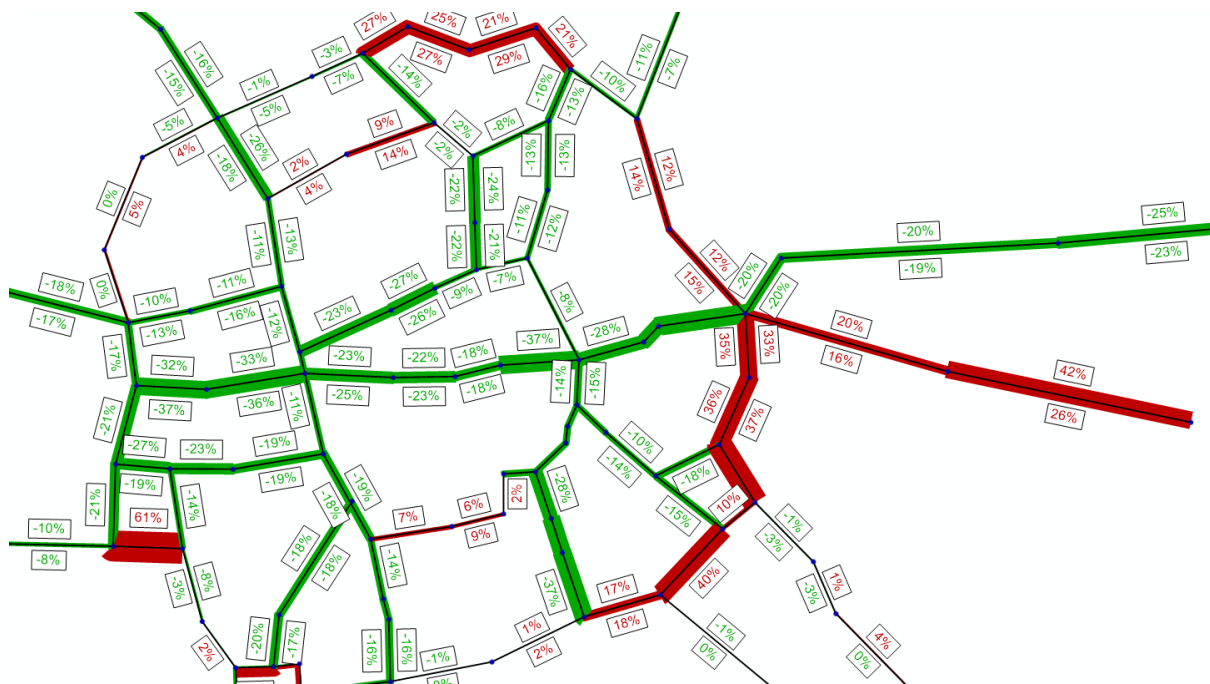
Rys. 6.26 Wariant 6, widok na obszar centrum aglomeracji

Rysunek 6.26 pokazuje całkiem dobry rezultat. Pomimo mocnych niedoborów na trasie W-Z oraz w południowych połączeniach obszaru centralnego, reszta połączeń (szczególnie obwodnica śródmiejska) ma dość zbilansowane wartości. Porównując do wariantu 1, uwzględniającego całą metodę, wyniki na obwodnicy śródmiejskiej prezentują bliższe wartości do modelu bazowego. Nadmiary są niższe, choć jest ich mniej. Na minus wypada obszar centralny, który patrząc na największe połączenie na trasie W-Z przedstawia się lepiej, lecz reszta trasy ma większe niedobory niż w wariantcie 1, na zachodniej części o około 1000 podróży na całej tej części trasy, po stronie wschodniej nawet do 3000 podróży. W dodatku połączenia obok mają zdecydowanie większe niedobory niż w wariantcie 1. Wyjaśnia to dlaczego sumaryczna liczba podróży w tym przypadku zdecydowanie odbiega od pełnej metody. Dodatkowo jest znacznie więcej połączeń z niedoborami niż połączeń z nadmiarami podróży, co jeszcze bardziej pogłębia różnicę.

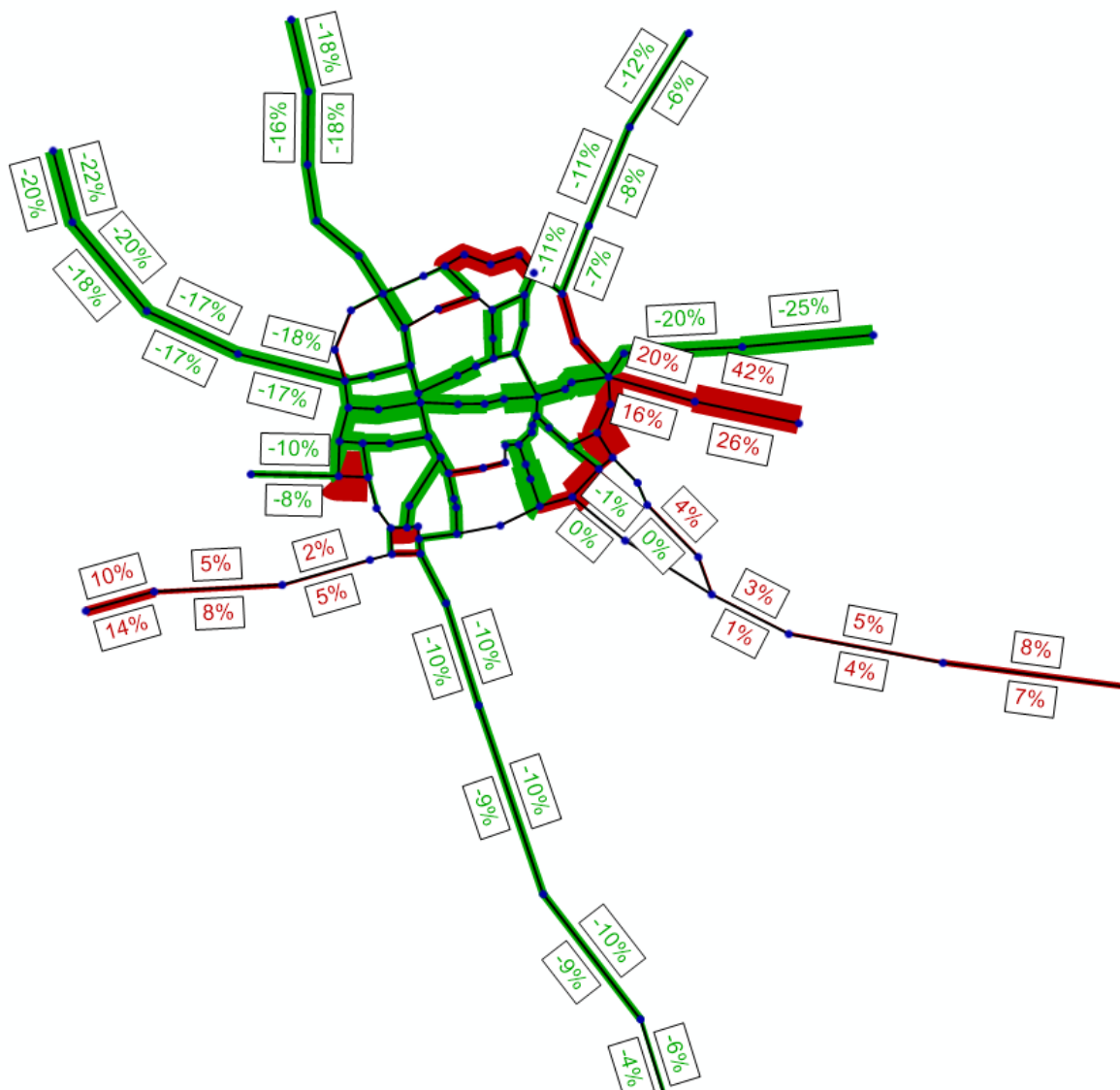


Rys. 6.27 Wariant 6, widok na obszar całej aglomeracji

Rysunek 6.27 jest jeszcze ciekawszy. Odnogi aglomeracji, pomimo znacznie większych niedoborów, „radzą” sobie lepiej niż w pełnej metodzie. Pokazuje to, że jest to źródło najlepiej szacujące drogi zewnętrzne. Niemniej, ze względu na charakter symulacyjny doktoratu, nie jest to wniosek ogólny dotyczący tego źródła Big Data, a tyczy się konkretnego przypadku z użytych źródeł w metodzie i tę hipotezę należałoby sprawdzić na rzeczywistych danych w różnych ośrodkach miejskich. W porównaniu do pełnej metody, odnoga południowa prezentuje wartości znacznie bliższe w stosunku do modelu bazowego, w wariantcie 1 wartości potrafiły miejscowo przekraczać 10000 nadmiarowych podróży, w tym przypadku żadne z dalszych połączeń nie przekracza niedoboru powyżej 4000 podróży. W przypadku odnóg wschodnich, dwie dolne nie przekraczają nawet 1000 podróży niedoboru, jednak górna prezentuje większe



W wartościach procentowych wnioski są takie same jak przy opisie rysunku 6.26. W porównaniu do pełnej metody, która dała podobny rezultat, trasa W-Z prezentuje sumarycznie większe wartości procentowe rozbieżności. Obwodnica śródmiejska prezentuje mniejsze wartości rozbieżności nie przekraczają one 50%, jednak nie na każdym połączeniu te wartości są niższe od wariantu 1. W wartościach procentowych widać również większe wartości odchyleń na połączeniach w obszarze centralnym, szczególnie na osi północ-południe, choć występują tam pojedyncze połączenia mające bliższe, lub znacznie bliższe wartości procentowe w porównaniu do modelu bazowego. Wartości nadmiarów podróży w tym wariantcie spadły do maksymalnie kilkunastu procent.



Rys. 6.29 Wariant 6, widok na obszar całej aglomeracji – wartości procentowe

Wartości procentowe na wschodnich odnogach aglomeracji mają mniejsze wartości niż w wariantie 1. Zwracając, jednak uwagę na wartości liczbowe oraz procentowe jest tak tylko w przypadku odnogi dolnej, gdzie spadły z kilkunastu-kilkudziesięciu procent do wartości jednocyfrowej, mniejsze wartości procentowe są również na środkowej odnodze, ale na połączeniu o minimalnej liczbie podróży (19 w obie strony). Na połączeniu gdzie podróży jest więcej, wartości procentowe w wariantie 1 mają jednocyfrowe wartości. Ogólnie rezultat wariantu 6 na odnogach można określić jako nieznacznie bliższy w stosunku do modelu bazowego. W przypadku pełnej metody jak i tego wariantu, na paru połączeniach występują wartości jednocyfrowe różnic procentowych, na korzyść wariantu 6. Na rysunku 6.30 pokazano

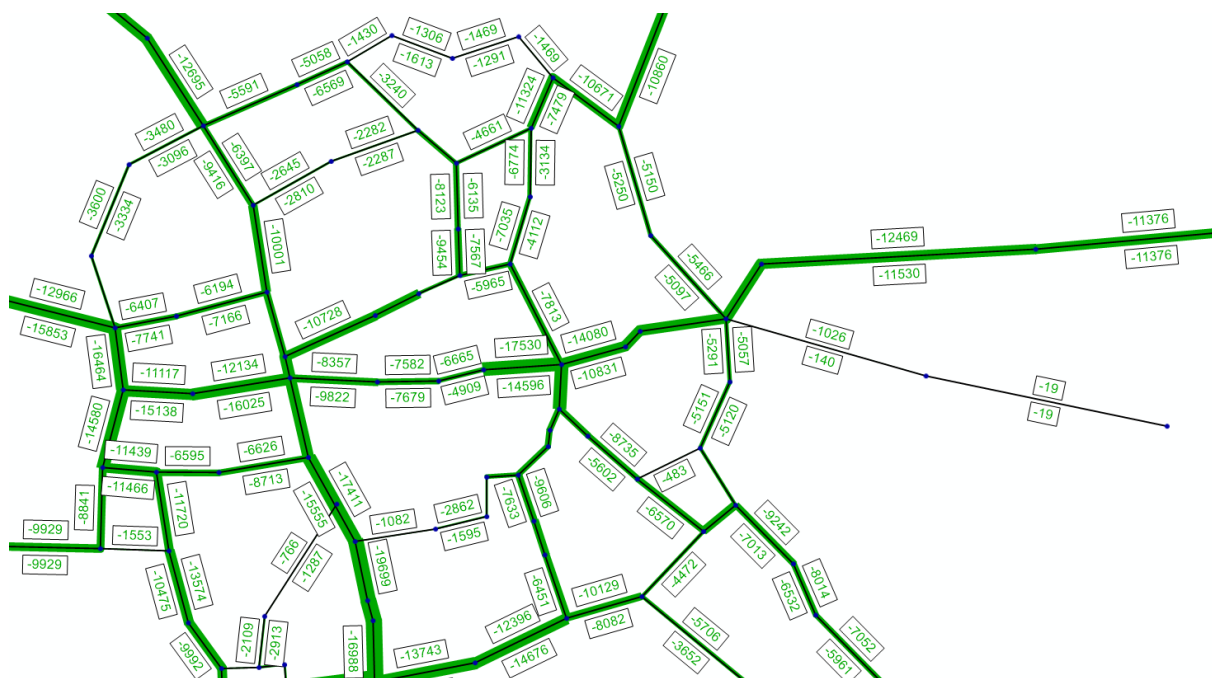
mapę ciepłą, w której również widać podobieństwo do modelu bazowego. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 9.



Rys. 6.30 Wariant 6, mapa ciepła macierzy podróży

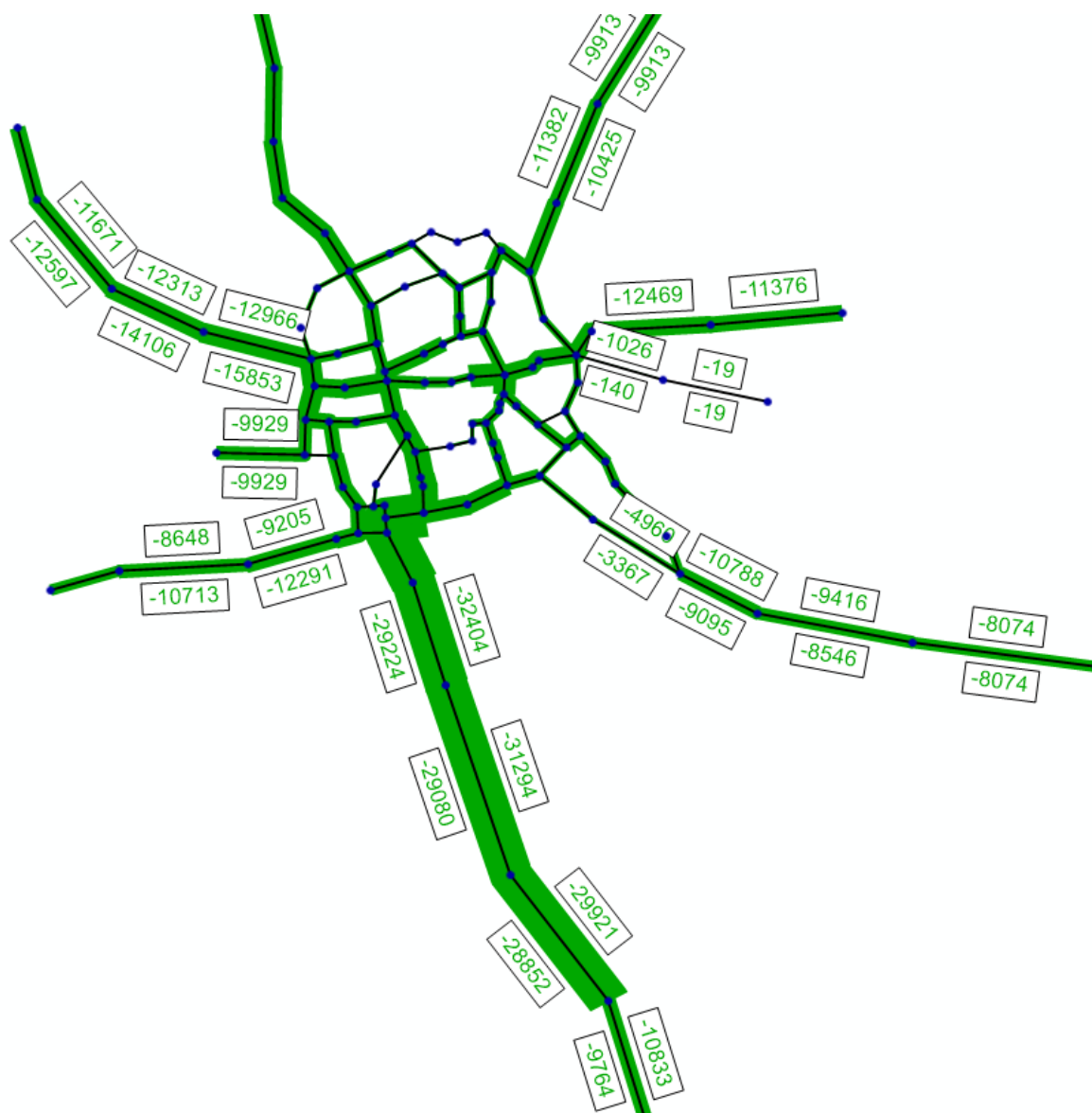
6.2.7. Wariant 7 – uzupełnienie danymi z Taxi

Wariant 7 podobnie jak wariant 5 dał niesatysfakcjonujący rezultat złożony z samych niedoborów ruchu. Jednak przyglądając się rysunkom od 6.31 do 6.34, widać, że spora część połączeń ma wartości zbliżone do modelu bazowego. Na odnogach aglomeracji są największe rozbieżności, jednak ta baza danych nie zawierała podróży poza aglomerację, co znacząco zredukowało liczbę podróży, ponieważ nie ma tutaj podróży tranzytowych oraz źródłowych i docelowych.



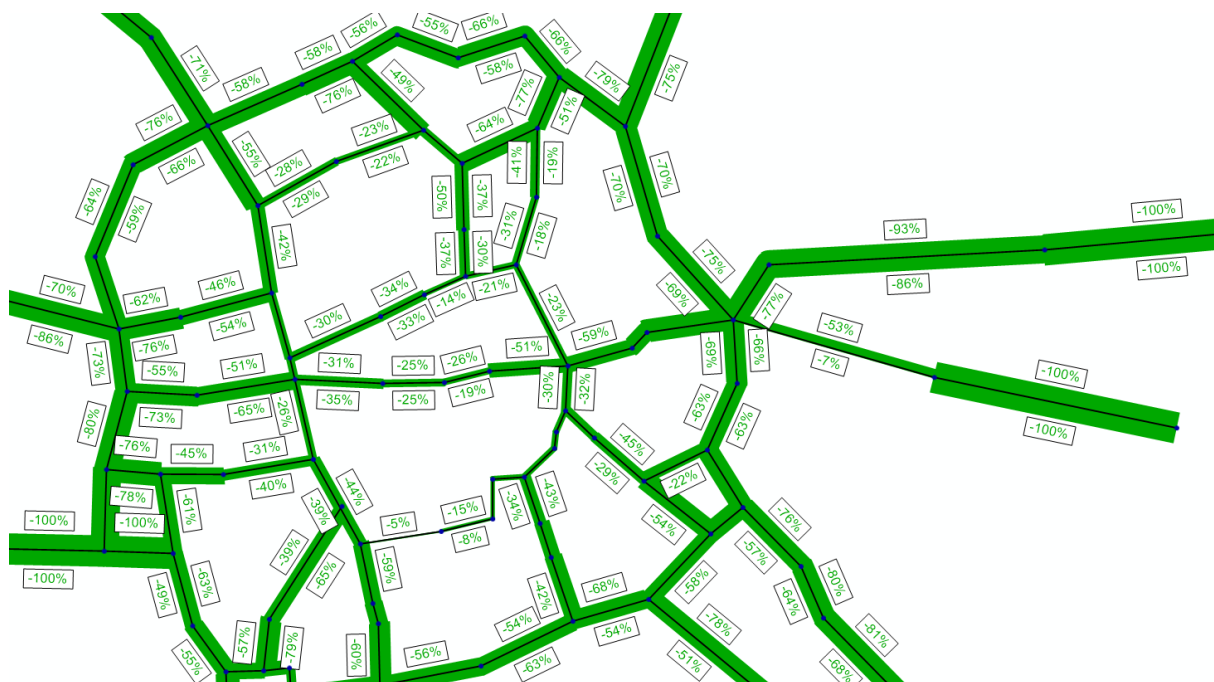
Rys. 6.31 Wariant 7, widok na obszar centrum aglomeracji

Ze względu na brak podróży tranzytowych na modelu widać, że najmniejsze niedobory podróży występują wewnątrz aglomeracji, na połączeniach oddalonych od odnóg aglomeracji. Największe liczby niedoborów sięgające ponad kilkanaście tysięcy podróży, podobnie jak w wariancie 6, widać na trasie W-Z oraz na połączeniach północ-południe, szczególnie blisko odnogi południowej. Pomimo samych niedoborów podróży na połączeniach, wynik ten jest znacząco bliższy modelowi bazowemu w porównaniu do wariantu 5. Wartości liczbowe są znacząco mniejsze, nawet ponad trzykrotnie. Połączenia, które wykorzystują tylko ruch miejscowy mają małe wartości niedoborów ruchu.



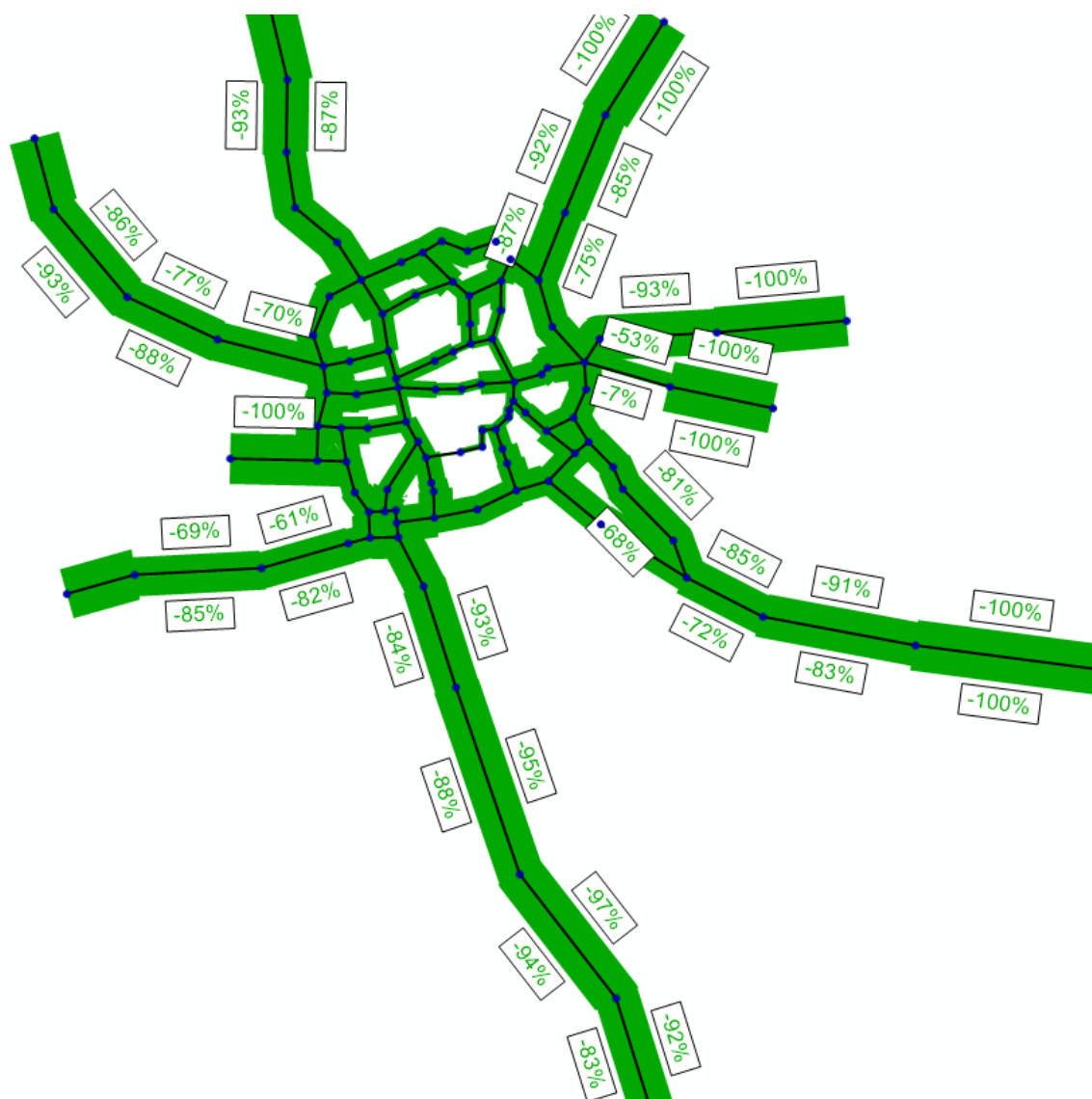
Rys. 6.32 Wariant 7, widok na obszar całej aglomeracji

Na odnogach aglomeracji, widocznych na rysunku 6.32 wartości są mniejsze niż w wariancie 5, ale znacznie wyższe niż w wariancie 1. Ze względu na bardziej bliskie modelowi bazowemu wyniki w podróżach miejscowych, wartości na połączeniach bliżej obszaru centralnego aglomeracji są korzystniejsze niż w dalszych rejonach odnog aglomeracji. Ogólnie wyniki tego wariantu są niezadowolające.



Rys. 6.33 Wariant 7, widok na obszar centrum aglomeracji – wartości procentowe

Wartości procentowe na rysunku 6.33 dają lepszy pogląd na wyniki na połączeniach. Nie wszystkie połączenia wewnętrzne pomiędzy rejonami obszaru centralnego aglomeracji pokazują wyniki bliskie modelowi bazowemu. Najbliższy modelowi bazowemu rezultat dało ściśle centrum aglomeracji, im dalsze połączenia, tym coraz większe rozbieżności. Obwodnica śródmiejska ma ponad 50% niedoborów podróży, czego nie było wyraźnie widać po samych wartościach na rysunku 6.31. Ring wokół obszaru ścisłego centrum miejscami na wartości nawet poniżej 10%.

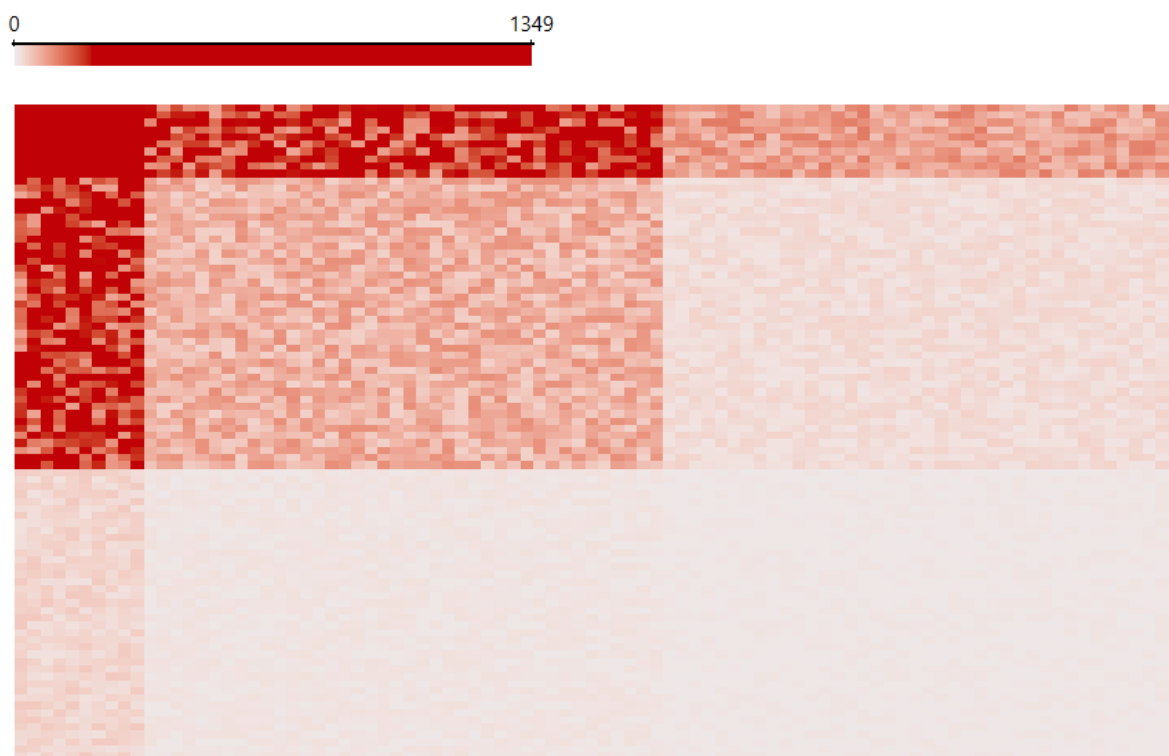


Rys. 6.34 Wariant 7, widok na obszar całej aglomeracji – wartości procentowe

Podobną tendencję jak na rysunku 6.33 widać również na rysunku 6.34. Im dalej od centrum aglomeracji tym bardziej rozbieżne wyniki, co można zaobserwować za pomocą grubości „wykresów” na połączeniach. Wyjątkiem jest tylko odnoga wschodnia środkowa, której relacja na zewnątrz aglomeracji ma dobry wynik, odbiegający od modelu bazowego o zaledwie 7% podróży. Na połączeniu dalej od centrum występuje brak podróży, podobnie jest na innych odnogach, wynika to z braku podróży poza aglomeracyjnych, więc wynik był spodziewany.

Takie wyniki biorą się z dwóch charakterystyk tej symulowanej bazy danych, ponieważ uwzględniane nie zostały podróże tranzytowe oraz źródłowe i docelowe, a same miejscowe w aglomeracji. Wpłynęło to szczególnie na odnogi aglomeracji oraz obwodnicę śródmiejską, ale również na obszar centralny, w którym wyniki były miejscami akceptowalne. Drugi czynnik to

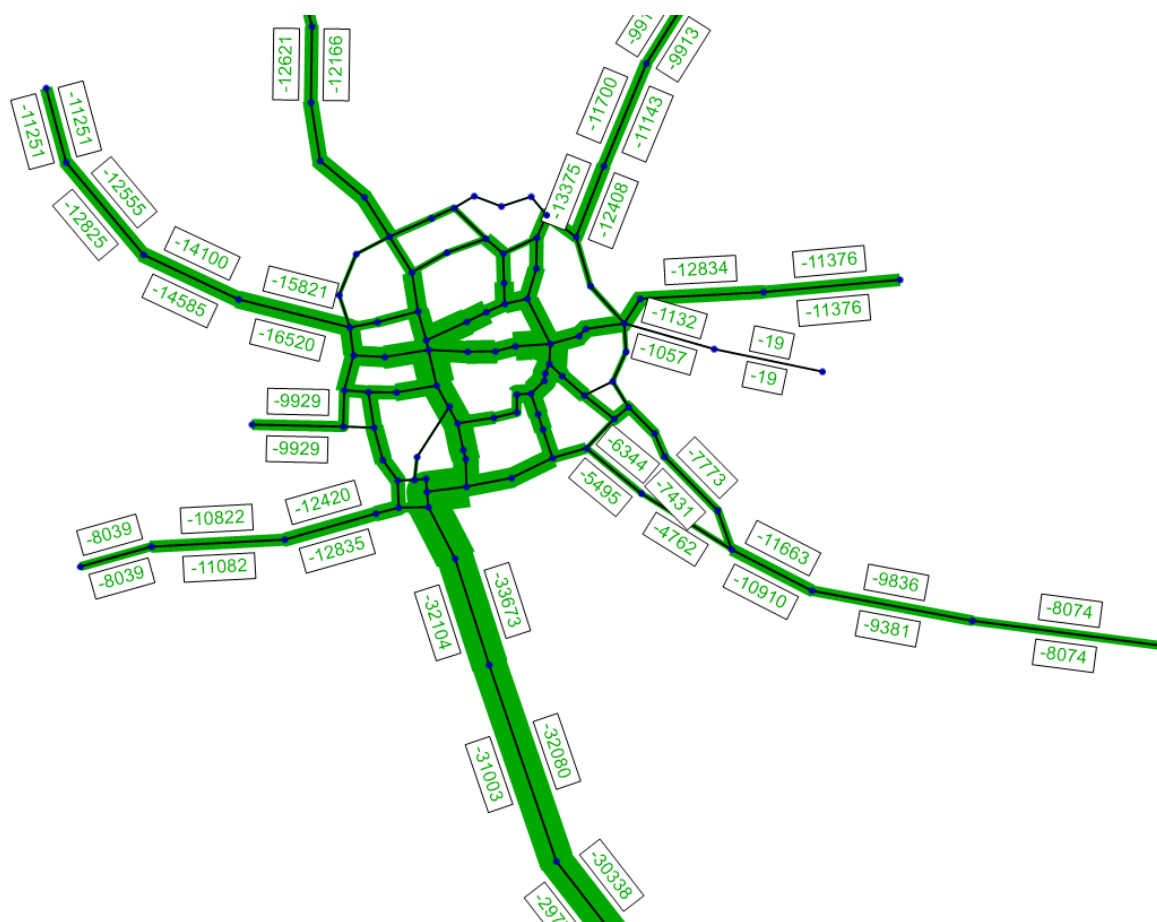
symulowanie tej bazy danych na bazie map ciepła i opisu. W tym przypadku więcej podróży dotyczy rejonów centralnych. Mapę ciepła przedstawiono na rysunku 6.35. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 10.



Rys. 6.35 Wariant 7, mapa ciepła macierzy podróży

6.2.8. Wariant 8 – dane z rowerów miejskich i Urzędzeń Transportu Osobistego

Wariant 8 to pojedyncze źródło symulowane jak w wariancie 7. W tym przypadku założeniem podczas tworzenia macierzy podróży był brak podróży poza obszar aglomeracji, nadmiar symulowanych podróży pomiędzy rejonami położonymi blisko siebie i niedomiar podróży pomiędzy rejonami położonymi dalej. Na wyniki modelu mogło wpłynąć założenie sieci transportowej jako samych głównych połączeń aglomeracyjnych pomiędzy rejonami transportowymi i brak dróg osiedlowych w modelu.



Rys. 6.37 Wariant 8, widok na obszar całej aglomeracji

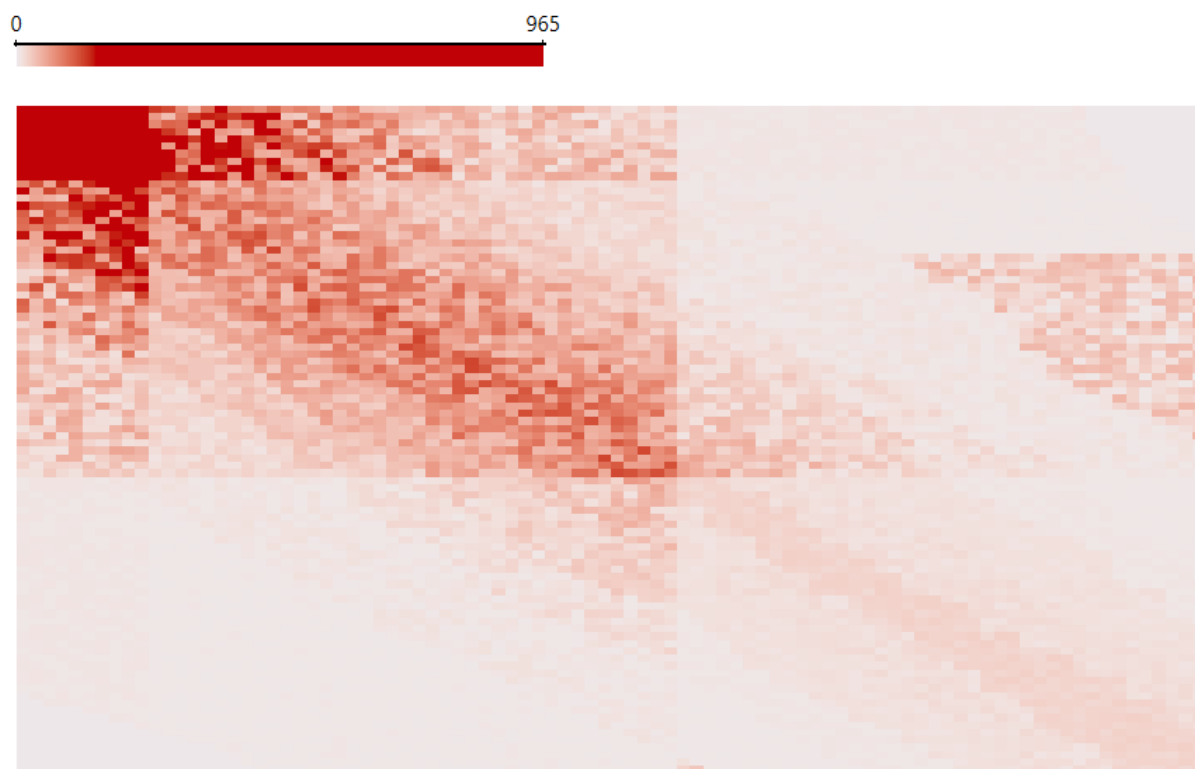
Porównując rysunek 6.37. z wariantem 5, rezultaty są zbliżone - nieznacznie większe rozbieżności w połączeniach w rejonach dalekich od centrum aglomeracji i nieznacznie mniejsze w większości połączeń bliskich centralnemu obszarowi aglomeracji. Ogółem wszystkie odnogi dają wynik niesatysfakcjonujący. Było to spodziewane, ponieważ w tym wariantcie również nie ma podróży związanych z rejonami zewnętrznymi (tranzytowych, źródłowych i docelowych).



Rys. 6.38 Wariant 8, widok na obszar centrum aglomeracji – wartości procentowe

Wartości procentowe widoczne na rysunku 6.38. mocno odbiegają od modelu bazowego. Nawet w obszarze centralnym aglomeracji potrafią wynosić około 90%, tylko na niektórych połączeniach niedobory są niższe 50%. Rezultat nie jest satysfakcjonujący, jednak rozbieżności są niższe niż w najgorszym wariancie z zestawienia, czyli wariancie 5, gdzie niedobory podróży przekraczały 80% na każdym z połączeń. Porównując do wariantu 7, wartości niedoborów są większe, tam niedobory podróży w większości wynosiły poniżej 50% w obszarze centralnym aglomeracji. Wartości procentowe pokazują również, że w wariancie 8 nie ma znacząco większych niedoborów na konkretnych typach połączeń, są jednak rozłożone równomiernie na połączeniach w centralnym obszarze aglomeracji.

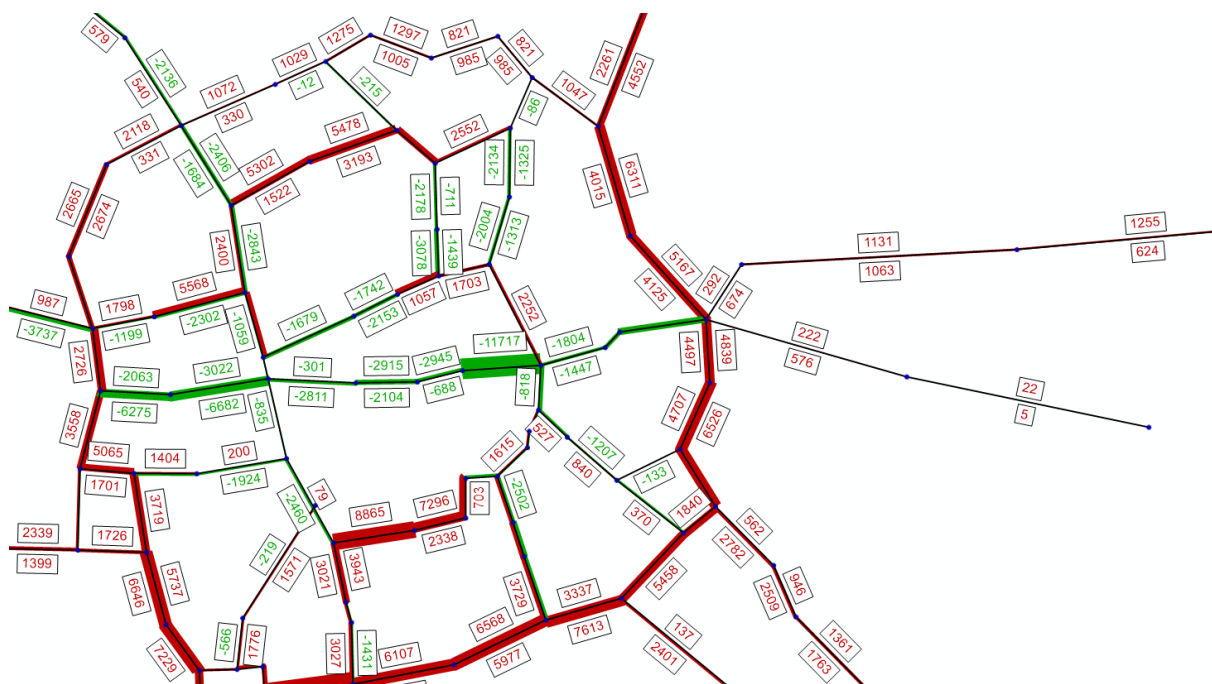
rowerów miejskich i UTO. Na rysunku 6.40 przedstawiono mapę ciepłą. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 11.



Rys. 6.40 Wariant 8, mapa ciepła macierzy podróży

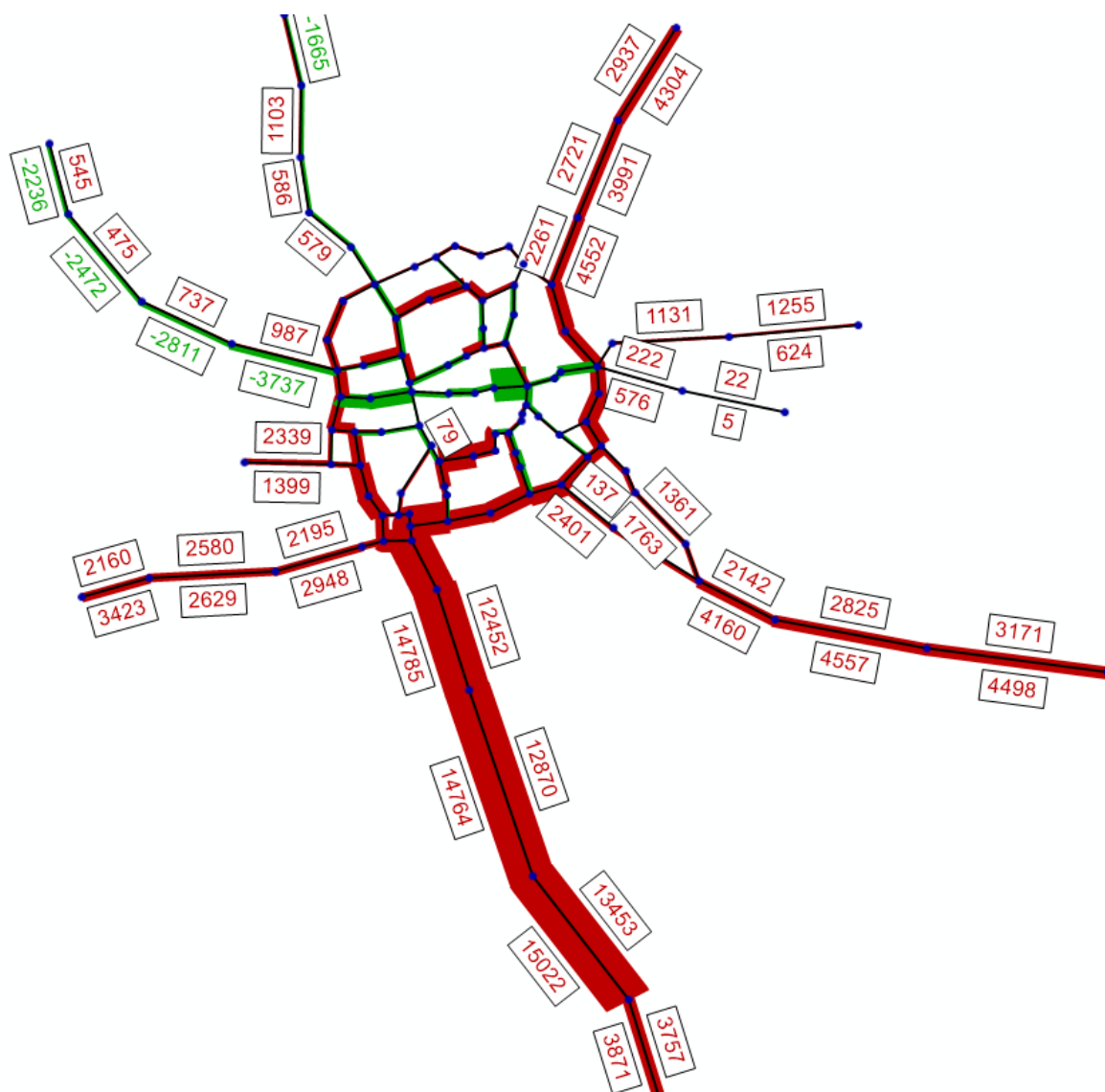
6.2.9. Wariant 9 – Dane z GSM, mediów społecznościowych i Taxi

Wariant 9 został stworzony jako ostatni. Jest to próba stworzenia „konkurencji” dla wariantu 1, który uwzględniał wszystkie źródła, więc pełną metodę. Celem tego wariantu jest sprawdzenie - czy lepsze jest uwzględnianie wielu źródeł, czy lepiej jest wyselekcjonować tylko te najlepsze. W wariantcie 9 uwzględniono najlepsze rezultaty wariantów zawierających jedno źródło danych. Nie sprawdzono wszystkich źródeł, ponieważ z arkusza kalkulacyjnego wynikało, że nie dadzą lepszych rezultatów, niż sprawdzone cztery. Ostatecznie wybrano trzy źródła na bazie wniosków z analizy wariantów 4, 6 i 7, czyli danych z GSM, mediów społecznościowych i Taxi.



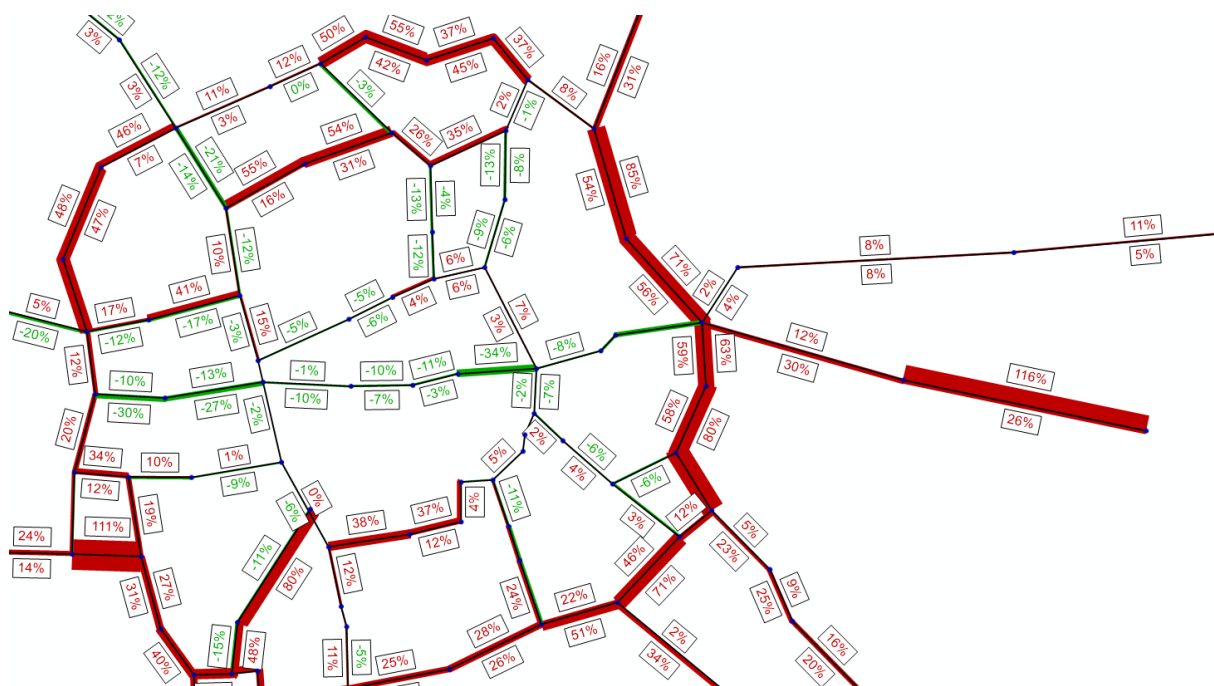
Rys. 6.41 Wariant 9, widok na obszar centrum aglomeracji

Rysunek 6.41. będzie porównywany z wariantem 1, czyli odpowiadającym mu rysunku 6.1. Przy pierwszej obserwacji widoczny jest większościowy udział połączeń z nadmiarową liczbą podróży. Na południowej części obwodnicy śródmiejskiej liczby podróży są wyższe - nawet około dwukrotnie na niektórych połączeniach. Na części wschodniej liczby są mocno do siebie zbliżone - niektóre połączenia w wariantie 9 dają wyniki bliższe modelowi bazowemu, niektóre dalsze. Podobnie jest na części północnej i zachodniej. W przypadku trasy W-Z, wariant 9 dał znacząco wyniki bliższe o około kilka tysięcy na większości połączeń. Może to być spowodowane ogólnym nadmiarem podróży w modelu, w wariantie 1 było to połączenie z największymi niedoborami ruchu. Nie wszystkie połączenia z niedoborami ruchu mają wartości bliższe modelowi bazowemu. W przypadku tego modelu sieci, na rysunku 6.1 można znaleźć połączenia z mniejszymi niedoborami ruchu oraz analogicznie z mniejszymi nadmiarami. Ogółem sieć obszaru centralnego w wariantie 1 daje bliższe wyniki z wyjątkiem trasy W-Z.



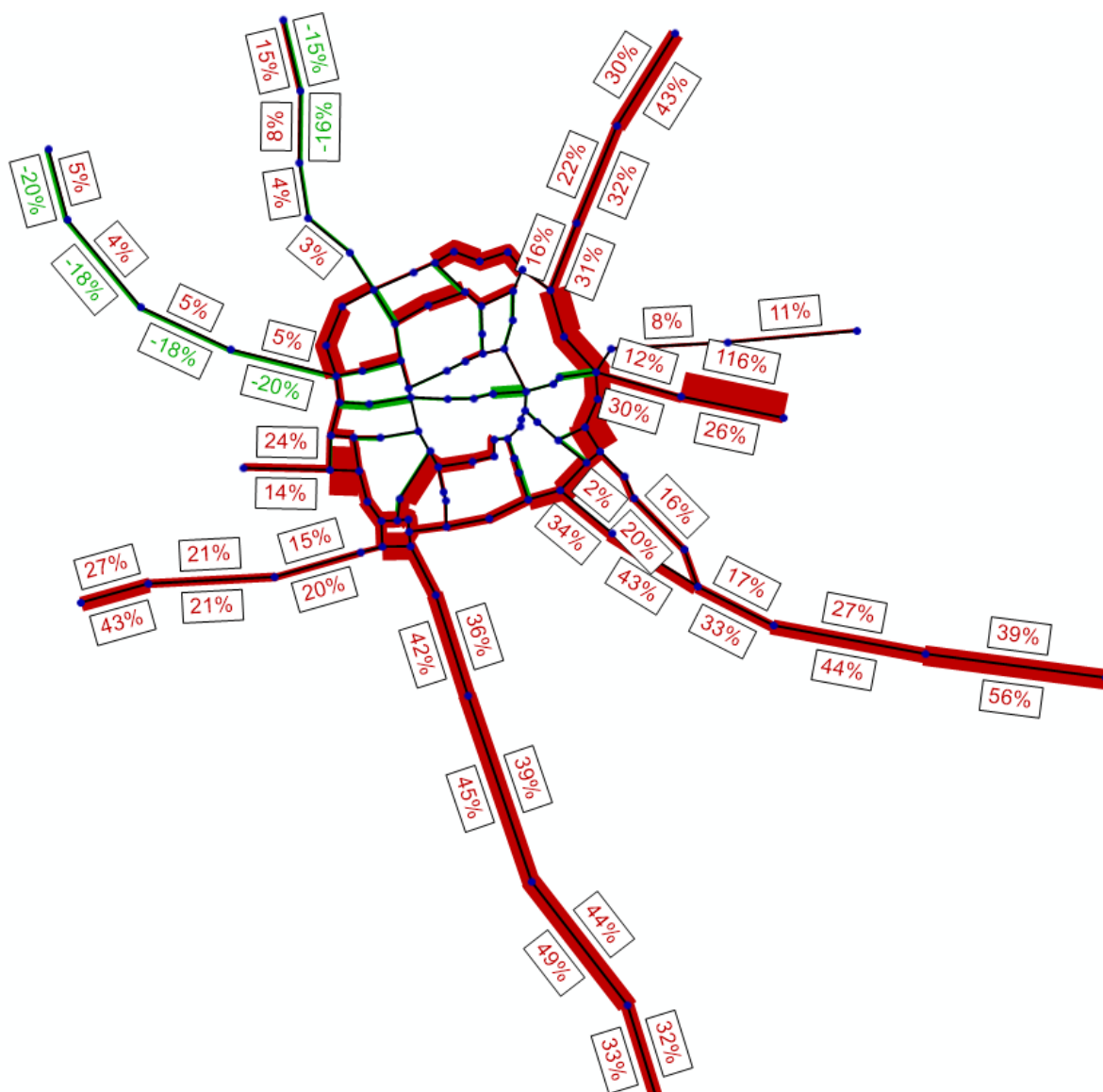
Rys. 6.42 Wariant 9, widok na obszar całej aglomeracji

Na odnogach aglomeracji, widocznych na rysunku 6.42, w większości przypadków nadmiary podróży są wyższe niż w wariantie 1. Lepiej wypada tylko odnoga zachodnia górna, co mogło wpłynąć na ogólną liczbę podróży. Jest to jedyny z wariantów wykazujący nadmiar podróży, choć różnica wynosi tylko 4,80% i po wariantie 1 daje drugi najbliższy rezultat w porównaniu do modelu bazowego. Różnice wynoszą od kilkuset do ponad trzech tysięcy na odnodze południowej. Rezultat jest ogólnie satysfakcjonujący, lecz nie daje tak dobrych wyników jak w przypadku wykorzystania pełnej metody.



Rys. 6.43 Wariant 9, widok na obszar centrum aglomeracji – wartości procentowe

Na rysunku 6.43 są pokazane wartości procentowe w obszarze centralnym. W porównaniu do rysunku 6.3, największe różnice są na północnej części obwodnicy śródmiejskiej, gdzie wartości spadły do około 50% z 78% na jednym połączeniu w wariantcie 1. W większości przypadków jest to maksymalnie kilkanaście procent różnicy w różne strony. Ze względu na dużą liczbę połączeń na trasie W-Z, zmiany które przy analizie rysunku 6.41 wydawały się znaczne, w tym przypadku nie wyróżniają się tak mocno jak wyżej wspomniana część północnej części obwodnicy śródmiejskiej.



Rys. 6.44 Wariant 9, widok na obszar całej aglomeracji – wartości procentowe

Rysunek 6.44 ze względu na dużą liczbę podróży na odnogach aglomeracji ma mniejsze różnice procentowe w wartościach procentowych niż mógłby sugerować rysunek 6.42. W porównaniu do wariantu 1 wyraźnie widać, że wariant 9 ma bardziej rozbieżne od modelu bazowego wyniki pomimo ich ogólnego podobieństwa w liczbach.

Na rysunku 4.45 przedstawiono macierz podróży w postaci mapy ciepła. Widać podobieństwo do modelu bazowego, jednak występują tam lokalne wartości ekstremalne, które są naleciałościami z adekwatnymi anomaliami z macierzy podróży pozyskanej z GSM.

Wyniki wariantu 9 dały rezultat zbliżony do pełnej metody. Pełna metoda opisana jako wariant 1, dała wyniki bliższe, pomimo tego, że uwzględniała również „słabe” źródła danych. Zdaje się to potwierdzać tezę o większej liczbie źródeł danych, nawet częściowych, które zwiększą dokładność modeli transportowych. Macierz podróży w formacie liczbowym załączono do pracy w Zał. 12.



Rys. 6.45 Wariant 9, mapa ciepła macierzy podróży

6.3. Analiza rezultatów i skuteczności uzupełniania danych

Na początku zestawiono wartości sumy podróży uzyskanych w modelu bazowym do każdego wariantu i zaprezentowano różnice. Symulowana macierz dała 659890 podróży. Najlepszy wariant zawierał wszystkie źródła, czyli sumarycznie 645242 podróże, co daje różnicę z modelem wzorcowym na poziomie 2,22%. Jest to najlepszy ze wszystkich wariantów, co wydaje się potwierdzać tezę „Wartościowe jest łączenie wielu źródeł Big Data, ponieważ pozwala uzyskać efekt synergiczny w uzupełnianiu brakujących lub eliminacji nieprawidłowych danych.” Drugi wariant, który dał zbliżone wartości, to wariant łączący źródła dające najlepsze wyniki. Jest to ostatni z dziewięciu wariantów - połączenie danych z GSM, mediów społecznościowych i Taxi. Pierwszy wariant to źródło dające dobre dane ze

względem na możliwość dostosowania do symulowanego ośrodka opisanego w Rozdziale 5. Dane z mediów społecznościowych i Taxi są źródłami symulowanymi na podstawie map ciepła, proces symulowania opisano w podrozdziale 5.3. Te same dane zestawiono jako osobny wariant oraz jako pojedyncze źródła. W stosunku do wariantu łączącego te trzy źródła danych dawały one większą rozbieżność. GSM dawał 25,34% różnicy względem modelu bazowego, media społecznościowe 14,27%, a Taxi 31,18%. Po połączeniu metodą zaproponowaną w rozprawie doktorskiej te trzy źródła dały tylko 4,8% rozbieżności z modelem bazowym. Jest to zdecydowanie lepszy wynik niż np. zestawienie połączenia źródeł pochodzących z symulowania baz danych na podstawie „map ciepła” (18,87% rozbieżności). Wariant łączący dane z wrocławskiego KBR dał duże rozbieżności ze względu na słabe dane GPS pochodzące z sondowania pojazdów. Wiele rejonów dawało bardzo małe dane, a ze względu na niewielką próbę niewielkie wartości wielokrotnie skalowano do wielkości ośrodka miejskiego. Była to macierz zawierająca dużą liczbę relacji pomiędzy rejonami bez danych - jest tam pustych aż 9438 na 10000 relacji. Dlatego zdecydowano się na uwzględnienie tylko samego GSM w następnym wariancie, ze względu na lepsze wyniki. Wariant BT+APC+SC+ANPR obejmował dane głównie cząstkowe, stąd spodziewany słaby wynik o dużej rozbieżności z modelem bazowym (55,91%). Gorszy wynik dał pojedynczy wariant symulowanych na podstawie charakterystyki danych i „map ciepła” danych z UTO i rowerów miejskich, gdzie rozbieżność wynosiła aż 63,64% z modelem bazowym. Pomimo stosunkowo małej liczby par połączeń z danymi zerowymi – 2119 na 10000. Jest to spowodowane uwzględnieniem charakterystyki danych i podróży rowerami i UTO. Symulując macierz zakładano znaczący spadek podróży na dalekie dystanse.

Tab. 6.1 Zestawienie sumy podróży w poszczególnych wariantach

Nr.	Wariant	Liczba podróży	Różnica	Różnica %	Liczba zer
-	Model Bazowy	659890	0	0,00%	0
1	Wszystkie źródła	645242	14648	2,22%	0
2	Źródła symulowane	535385	124506	18,87%	0
3	Wrocławski KBR	488576	171315	25,96%	7066
4	Tylko GSM	492660	167230	25,34%	7405
5	BT APC SC ANPR	290971	368919	55,91%	7072
6	Tylko Social media	565712	94178	14,27%	0
7	Tylko Taxi	454128	205763	31,18%	1900
8	Tylko Rowery+UTO	239908	419982	63,64%	2119
9	GSM Media Społ. Taxi	691552	31661	4,80%	0

W celu pogłębionej analizy statystycznej zdecydowano się na zbadanie odchyłeń standardowych każdego z wariantów. Wyniki zestawiono z wariantami w tabeli 6.2. Zdecydowanie najlepsze wyniki o najniższych odchyleniach dały warianty zawierające tylko źródła symulowane. Najlepszy wynik dała macierz UTO i rowerów miejskich, która ma aż 63% mniej podróży niż model bazowy i w różnicy procentowej daje najgorszy rezultat ze wszystkich wariantów. Może mieć to dużo wspólnego z rejonami bez danych liczbowych oraz z tym, że są to dane uśrednione poprzez symulowanie podobne do metody Monte Carlo, opisane w rozdziale 5.1. Na korzyść metody zdecydowanie przemawia najlepszy wynik ze wszystkich zaproponowanych wariantów, które zawierają przynajmniej jedno ze źródeł, które nie jest symulowane. W porównaniu do wariantu 9 wynik jest podobny do analizy z tabeli 6.1, czyli dość zbliżony, jednak pełna metoda daje lepszy rezultat. Zaskakujące jest również odchylenie w wariancie 3, gdzie użyto danych z wrocławskiego KBR. Wariant 4 zawierał znacznie bogatszą w dane macierz z zestawienia dwóch z wariantu 3 i pomimo tego prezentuje nieznacznie większe odchylenie. Najbardziej wybrakowana macierz z wariantu 5, zestawiona próbnie spodziewając się słabych wyników, dała rezultat niewiele gorszy niż rezultat wariantu 9, mającego „konkurować” z wariantem 1 testującym metodę. Stwierdzono, że bardziej wiarygodną miarą, pokrywającą się z rysunkami z podrozdziału 6.2 będzie średnia z tabeli 6.1. Analizując tabelę 6.2. należy przyjąć, że źródła symulowane dały lepiej uśrednione wyniki, co widać po lepszych wartościach odchylenia standardowego. Niemniej są to wartości, które nie odstają mocno od odchylenia standardowego pełnej metody w wariancie 1.

Tab. 6.2 Zestawienie odchyłeń standardowych w poszczególnych wariantach

Nr.	Wariant	Liczba zer	Odch. Stand.
-	Model Bazowy	0	-
1	Wszystkie źródła	0	129
2	Źródła symulowane	0	95
3	Wrocławski KBR	7066	253
4	Tylko GSM	7405	262
5	BT APC SC ANPR	7072	144
6	Tylko Social media	0	92
7	Tylko Taxi	1900	108
8	Tylko Rowery+UTO	2119	90
9	GSM Media Społ. Taxi	0	140

6.4. Podsumowanie

Analiza wyników potwierdza hipotezy rozprawy „Wykorzystanie źródeł Big Data w modelowaniu podróży pozwala uzyskać dokładniejsze wyniki” oraz „Wartościowe jest łączenie

wielu źródeł Big Data, ponieważ pozwala uzyskać efekt synergiczny w uzupełnianiu brakujących lub eliminacji nieprawidłowych danych.”. Zarówno rezultaty statystyczne z podrozdziału 6.3., oraz ocena rysunków i liczb z podrozdziału 6.2. przedstawiają wariant będący wykorzystaniem pełnej metody jako najbardziej zbliżony do modelu bazowego z dziewięciu wytypowanych wariantów, które potencjalnie miały dawać najlepsze lub najbardziej charakterystyczne dane. Drugim najbliższym modelowi bazowemu był wariant zestawiający 3 źródła danych dające najlepsze wyniki. Pomimo zestawienia źródeł w sposób mający „konkurować” z pełnym wykorzystaniem zaproponowanej w niniejszej rozprawie metody, wyniki pełnej metody prezentowały wartości bliższe modelowi bazowemu. Tylko wartości odchyleń standardowych nie dawały najlepszego rezultatu w przypadku zastosowania pełnej metody, jednak w przypadku tego wyznacznika statystycznego wystąpiły anomalie opisane w podrozdziale 6.3. Należy również podkreślić, że niektóre zestawienia źródeł danych na pojedynczych połączeniach dawały lepsze rezultaty niż pełna metoda. Nie jest jednak możliwe poprawnie merytorycznie selekcionowanie wartości w celu uzyskania wariantu będącego kombinacją najlepszych wartości na poszczególnych połączeniach, ze względu na to, że są to składowe łączone połączeń z macierzy podróży.

Gdy pojawią się ośrodki miejskie, z których będzie się dało pozyskać zbliżoną liczbę źródeł big data, metoda zostanie ostatecznie przetestowana na danych rzeczywistych. Na bazie dostępnych danych i symulowanego charakteru ośrodka uzyskano jednak najbliższy rezultat do modelu bazowego, co potwierdza tezę rozprawy.

7. Dyskusja

7.1. Ogólny opis rezultatów

W rozprawie stworzono dwie metody do badania źródeł danych Big Data w modelowaniu transportu. Pierwsza metoda łączenia źródeł z dużych wolumenów i zróżnicowanego pochodzenia na poziomie liczby podróży pomyślnie posłużyła do stworzenia macierzy podróży z wykorzystaniem dziewięciu zróżnicowanych źródeł danych. Żaden z ośmiu wariantów „konkurencyjnych” nie prezentował wartości bliższych do modelu bazowego. Nawet ostatni, który na bazie doświadczeń z tworzenia wszystkich poprzednich był zestawieniem trzech baz danych dających najlepsze rezultaty i pomimo podobnego wyniku do wariantu wykorzystującego pełną metodę łączenia dziewięciu źródeł, wyniki były gorsze. Zarówno w wartościach różnicy sumy liczby podróży (2,22% do 4,80%), odchylen standardowych (129 do 140) i analizy rozłożenia podróży na ogólną sieć połączeń pomiędzy rejonami. Polem do poprawy tej metody jest tylko zmniejszenie rozbieżności na poszczególnych połączeniach na uproszczonej sieci transportowej symulowanego ośrodka miejskiego i jest to zaplanowane jako kierunek dalszych prac badawczych.

Druga metoda - symulacji ośrodka miejskiego pozwoliła stworzyć ośrodek miejski o racjonalnych parametrach dotyczących generacji i rozkładu przestrzennego podróży. W tym przypadku nie była możliwa porównawcza analiza liczbowa, jednak poprzez stworzenie poglądowego modelu uproszczonej sieci transportowej natężenie potoków podróży prezentuje wartości odpowiadające rzeczywistym przemieszczeniom podróży w ośrodku miejskim tej wielkości. Dość duże różnice w ruchotwórczości rejonów tej samej kategorii (np. ścisłego centrum) mogły wpłynąć na rozbieżności potoków podróży rozłożonych na uproszczoną sieć aglomeracji analizowanych w podrozdziale 6.2. Sieć nie uwzględnia podziału modalnego, co również mogło wpłynąć na rozbieżności analizowane w menedżerze scenariuszy programu PTV Visum. Pomimo tych problemów wyniki rozprawy są satysfakcjonujące i potwierdzają tezy badawcze. Na ich podstawie można stwierdzić, że metoda ma sens i jest dobrym kierunkiem przyszłych badań, które opisano w podrozdziale 8.2.

7.2. Ocena metody

Zaletą metody jest zdecydowanie jej niewielki koszt. Jest to metoda, która wymaga tylko dostępu do oprogramowania w postaci arkusza kalkulacyjnego, opcjonalnie do programu do

modelowania ruchu, co ostatecznie zdeterminowało temat niniejszej rozprawy doktorskiej. Możliwe jest wykonanie obliczeń odręcznie, jednak jest to nieefektywne czasowo. Metoda jest również elastyczna, pozwala uwzględniać i wykluczać poszczególne bazy danych i nie ogranicza się tylko do wykorzystanych w tej rozprawie dziewięciu ze źródeł Big Data. Nie ma również problemu z wykorzystaniem źródeł szczytkowych lub niepokrywających całej macierzy podróży. Metoda jest skonstruowana w ten sposób, że braki w macierzach nie wpływają na wynik w żaden sposób, na przykład sztucznie zaniżając liczbę podróży w danej relacji międzyrejonowej.

Elementem nieuwzględnionym w metodzie łączenia danych jest wstępne przygotowanie macierzy podróży. W przypadku wykorzystania źródła, które dzieli dany ośrodek miejski na większą liczbę rejonów, należy je zagregować. Jest to pole do dalszej pracy nad metodą. Metoda jest również ograniczona do macierzy podróży, czyli tylko rozkładu przestrzennego podróży. W obecnej formie możliwe jest wpływanie na wyniki baz danych, które mają mocno zawyżone lub zaniżone wyniki, niemniej można to poprawić ręcznie, na przykład wykluczając nierealistyczną macierz podróży lub jej części. W kwestii wyników rozprawy, a dokładnie liczby podróży w stworzonej metodą macierzy, wyniki prezentują dobre i realistyczne wartości. Jednak przy rozłożeniu potoków podróży na wygenerowaną uproszczoną sieć transportową miasta, odchylenia na poszczególnych połączeniach prezentują nawet kilkudziesięcioprocentowe wartości. W przypadku metody symulowania syntetycznego ośrodka miejskiego niekorzystna jest konieczność odrębnego zasymulowania układu miasta, pomimo generacji wartości liczbowych.

Kierunki dalszych, planowanych prac badawczych mających na celu ulepszenie metody i uzupełnienie jej braków opisano w podrozdziale 8.2. Nie wprowadzono ich ze względu na ograniczone ramy czasowe realizacji rozprawy i konieczności zaangażowania większych zasobów do realizacji poszczególnych propozycji rozwinięcia metod stworzonych w rozprawie.

8. Wnioski końcowe i kierunki dalszych prac

8.1. Realizacja celu

By zrealizować cel rozprawy doktorskiej dokonano przeglądu literatury. Zasięgnięto do polskich i anglojęzycznych pozycji opisując ogólne pojęcia i tematykę, a w celu analizy aktualnej wiedzy na temat modelowania ruchu i wykorzystania w nim źródeł Big Data przeanalizowano anglojęzyczne pozycje literaturowe z ostatnich lat oraz w przypadku braku pozycji „świeższych” sięgano po pozycje starsze. Oprócz tego dokonano poszukiwań pozycji zawierających bazy danych i macierze podróży lub sam ich opis i mapy ciepła macierzy podróży. W jednym z przypadków opisanych w rozprawie pozyskano pełne robocze arkusze kalkulacyjne bezpośrednio od autora jednej z prac. Na tej podstawie w metodzie wykorzystano sześć źródeł Big Data – z sieci telefonii komórkowej, sondowania pojazdów (GPS), APC, ANPR, Smart Card Data i skanerów Bluetooth, oraz zasymulowano trzy – dane z Taxi, floty UTO i rowerów miejskich oraz mediów społecznościowych.

By zrealizować główny cel doktoratu w postaci stworzenia metody łączenia wielu zróżnicowanych źródeł danych z dużych wolumenów oprócz pozyskania dziewięciu źródeł danych do badań, dodatkowo założono cel poboczny – stworzenie metody symulacji ośrodka miejskiego do modelowania ruchu. Wykorzystując sposób symulowania danych podobny do metody Monte Carlo stworzono realistyczną macierz podróży ośrodka miejskiego. Pomyślna realizacja tego celu pozwoliła zredukować tendencyjność danych poprzez brak oparcia badań o charakterystyczny istniejący ośrodek miejski. Za pomocą modelu symulowanego ośrodka dokonano również wizualizacji wyników działania metody.

Wykorzystując dziewięć macierzy podróży z różnych źródeł danych wykorzystano stworzoną metodę agregacji danych. Wykorzystanie metody było skuteczne, w liczbie podróży uzyskano wynik różniący się o 2,22% od modelu bazowego aglomeracji, choć różnice na poszczególnych połączeniach w modelu były większe. W celu sprawdzenia, czy zasadnym jest łączenie tak dużej liczby źródeł stworzono jeszcze osiem wariantów zestawienia mniejszej liczby źródeł danych łącząc źródła „rokujące” na lepsze wyniki. Ostatecznie warianty odbiegały od wariantu bazowego aż od 63,64%, do zaledwie 4,80%. Ze względu na brak sukcesu w stworzeniu wariantu konkurencyjnego do wykorzystania wszystkich źródeł, udało się potwierdzić tezy badawcze.

8.2. Kierunki dalszych prac

Pierwszym z kierunków dalszych prac badawczych wydaje się rozwinięcie metody w celu uwzględnienia innych stadiów czterostopniowego klasycznego modelu ruchu. Szczególnie podział na środki lokomocji oraz podział ruchu na sieć transportową, ponieważ w rozprawie sieć transportowa zasymulowanego ośrodka miejskiego została stworzona do celów prezentacji i analizy wyników, a nie szczegółowej analizy i jest ze względu czytelności uproszczona.

Drugim kierunkiem prac byłoby rozbudowanie metody o dodatkowe wagi nadane poszczególnym źródłom danych. Źródła danych mające częściej tendencyjne i bardziej zaburzone dane byłyby uwzględniane w metodzie ze zmniejszoną wagą. W aktualnej wersji rozprawy nie uwzględniono tych wag, ponieważ wymagałoby to większego doświadczenia autora na temat poszczególnych studiów przypadków wykorzystania każdego z dziewięciu źródeł danych użytych w rozprawie lub stworzenia czegoś na miarę metaanalizy dotyczącej wiarygodności każdego z tych źródeł. Ze względu na ograniczone ramy czasowe realizacji rozprawy jest to przewidziane na inną przyszłą pracę badawczą autora.

Najbardziej prostym kierunkiem dalszych prac będzie sprawdzenie metody na realnych źródłach danych, gdy dla konkretnego istniejącego ośrodka miejskiego lub aglomeracji będzie dostępna większa liczba źródeł danych. Wiąże się to jednak z zaangażowaniem większych zasobów w celu pozyskania źródeł danych oraz aktualnym brakiem dostępności dużej liczby źródeł danych dla jednego ośrodka miejskiego. Autor pokłada nadzieję m. in. na wprowadzenie systemów ANPR we Wrocławiu przy planowanej Strefie Czystego Transportu.

Inne kierunki dalszych badań dotyczyć będą metody symulacji ośrodka miejskiego. Po pierwsze należy przetestować ją w dosymulowaniu brakujących danych niezbędnych do stworzenia modelu ruchu dla miast, które nie posiadają szczegółowych Kompleksowych Badań Ruchu. Pierwsza może być Jelenia Góra, dla której autor stworzył prosty model ruchu w pracy magisterskiej pt. „Model sieci transportowej w Jeleniej Górze”, której promotorem również był Promotor niniejszej rozprawy Prof. Maciej Kruszyna.

Drugim kierunkiem dotyczącym tej metody może być wykorzystanie narzędzi heurystycznych by generować losowe sieci transportowe miast. Pozwoli to na inne badania źródeł Big Data na wielu różnych sieciach transportowych, by upewnić się, że stworzona ręcznie sieć nie ma

elementów tendencyjności. Przetestowany w ten sposób może być ponownie przykład z niniejszej rozprawy przed uzyskaniem zasobów na pozyskanie rzeczywistych źródeł danych.

Spis rysunków

Rys. 1.1 Poglądowy wzrost liczby źródeł Big Data wraz z upływem lat	10
Rys. 2.1 Skanowanie kodu QR na przystanku	16
Rys. 2.2 Diagram Woronoja – rejonny stworzone przez granice oddziaływania stacji bazowych (czerwone trójkąty) w obszarze ulicy Powstańców Śląskich (zaznaczonej na zielono) (Opracowanie własne na podkładzie mapowym [169])	19
Rys. 4.1 Formuła uzupełniania danych w formie schematu blokowego	41
Rys. 4.2 Graficzna forma metody. Utworzona macierz w kolorze zielonym. Uzupełniana komórka w danym kroku kolorem niebieskim	42
Rys. 5.1 Układ rejonów transportowych w centrum aglomeracji	50
Rys. 5.2 Układ rejonów w całej aglomeracji	51
Rys. 5.3 Układ węzłów w całej aglomeracji	52
Rys. 5.4 Układ połączeń w obszarze centralnym aglomeracji	53
Rys. 5.5 Układ połączeń w całej aglomeracji	54
Rys. 5.6 Układ połączeń bez siatki rejonów transportowych	55
Rys. 5.7 Wycinek modelu pokazujący sposób przyłączania rejonów do sieci	56
Rys. 5.8 Obszar centralny zasymulowanej aglomeracji	57
Rys. 5.9 Widok na obszar całej aglomeracji	57
Rys. 5.10 Mapa ciepła macierzy podróży modelu bazowego	58
Rys. 5.11 Macierz podróży z sieci telefonii komórkowej – mapa ciepła	59
Rys. 5.12 Macierz podróży z GPS – mapa ciepła	60
Rys. 5.13 Macierz podróży z Bluetooth – mapa ciepła 29x29 rejonów z danymi	61
Rys. 5.14 Macierz podróży z APC – mapa ciepła	62
Rys. 5.15 Macierz podróży z kart miejskich obszar miejski – mapa ciepła	62
Rys. 5.16 Macierz podróży z kart miejskich obszar podmiejski – mapa ciepła	63
Rys. 5.17 Macierz podróży z ANPR – mapa ciepła	64
Rys. 5.18 Macierz podróży z UTO i rowerów miejskich – mapa ciepła	65
Rys. 5.19 Macierz podróży z Taxi – mapa ciepła	66
Rys. 5.20 Macierz podróży z mediów społecznościowych – mapa ciepła	67
Rys. 6.1 Wariant 1, widok na obszar centrum aglomeracji	71
Rys. 6.2 Wariant 1, widok na obszar całej aglomeracji	72
Rys. 6.3 Wariant 1, widok na obszar centrum aglomeracji – wartości procentowe	73
Rys. 6.4 Wariant 1, widok na obszar całej aglomeracji – wartości procentowe	74
Rys. 6.5 Wariant 1, mapa ciepła macierzy podróży	75
Rys. 6.6 Wariant 2, widok na obszar centrum aglomeracji	76
Rys. 6.7 Wariant 2, widok na obszar całej aglomeracji	77
Rys. 6.8 Wariant 2, widok na obszar centrum aglomeracji – wartości procentowe	78
Rys. 6.9 Wariant 2, widok na obszar całej aglomeracji – wartości procentowe	79
Rys. 6.10 Wariant 2, mapa ciepła macierzy podróży	80
Rys. 6.11 Wariant 3, widok na obszar centrum aglomeracji	81
Rys. 6.12 Wariant 3, widok na obszar całej aglomeracji	82
Rys. 6.13 Wariant 3, widok na obszar centrum aglomeracji – wartości procentowe	83
Rys. 6.14 Wariant 3, widok na obszar całej aglomeracji – wartości procentowe	84

Rys. 6.15	Wariant 3, mapa ciepła macierzy podróży	85
Rys. 6.16	Wariant 4, widok na obszar centrum aglomeracji.....	86
Rys. 6.17	Wariant 4, widok na obszar całej aglomeracji	87
Rys. 6.18	Wariant 4, widok na obszar centrum aglomeracji – wartości procentowe	88
Rys. 6.19	Wariant 4, widok na obszar całej aglomeracji – wartości procentowe	89
Rys. 6.20	Wariant 4, mapa ciepła macierzy podróży	90
Rys. 6.21	Wariant 5, widok na obszar centrum aglomeracji.....	91
Rys. 6.22	Wariant 5, widok na obszar całej aglomeracji	92
Rys. 6.23	Wariant 5, widok na obszar centrum aglomeracji – wartości procentowe	93
Rys. 6.24	Wariant 5, widok na obszar całej aglomeracji – wartości procentowe	94
Rys. 6.25	Wariant 5, mapa ciepła macierzy podróży	95
Rys. 6.26	Wariant 6, widok na obszar centrum aglomeracji.....	96
Rys. 6.27	Wariant 6, widok na obszar całej aglomeracji	97
Rys. 6.28	Wariant 6, widok na obszar centrum aglomeracji – wartości procentowe	98
Rys. 6.29	Wariant 6, widok na obszar całej aglomeracji – wartości procentowe	99
Rys. 6.30	Wariant 6, mapa ciepła macierzy podróży	100
Rys. 6.31	Wariant 7, widok na obszar centrum aglomeracji.....	101
Rys. 6.32	Wariant 7, widok na obszar całej aglomeracji	102
Rys. 6.33	Wariant 7, widok na obszar centrum aglomeracji – wartości procentowe	103
Rys. 6.34	Wariant 7, widok na obszar całej aglomeracji – wartości procentowe	104
Rys. 6.35	Wariant 7, mapa ciepła macierzy podróży	105
Rys. 6.36	Wariant 8, widok na obszar centrum aglomeracji.....	106
Rys. 6.37	Wariant 8, widok na obszar całej aglomeracji	107
Rys. 6.38	Wariant 8, widok na obszar centrum aglomeracji – wartości procentowe	108
Rys. 6.39	Wariant 8, widok na obszar całej aglomeracji – wartości procentowe	109
Rys. 6.40	Wariant 8, mapa ciepła macierzy podróży	110
Rys. 6.41	Wariant 9, widok na obszar centrum aglomeracji.....	111
Rys. 6.42	Wariant 9, widok na obszar całej aglomeracji	112
Rys. 6.43	Wariant 9, widok na obszar centrum aglomeracji – wartości procentowe	113
Rys. 6.44	Wariant 9, widok na obszar całej aglomeracji – wartości procentowe	114
Rys. 6.45	Wariant 9, mapa ciepła macierzy podróży	115

Spis tabel

Tab. 2.1. Macierz podróży oparta na bazie danych pochodzącej ze skanerów Bluetooth [62]	25
Tab. 2.2 Przykład luk w detekcji za pomocą skanera Bluetooth, białe pola to brak wykrycia, czerwone to wykrycie [10]	26
Tab. 2.3 Przykład bazy danych ze Smart Card Data [78]	27
Tab. 2.4 Pierwszy przykład bazy danych z sieci telefonii komórkowej [27]	27
Tab. 2.5 Drugi przykład bazy danych z sieci telefonii komórkowej [27]	27
Tab. 2.6 Trzeci przykład bazy danych z sieci telefonii komórkowej [68]	27
Tab. 2.7 Czwarty przykład bazy danych z sieci telefonii komórkowej. Tabela dzieli się na trzy części, pierwsza dotyczy podróży od stacji bazowej do stacji bazowej, druga jest pośrednia pomiędzy rejonem transportowym a stacją bazową, trzecia to podróże zakwalifikowane do konkretnych rejonów transportowych [68]	27
Tab. 2.8 Przykład bazy danych z AFC [139]	28
Tab. 2.9 Piąty przykład bazy danych z CDR [139]	28
Tab. 2.10 Przetworzone dane z taksówek w Nowym Jorku [124]	28
Tab. 5.1 Zestawienie sposobu generacji liczby mieszkańców dla rejonów	47
Tab. 5.2 Zestawienie współczynników ruchliwości	47
Tab. 5.3 Maksymalne wartości procentowe atrakcyjności rejonów	48
Tab. 5.4 Minimalne wartości procentowe atrakcyjności rejonów	48
Tab. 6.1 Zestawienie sumy podróży w poszczególnych wariantach	116
Tab. 6.2 Zestawienie odchyleń standardowych w poszczególnych wariantach	117

Spis załączników

Załącznik 1 - Pseudokod uzupełniania jednej relacji w macierzy podróży z rejonu i-tego do rejonu j-tego

Załącznik 2 - Formuła uzupełniania jednej relacji w macierzy podróży z rejonu i-tego do rejonu j-tego z programu MS Excel , oraz forma w postaci kodu programowania tej samej formuły

Załącznik 3 - Macierz podróży symulowanego ośrodka miejskiego – Model Bazowy

Załącznik 4 – Macierz podróży – Wariant 1

Załącznik 5 – Macierz podróży – Wariant 2

Załącznik 6 – Macierz podróży – Wariant 3

Załącznik 7 – Macierz podróży – Wariant 4

Załącznik 8 – Macierz podróży – Wariant 5

Załącznik 9 – Macierz podróży – Wariant 6

Załącznik 10 – Macierz podróży – Wariant 7

Załącznik 11 – Macierz podróży – Wariant 8

Załącznik 12 – Macierz podróży – Wariant 9

Literatura

1. Abbott-Jard M, Shah H, Bhaskar A (2013) Empirical evaluation of Bluetooth and Wifi scanning for road transport. Australasian Transport Research Forum 2013 Proceedings, Brisbane, Australia
2. Abedi N, Bhaskar A, Chung E (2013) Bluetooth and Wi-Fi MAC Address Based Crowd Data Collection and Monitoring: Benefits, Challenges and Enhancement. Australasian Transport Research Forum 2013 Proceedings, Brisbane, Australia
3. Ackaah W, Bogenberger K, Bertini RL (2019) Empirical evaluation of real-time traffic information for in-vehicle navigation and the variable speed limit system. *Journal of Intelligent Transportation Systems* 23:499–512. doi: 10.1080/15472450.2018.1563864
4. Afandizadeh Zargari S, Memarnejad A, Mirzahosseini H (2021) Hourly Origin–Destination Matrix Estimation Using Intelligent Transportation Systems Data and Deep Learning. *Sensors* 21:7080. doi: 10.3390/s21217080
5. Ahmad F, Samui P, Mishra SS (2024) Machine learning-enhanced Monte Carlo and subset simulations for advanced risk assessment in transportation infrastructure. *J Mt Sci* 21:690–717. doi: 10.1007/s11629-023-8388-8
6. Aldrich PR, El-Zabet J, Hassan S, Briguglio J, Aliaj E, Radcliffe M, Mirza T, Comar T, Nadolski J, Huebner CD (2015) Monte Carlo tests of small-world architecture for coarse-grained networks of the United States railroad and highway transportation systems. *Physica A: Statistical Mechanics and its Applications* 438:32–39. doi: 10.1016/j.physa.2015.06.013
7. Aliari S, Nasri A, Motalleb Nejad M, Haghani A (2020) Toward sustainable travel: An analysis of campus bikeshare use. *Transportation Research Interdisciplinary Perspectives* 6:100162. doi: 10.1016/j.trip.2020.100162
8. Alrawad M, Lutfi A, Almaiah MA, Elshaer IA (2023) Examining the influence of trust and perceived risk on customers intention to use NFC mobile payment system. *Journal of Open Innovation: Technology, Market, and Complexity* 9:100070. doi: 10.1016/j.joitmc.2023.100070
9. Anderson JW, Kennedy KE, Ngo LB, Luckow A, Apon AW (2014) Synthetic data generation for the internet of things. In: 2014 IEEE International Conference on Big Data (Big Data). pp 171–176
10. Araghi BN, Hammershøj Olesen J, Krishnan R, Tørholm Christensen L, Lahrmann H (2015) Reliability of Bluetooth Technology for Travel Time Estimation. *Journal of Intelligent Transportation Systems* 19:240–255. doi: 10.1080/15472450.2013.856727
11. Babu EK, Bashir I, Pillai G, Jyothi KC (2021) Generation of synthetic datasets for transformer’s dissolved gas analysis using Monte-Carlo Simulation. In: 2021 56th International Universities Power Engineering Conference (UPEC). pp 1–4
12. Bao L, Wang Q, Jiang Y (2021) Review of Digital twin for intelligent transportation system. In: 2021 International Conference on Information Control, Electrical Engineering and Rail Transit (ICEERT). IEEE, Lanzhou, China, pp 309–315

13. Barceló J (2010) Fundamentals of Traffic Simulation. Springer, New York, NY
14. Beeston L, Blewitt R, Bulmer S, Wilson J (2021) Traffic Modelling Guidelines v4
15. Bhaskar A, Chung E (2013) Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. *Transportation Research Part C: Emerging Technologies* 37:42–72. doi: 10.1016/j.trc.2013.09.013
16. Bhaskar A, Kieu LM, Qu M, Nantes A, Chung E (2013) On the use of Bluetooth MAC Scanners for live reporting of the transport network. *Proceedings of the Eastern Asia Society for Transportation Studies, Volume 9*
17. Bieler M, Skretting A, Büdinger P, Grønli T-M (2022) Survey of Automated Fare Collection Solutions in Public Transportation. *IEEE Transactions on Intelligent Transportation Systems* 23:14248–14266. doi: 10.1109/TITS.2022.3161606
18. Birr K (2018) Modelowanie podziału zadań przewozowych w obszarach zurbanizowanych
19. Bjerre-Nielsen A, Minor K, Sapieżyński P, Lehmann S, Lassen DD (2020) Inferring transportation mode from smartphone sensors: Evaluating the potential of Wi-Fi and Bluetooth. *PLoS ONE* 15:e0234003. doi: 10.1371/journal.pone.0234003
20. Bonnel P, Hombourger E, Olteanu-Raimond A-M, Smoreda Z (2015) Passive Mobile Phone Dataset to Construct Origin-destination Matrix: Potentials and Limitations. *Transportation Research Procedia* 11:381–398. doi: 10.1016/j.trpro.2015.12.032
21. Bouaru R, Peculea A, Iancu B, Buzura S, Cebuc E, Dadarlat V (2024) Analysis of a Bluetooth Traffic Dataset Obtained during University Examination Sessions. *Data* 9. doi: 10.3390/data9040049
22. Bruni Prenestino F, Barbierato E, Gatti A (2025) Robust Synthetic Data Generation for Sequential Financial Models Using Hybrid Variational Autoencoder–Markov Chain Monte Carlo Architectures. *Future Internet* 17:95. doi: 10.3390/fi17020095
23. Cai Z, Vagena Z, Perez L, Arumugam S, Haas PJ, Jermaine C (2013) Simulation of database-valued markov chains using SimSQL. In: *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. Association for Computing Machinery, New York, NY, USA, pp 637–648
24. Campos Ferreira M, Dias TG, Falcão E Cunha J (2020) Is Bluetooth Low Energy feasible for mobile ticketing in urban passenger transport? *Transportation Research Interdisciplinary Perspectives* 5:100120. doi: 10.1016/j.trip.2020.100120
25. Carrel A, Lau PSC, Mishalani RG, Sengupta R, Walker JL (2015) Quantifying transit travel experiences from the users’ perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. *Transportation Research Part C: Emerging Technologies* 58:224–239. doi: 10.1016/j.trc.2015.03.021
26. Cebecauer M, Gundlegård D, Jenelius E, Burghout W (2023) Spatio-Temporal Public Transport Mode Share Estimation and Analysis Using Mobile Network and Smart Card

- Data. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). pp 2543–2548
27. Chen C, Ma J, Susilo Y, Liu Y, Wang M (2016) The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies* 68:285–299. doi: 10.1016/j.trc.2016.04.005
 28. Chen X, Wang Z, Di X (2023) Sentiment Analysis on Multimodal Transportation during the COVID-19 Using Social Media Data. *Information* 14. doi: 10.3390/info14020113
 29. Cheng Z, Jian S, Rashidi TH, Maghrebi M, Waller ST (2020) Integrating Household Travel Survey and Social Media Data to Improve the Quality of OD Matrix: A Comparative Case Study. *IEEE Transactions on Intelligent Transportation Systems* 21:2628–2636. doi: 10.1109/TITS.2019.2958673
 30. Chiva I-C, Minea M, Minea V-L, Semenescu A, Chiva I-C, Minea M, Minea V-L, Semenescu A (2020) Anonymous Detection of Traveler Flows Employing Bluetooth Technologies. *Proceedings* 63. doi: 10.3390/proceedings2020063061
 31. Choi M, Lee J, Kim S, Jeong Y-S, Park J-H (2016) Location based authentication scheme using BLE for high performance digital content management system. *Neurocomputing* 209:25–38. doi: 10.1016/j.neucom.2015.05.142
 32. Couto R, Leal J, Costa PM, Galvao T (2015) Exploring Ticketing Approaches Using Mobile Technologies: QR Codes, NFC and BLE. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, Gran Canaria, Spain, pp 7–12
 33. Crawford F, Watling DP, Connors RD (2018) Identifying road user classes based on repeated trip behaviour using Bluetooth data. *Transportation Research Part A: Policy and Practice* 113:55–74. doi: 10.1016/j.tra.2018.03.027
 34. Croce AI, Musolino G, Rindone C, Vitetta A (2019) Transport System Models and Big Data: Zoning and Graph Building with Traditional Surveys, FCD and GIS. *IJGI* 8:187. doi: 10.3390/ijgi8040187
 35. Dabiri S, Marković N, Heaslip K, Reddy CK (2020) A deep convolutional neural network based approach for vehicle classification using large-scale GPS trajectory data. *Transportation Research Part C: Emerging Technologies* 116:102644. doi: 10.1016/j.trc.2020.102644
 36. Daniş F, Cemgil A (2017) Model-Based Localization and Tracking Using Bluetooth Low-Energy Beacons. *Sensors* 17:2484. doi: 10.3390/s17112484
 37. Daniş FS, Cemgil AT, Daniş FS, Cemgil AT (2017) Model-Based Localization and Tracking Using Bluetooth Low-Energy Beacons. *Sensors* 17. doi: 10.3390/s17112484
 38. Delipei GK, Rouxelin P, Avramova M, Ivanov E, Ivanov K (2024) Doppler effective temperature database based on Monte Carlo simulations
 39. Deriba AZ, Yang DY (2025) Performance-Based Risk Assessment for Large-Scale Transportation Networks Using the Transitional Markov Chain Monte Carlo Method.

-
- ASCE-ASME J Risk Uncertainty Eng Syst, Part A: Civ Eng 11:04024090. doi: 10.1061/AJRUA6.RUENG-1402
40. Dong H, Wu M, Ding X, Chu L, Jia L, Qin Y, Zhou X (2015) Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C: Emerging Technologies* 58:278–291. doi: 10.1016/j.trc.2015.06.007
 41. Dong H, Wu M, Ding X, Chu L, Jia L, Qin Y, Zhou X (2015) Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C: Emerging Technologies* 58:278–291. doi: 10.1016/j.trc.2015.06.007
 42. Faghih-Roohi S, Xie M, Ng KM (2014) Accident risk assessment in marine transportation via Markov modelling and Markov Chain Monte Carlo simulation. *Ocean Engineering* 91:363–370. doi: 10.1016/j.oceaneng.2014.09.029
 43. Fan W (David), Chen Z (2018) Estimation of Origin-Destination Matrix and Identification of User Activities Using Public Transit Smart Card Data. <https://rosap.ntl.bts.gov>. Accessed 22 Jan 2026
 44. Ferreira MC, Fontes T, Costa V, Dias TG, Borges JL, E Cunha JF (2017) Evaluation of an integrated mobile payment, route planner and social network solution for public transport. *Transportation Research Procedia* 24:189–196. doi: 10.1016/j.trpro.2017.05.107
 45. Forghani M, Karimipour F, Claramunt C (2020) From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transportation Research Part C: Emerging Technologies* 117:102666. doi: 10.1016/j.trc.2020.102666
 46. Gaca S, Suchorzewski W, Tracz M (2014) *Inżynieria ruchu drogowego: teoria i praktyka*
 47. Gao Y, Qian S, Li Z, Wang P, Wang F, He Q (2021) Digital Twin and Its Application in Transportation Infrastructure. In: 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI). IEEE, Beijing, China, pp 298–301
 48. Gáspár P, Szalay Z, Aradi S (2014) Highly Automated Vehicle Systems. *BME MOGI*
 49. Gerpott TJ, Meinert P (2017) Who signs up for NFC mobile payment services? Mobile network operator subscribers in Germany. *Electronic Commerce Research and Applications* 23:1–13. doi: 10.1016/j.eleap.2017.03.002
 50. Gheorghiu RA, Iordache V, Cormoș AC (2021) Analysis of the Possibility to Detect Road Vehicles via Bluetooth Technology. *Sensors* 21. doi: 10.3390/s21217281
 51. Ghosh B, Dauwels J (2022) Comparison of different Bayesian methods for estimating error bars with incident duration prediction. *Journal of Intelligent Transportation Systems* 26:420–431. doi: 10.1080/15472450.2021.1894936
 52. Girolami M, La Rosa D, Barsocchi P (2022) A Bluetooth Dataset for Proximity Detection in Indoor Environments Collected with Smartphones. 1. doi: 10.17632/sbhch9pxwh.1
-

53. Girolami M, La Rosa D, Barsocchi P (2024) Bluetooth dataset for proximity detection in indoor environments collected with smartphones. *Data in Brief* 53:110215. doi: 10.1016/j.dib.2024.110215
54. Gkiotsalitis K, Stathopoulos A (2015) A utility-maximization model for retrieving users' willingness to travel for participating in activities from big-data. *Transportation Research Part C: Emerging Technologies* 58:265–277. doi: 10.1016/j.trc.2014.12.006
55. Gołębiowska A (2021) Wyniki WBR 2015 | Mobilna Warszawa. <https://web.archive.org/web/20210418151426/http://transport.um.warszawa.pl/warszawskie-badanie-ruchu-2015/wyniki-wbr-2015>. Accessed 22 Jan 2026
56. Gomez C, Oller J, Paradells J (2012) Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. *Sensors* 12:11734–11753. doi: 10.3390/s120911734
57. Griffin GP, Mulhall M, Simek C, Riggs WW (2020) Mitigating Bias in Big Data for Transportation. *SocArXiv*
58. Gudishala R, Wilmot C, Mokkapatti A (2016) Travel Time Estimation Using Bluetooth. Department of Civil and Environmental Engineering Louisiana State University Baton Rouge, LA 70803
59. Guimarães VDA, Leal Junior IC, Da Silva MAV (2018) Evaluating the sustainability of urban passenger transportation by Monte Carlo simulation. *Renewable and Sustainable Energy Reviews* 93:732–752. doi: 10.1016/j.rser.2018.05.015
60. Gunda PK (2021) Network-wide traffic congestion visual analytics: A case study for Brisbane Bluetooth MAC scanner data. Master of Philosophy, Queensland University of Technology
61. Hagen-Zanker A, Jin Y (2015) Adaptive Zoning for Efficient Transport Modelling in Urban Models. In: Gervasi O, Murgante B, Misra S, Gavrilova ML, Rocha AMAC, Torre C, Tanir D, Apduhan BO (eds) *Computational Science and Its Applications -- ICCSA 2015*. Springer International Publishing, Cham, pp 673–687
62. Hainen AM, Wasson JS, Hubbard SML, Remias SM, Farnsworth GD, Bullock DM (2011) Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. *Transportation Research Record: Journal of the Transportation Research Board* 2256:43–50. doi: 10.3141/2256-06
63. Hosford K, Therrien S, Fuller D, Gauvin L, Teschke K, Brauer M, Winters M (2017) Assessing the Modal Impacts of Public Bike Share Systems: A Comparison of Survey Tools
64. Huang A, Levinson D (2015) Axis of travel: Modeling non-work destination choice with GPS data. *Transportation Research Part C: Emerging Technologies* 58:208–223. doi: 10.1016/j.trc.2015.03.022
65. Huo J, Yang H, Li C, Zheng R, Yang L, Wen Y (2021) Influence of the built environment on E-scooter sharing ridership: A tale of five cities. *Journal of Transport Geography* 93:103084. doi: 10.1016/j.jtrangeo.2021.103084

-
66. Hussain E, Behara KNS, Bhaskar A, Chung E (2019) A framework to quantify the differences between multi-modal travel demand: A case study on Brisbane network using Bluetooth and smartcard data
 67. info@qrtiger.com BC QR Code Generator Statistics 2025: Latest Global Report. <https://www.qrcode-tiger.com/qr-code-statistics>. Accessed 28 Nov 2025
 68. Iqbal MdS, Choudhury CF, Wang P, González MC (2014) Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies* 40:63–74. doi: 10.1016/j.trc.2014.01.002
 69. Irfan MS, Dasgupta S, Rahman M (2024) Toward Transportation Digital Twin Systems for Traffic Safety and Mobility: A Review. *IEEE Internet Things J* 11:24581–24603. doi: 10.1109/JIOT.2024.3395186
 70. Jampani R, Xu F, Wu M, Perez LL, Jermaine C, Haas PJ (2008) MCDB: a monte carlo approach to managing uncertain data. In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, Vancouver Canada, pp 687–700
 71. Jenelius E (2020) Personalized predictive public transport crowding information with automated data sources. *Transportation Research Part C: Emerging Technologies* 117:102647. doi: 10.1016/j.trc.2020.102647
 72. Ji Y, Mishalani RG, McCord MR (2015) Transit passenger origin–destination flow estimation: Efficiently combining onboard survey and large automatic passenger count datasets. *Transportation Research Part C: Emerging Technologies* 58:178–192. doi: 10.1016/j.trc.2015.04.021
 73. Jin PJ, Cebelak M, Yang F, Zhang J, Walton CM, Ran B (2014) Location-Based Social Networking Data: Exploration into Use of Doubly Constrained Gravity Model for Origin–Destination Estimation. *Transportation Research Record* 2430:72–82. doi: 10.3141/2430-08
 74. Kammouh O, Cardoni A, Marasco S, Cimellaro GP, Mahin S (2018) Resilience Assessment of City-Scale Transportation Networks Using Monte Carlo Simulation. In: Powers N, Frangopol DM, Al-Mahaidi R, Caprani C (eds) *Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges*, 1st ed. CRC Press, pp 1902–1910
 75. Karaca H (2021) Generation of synthetic catalog by using Markov chain Monte Carlo simulation and inverse Poisson distribution. *J Seismol* 25:1103–1114. doi: 10.1007/s10950-021-10018-z
 76. Kasemsri R, Ishizaka T, Fukuda A *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol.12,2019
 77. Kazemzadeh K, Sprei F (2024) The effect of shared e-scooter programs on modal shift: Evidence from Sweden. *Sustainable Cities and Society* 101:105097. doi: 10.1016/j.scs.2023.105097
-

78. Kieu L-M, Bhaskar A, Chung E (2015) A modified Density-Based Scanning Algorithm with Noise for spatial travel pattern analysis from Smart Card AFC data. *Transportation Research Part C: Emerging Technologies* 58:193–207. doi: 10.1016/j.trc.2015.03.033
79. Konarski A (2014) Wykorzystanie skryptów w programie VISUM, na przykładzie modelu ruchu województwa śląskiego. *Zeszyty Naukowo-Techniczne Stowarzyszenia Inżynierów i Techników Komunikacji w Krakowie Seria: Materiały Konferencyjne* Nr 1(103)
80. Kruszyna M (2021) Investment challenges pertaining to the achievement of the goals of the Mobility Policy based on the analysis of the results of traffic surveys in Wrocław. *Archives of Civil Engineering*; 2021; vol 67; No 3; 505-523
81. Krych A (2018) *Badania Kompleksowe, Modelowanie i Planowanie Ruchu - Słownik Terminologiczny*
82. Krystek R, Szczuraszek T, Hoppe LN, Chodur J, Gondek S, Tracz M, Kroman T, Rakiewicz M, Molisz W, Sosin J, Kaczmarek J, Barne B, Urbański A (1980) *Symulacja ruchu potoku pojazdów - Wybrane zagadnienia*. Wydawnictwa Komunikacji i Łączności, Warszawa
83. Kucharski R (2010) Otwieranie Oprogramowania – skrypty i programy w programie Visum. Nr 153
84. Kuśić K, Schumann R, Ivanjko E (2023) A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Advanced Engineering Informatics* 55:101858. doi: 10.1016/j.aei.2022.101858
85. Larijani AN, Olteanu-Raimond A-M, Perret J, Brédif M, Ziemlicki C (2015) Investigating the Mobile Phone Data to Estimate the Origin Destination Flow and Analysis; Case Study: Paris Region. *Transportation Research Procedia* 6:64–78. doi: 10.1016/j.trpro.2015.03.006
86. Le TA, Baydin AG, Zinkov R, Wood F (2017) Using synthetic data to train neural networks is model-based reasoning. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, Anchorage, AK, USA, pp 3514–3521
87. Lee M, Jeon I, Jun C (2022) A Deterministic Methodology Using Smart Card Data for Prediction of Ridership on Public Transport. *Applied Sciences* 12. doi: 10.3390/app12083867
88. Lei D, Chen X, Cheng L, Zhang L, Ukkusuri SV, Witlox F (2020) Inferring temporal motifs for travel pattern analysis using large scale smart card data. *Transportation Research Part C: Emerging Technologies* 120:102810. doi: 10.1016/j.trc.2020.102810
89. Leurent F, Sun D, Xie X (2023) On Heterogenous Sampling Rates in Origin–Destination Matrix Estimation Based on Trajectory Data and Link Counts. *Transportation Research Record* 2677:1169–1180. doi: 10.1177/03611981221103589
90. Li C, Zheng L, Jia N (2024) Network-wide ride-sourcing passenger demand origin-destination matrix prediction with a generative adversarial network. *Transportmetrica A: Transport Science* 20:2109774. doi: 10.1080/23249935.2022.2109774

-
91. Li L, Su X, Wang Y, Lin Y, Li Z, Li Y (2015) Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C: Emerging Technologies* 58:292–307. doi: 10.1016/j.trc.2015.03.003
 92. Libes D, Lechevalier D, Jain S (2017) Issues in synthetic data generation for advanced manufacturing. In: 2017 IEEE International Conference on Big Data (Big Data). IEEE, Boston, MA, pp 1746–1754
 93. Lin L, Wang Y, Zhang Q, Li Y (2022) A Monte Carlo-based adaptive weighted reduced order model for gamma density measurement. *Annals of Nuclear Energy* 178:109341. doi: 10.1016/j.anucene.2022.109341
 94. Litman T (2003) Measuring transportation: Traffic, mobility and accessibility. *Social Research in Transport (SORT) Clearinghouse* 73(10)
 95. López D, Farooq B (2020) A multi-layered blockchain framework for smart mobility data-markets. *Transportation Research Part C: Emerging Technologies* 111:588–615. doi: 10.1016/j.trc.2020.01.002
 96. Luber A, Bauer S, Junghans M, Schulz J (2011) On Measuring Traffic with Wi-Fi and Bluetooth
 97. Ma X, Yuan Y, Van Oort N, Hoogendoorn S (2020) Bike-sharing systems' impact on modal shift: A case study in Delft, the Netherlands. *Journal of Cleaner Production* 259:120846. doi: 10.1016/j.jclepro.2020.120846
 98. Malinovskiy Y, Saunier N, Wang Y (2012) Analysis of Pedestrian Travel with Static Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board* 2299:137–149. doi: 10.3141/2299-15
 99. Mayne K, Fried L, Ducellier C (2024) Shared Ambition: The potential for bike sharing in Europe: Benchmarking 148 cities. *Cycling Industries Europe*
 100. Mhaske DAS (2021) Difference between Fuzzy and Crisp Transportation Problem Using Pentagonal Fuzzy Numbers with ranking by α -cut Method. *Adv Fuzzy Math* 8
 101. Michau G, Nantes A, Chung E, Abry P, Borgnat P (2014) Retrieving Trip Information from a Discrete Detectors Network: the case of Brisbane Bluetooth detectors. 32nd Conference of Australian Institutes of Transport Research
 102. Mirzahosseini H, Gholampour I, Sedghi M, Zhu L (2021) How realistic is static traffic assignment? Analyzing automatic number-plate recognition data and image processing of real-time traffic maps for investigation. *Transportation Research Interdisciplinary Perspectives* 9:100320. doi: 10.1016/j.trip.2021.100320
 103. Mirzahosseini H, Gholampour I, Sedghi M, Zhu L (2021) How realistic is static traffic assignment? Analyzing automatic number-plate recognition data and image processing of real-time traffic maps for investigation. *Transportation Research Interdisciplinary Perspectives* 9:100320. doi: 10.1016/j.trip.2021.100320
-

104. Nag D, Brandel-Tanis F, Pramestri ZA, Pitera K, Frøyen YK (2025) Exploring digital twins for transport planning: a review. *Eur Transp Res Rev* 17:15. doi: 10.1186/s12544-025-00713-0
105. Ni D, Leonard JD (2005) Markov Chain Monte Carlo Multiple Imputation Using Bayesian Networks for Incomplete Intelligent Transportation Systems Data. *Transportation Research Record: Journal of the Transportation Research Board* 1935:57–67. doi: 10.1177/0361198105193500107
106. Nourmohammadi F, Mansourianfar M, Shafiei S, Gu Z, Saberi M (2021) An Open GMNS Dataset of a Dynamic Multi-Modal Transportation Network Model of Melbourne, Australia. *Data* 6:21. doi: 10.3390/data6020021
107. Okmi M, Por LY, Ang TF, Ku CS (2023) Mobile Phone Data: A Survey of Techniques, Features, and Applications. *Sensors* 23. doi: 10.3390/s23020908
108. Ortúzar S. J de D, Willumsen LG (2024) *Modelling transport*, Fifth edition. Wiley, Hoboken, NJ
109. Park ES, Rilett LR, Spiegelman CH (2008) A Markov Chain Monte Carlo-Based Origin Destination Matrix Estimator that is Robust to Imperfect Intelligent Transportation Systems Data. *Journal of Intelligent Transportation Systems* 12:139–155. doi: 10.1080/15472450802262364
110. Patire AD, Wright M, Prodhomme B, Bayen AM (2015) How much GPS data do we need? *Transportation Research Part C: Emerging Technologies* 58:325–342. doi: 10.1016/j.trc.2015.02.011
111. Pelletier M-P, Trépanier M, Morency C (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19:557–568. doi: 10.1016/j.trc.2010.12.003
112. Peng Y, Mo Z, Liu S (2021) Passenger's routes planning in stochastic common-lines' multi-modal transportation network through integrating Genetic Algorithm and Monte Carlo simulation. *AoT* 59:73–92. doi: 10.5604/01.3001.0015.0123
113. Phua P, Page B, Bogomolova S (2015) Validating Bluetooth logging as metric for shopper behaviour studies. *Journal of Retailing and Consumer Services* 22:158–163. doi: 10.1016/j.jretconser.2014.10.009
114. Plačiak J (2023) Single-route od matrix estimation USING AUTOMATIC PASSENGER COUNTING DATA: A CASE STUDY IN GENEVA
115. Polańczyk M, Strumillo P, Barański P (2011) Fusion of Data from Inertial Sensors, Raster Maps and GPS for Estimation of Pedestrian Geographic Location in Urban Terrain. *Metrology and Measurement Systems*; 2011; No 1; 145-158
116. Putatunda S, Laha AK (2023) Travel Time Prediction in Real time for GPS Taxi Data Streams and its Applications to Travel Safety. *Hum-Cent Intell Syst* 3:381–401. doi: 10.1007/s44230-023-00028-0

-
117. Radzimski A, Dzięcielski M (2021) Exploring the relationship between bike-sharing and public transport in Poznań, Poland. *Transportation Research Part A: Policy and Practice* 145:189–202. doi: 10.1016/j.tra.2021.01.003
 118. Raghunathan TE (2021) Synthetic Data. *Annu Rev Stat Appl* 8:129–140. doi: 10.1146/annurev-statistics-040720-031848
 119. Rahmani M, Jenelius E, Koutsopoulos HN (2015) Non-parametric estimation of route travel time distributions from low-frequency floating car data. *Transportation Research Part C: Emerging Technologies* 58:343–362. doi: 10.1016/j.trc.2015.01.015
 120. Rezaeifar B, Wolfs CJA, Lieuwes NG, Biemans R, Reniers B, Dubois LJ, Verhaegen F (2022) A deep learning and Monte Carlo based framework for bioluminescence imaging center of mass-guided glioblastoma targeting. *Phys Med Biol* 67:144003. doi: 10.1088/1361-6560/ac79f8
 121. Riascos AP, Mateos JL (2020) Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City. *Sci Rep* 10:4022. doi: 10.1038/s41598-020-60875-w
 122. Ribeiro P, Dias G, Pereira P (2021) Transport Systems and Mobility for Smart Cities. *ASI* 4:61. doi: 10.3390/asi4030061
 123. Richter MA, Hagenmaier M, Bandte O, Parida V, Wincent J (2022) Smart cities, urban mobility and autonomous vehicles: How different cities needs different sustainable investment strategies. *Technological Forecasting and Social Change* 184:121857. doi: 10.1016/j.techfore.2022.121857
 124. Robinson A, Venter C (2019) Validating traffic models using large-scale Automatic Number Plate Recognition (ANPR) data. *Journal of the South African Institution of Civil Engineering* 61:45–57. doi: 10.17159/2309-8775/2019/v61n3a5
 125. Rodriguez L, Salimi B, Ping H, Stoyanovich J, Howe B (2019) MobilityMirror: Bias-Adjusted Transportation Datasets. In: Oliveira J, Farias CM, Pacitti E, Fortino G (eds) *Big Social Data and Urban Computing*. Springer International Publishing, Cham, pp 18–39
 126. Roig-Costa O, Miralles-Guasch C, Marquet O (2024) Shared bikes vs. private e-scooters. Understanding patterns of use and demand in a policy-constrained micromobility environment. *Transport Policy* 146:116–125. doi: 10.1016/j.tranpol.2023.11.010
 127. Roucairol M, Georgiou A, Cazenave T, Prischi F, Pardo OE (2024) DrugSynthMC: An Atom-Based Generation of Drug-like Molecules with Monte Carlo Search. *J Chem Inf Model* 64:7097–7107. doi: 10.1021/acs.jcim.4c01451
 128. Sandhya S, k.a S, Paramasivam S (2015) Augmenting Bluetooth Security by Behavior Based User Categorization. *Procedia Computer Science* 57:1424–1431. doi: 10.1016/j.procs.2015.07.462
-

129. Sarkar C, Treurniet JJ, Narayana S, Prasad RV, De Boer W (2018) SEAT: Secure Energy-Efficient Automated Public Transport Ticketing System. *IEEE Trans on Green Commun Netw* 2:222–233. doi: 10.1109/TGCN.2017.2779833
130. Shi Q, Abdel-Aty M (2015) Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies* 58:380–394. doi: 10.1016/j.trc.2015.02.022
131. Sinclair M, Maadi S, Zhao Q, Hong J, Ghermandi A, Bailey N (2023) Assessing the socio-demographic representativeness of mobile phone application data. *Applied Geography* 158:102997. doi: 10.1016/j.apgeog.2023.102997
132. Siwek K (2017) Modele mikrosymulacyjne na Wyspach Brytyjskich : teoria i praktyka. *Transport Miejski i Regionalny*
133. St-Aubin P, Saunier N, Miranda-Moreno L (2015) Large-scale automated proactive road safety analysis using video data. *Transportation Research Part C: Emerging Technologies* 58:363–379. doi: 10.1016/j.trc.2015.04.007
134. Szczudło G (2021) iOS vs. Android – porównanie systemów operacyjnych smartfonów od Apple’a i Google’a. In: *Poradnik Orange*. <https://www.orange.pl/poradnik/ios-vs-android-porownanie-systemow-operacyjnych-smartfonow-od-apple-i-googlea/>. Accessed 28 Nov 2025
135. Tarry S (2016) Study on Public Transport Smart Cards - Final Report
136. Teixeira JF, Silva C, Moura e Sá F (2023) Factors influencing modal shift to bike sharing: Evidence from a travel survey conducted during COVID-19. *Journal of Transport Geography* 111:103651. doi: 10.1016/j.jtrangeo.2023.103651
137. Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, González MC (2015) The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies* 58:162–177. doi: 10.1016/j.trc.2015.04.022
138. Tsubota T, Yoshii T (2017) An Analysis of the Detection Probability of MAC Address from a Moving Bluetooth Device. *Transportation Research Procedia* 21:251–256. doi: 10.1016/j.trpro.2017.03.094
139. Ukkusuri SV, Yang C (2019) *Transportation Analytics in the Era of Big Data*. Springer International Publishing, Cham
140. Universidade do Porto – Faculdade de Engenharia, Porto, Portugal, Ferreira MC, Dias TG, Cunha JFE (2019) Codesign of a Mobile Ticketing Service Solution Based on BLE. *jtle* 10–17. doi: 10.18178/jtle.7.1.10-17
141. Vahidi M, Shafahi Y (2025) Time-dependent estimation of origin–destination matrices using partial path data and link counts. *Transportation* 52:55–92. doi: 10.1007/s11116-023-10412-1
142. Wang F, Ross CL (2019) New potential for multimodal connection: exploring the relationship between taxi and transit in New York City (NYC). *Transportation* 46:1051–1072. doi: 10.1007/s11116-017-9787-x

-
143. Wang L, Wang Y, Sun X, Wu Y, Peng F, Chen C-HP, Song G (2023) Public Transit Passenger Profiling by Using Large-Scale Smart Card Data. *Journal of Transportation Engineering, Part A: Systems* 149:04023013. doi: 10.1061/JTEPBS.TEENG-7634
 144. Wang M, Yu S (2025) Application of Bayesian network and genetic algorithm enhanced Monte Carlo method in urban transportation infrastructure investment. *Journal of Computational Methods in Sciences and Engineering* 25:2442–2458. doi: 10.1177/14727978241313262
 145. Wang X, Khattak AJ, Liu J, Masghati-Amoli G, Son S (2015) What is the level of volatility in instantaneous driving decisions? *Transportation Research Part C: Emerging Technologies* 58:413–427. doi: 10.1016/j.trc.2014.12.014
 146. Wang X, Qian Y, Gao H, Coley CW, Mo Y, Barzilay R, Jensen KF (2020) Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning. *Chem Sci* 11:10959–10972. doi: 10.1039/D0SC04184J
 147. Wang Y, Xiao Y, Lai J, Chen Y (2020) An adaptive k nearest neighbour method for imputation of missing traffic data based on two similarity. *Archives of Transport* 54:59–73. doi: 10.5604/01.3001.0014.2968
 148. Wang Z, Zhang Y, Jia B, Gao Z (2024) Comparative Analysis of Usage Patterns and Underlying Determinants for Ride-hailing and Traditional Taxi Services: A Chicago Case Study. *Transportation Research Part A: Policy and Practice* 179:103912. doi: 10.1016/j.tra.2023.103912
 149. Wismans LJJ, Friso K, Rijdsdijk J, De Graaf SW, Keij J (2018) Improving A Priori Demand Estimates Transport Models using Mobile Phone Data: A Rotterdam-Region Case. *Journal of Urban Technology* 25:63–83. doi: 10.1080/10630732.2018.1442075
 150. Wu Y, Liu T, Gong L, Luo Q, Du B (2024) Mining smart card data to estimate transfer passenger flow in a metro network. *IET Intelligent Transport Systems* 18:1830–1846. doi: 10.1049/itr2.12481
 151. Xu C, Luo H, Bao H, Wang P (2020) STEIM: A Spatiotemporal Event Interaction Model in V2X Systems Based on a Time Period and a Raster Map. *Mobile Information Systems* 2020:1–20. doi: 10.1155/2020/1375426
 152. Yang F Urban Origin-Destination Travel Demand Analysis Using Location-based Social Networking (LBSN) Data. Ph.D.
 153. Younes H, Zou Z, Wu J, Baiocchi G (2020) Comparing the Temporal Determinants of Dockless Scooter-share and Station-based Bike-share in Washington, D.C. *Transportation Research Part A: Policy and Practice* 134:308–320. doi: 10.1016/j.tra.2020.02.021
 154. Yu J, Stettler MEJ, Angeloudis P, Hu S, Chen X (Michael) (2020) Urban network-wide traffic speed estimation with massive ride-sourcing GPS traces. *Transportation Research Part C: Emerging Technologies* 112:136–152. doi: 10.1016/j.trc.2020.01.023
-

155. Zawodny M, Kruszyna M (2025) Data Sources for Traffic Analysis in Urban Canyons—The Comprehensive Literature Review. *Applied Sciences* 15:10686. doi: 10.3390/app151910686
156. Zhang C, Chen J, Fu X (2023) A multilevel analysis of adoption intention of travel information on social media: Evidence from China. *Travel Behaviour and Society* 32:100582. doi: 10.1016/j.tbs.2023.100582
157. Zhang Q, Ma Z, Zhang P, Jenelius E, Ma X, Wen Y (2023) User-station attention inference using smart card data: a knowledge graph assisted matrix decomposition model. *Appl Intell* 53:21944–21960. doi: 10.1007/s10489-023-04678-2
158. Zhao X, Cui M, Levinson D (2023) Exploring temporal variability in travel patterns on public transit using big smart card data. *Environment and Planning B: Urban Analytics and City Science* 50:198–217. doi: 10.1177/23998083221089662
159. Zhao Z, Koutsopoulos HN, Zhao J (2020) Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model. *Transportation Research Part C: Emerging Technologies* 116:102627. doi: 10.1016/j.trc.2020.102627
160. Zhuang D, Hao S, Lee D-H, Jin JG (2020) From compound word to metropolitan station: Semantic similarity analysis using smart card data. *Transportation Research Part C: Emerging Technologies* 114:322–337. doi: 10.1016/j.trc.2020.02.017
161. (2013) Transport analysis guidance (TAG)
162. (2014) Transport model development guidelines, First edition, Amendment 01. NZ Transport Agency, Wellington
163. (2021) How Covid turbocharged the QR revolution. <https://www.bbc.com/news/business-55579480>. Accessed 28 Nov 2025
164. (2025) Pażerski Model Transportowy 1.6. Centralny Port Komunikacyjny
165. (2025) Jaka będzie strefa czystego transportu we Wrocławiu i dlaczego o niej mówimy? In: Wrocław.pl. <https://www.wroclaw.pl/komunikacja/strefa-czystego-transportu-we-wroclawiu-jaka-ma-byc>. Accessed 12 Dec 2025
166. The Australian Transport Assessment and Planning (ATAP) Guidelines. <https://www.atap.gov.au/>. Accessed 7 Feb 2026
167. Kompleksowe Badania Ruchu we Wrocławiu i otoczeniu - KBR 2018 - Kompleksowe Badania Ruchu - Biuletyn Informacji Publicznej Urzędu Miejskiego Wrocławia. <https://bip.um.wroc.pl/artukul/565/37499/kompleksowe-badania-ruchu-we-wroclawiu-i-otoczeniu-kbr-2018>. Accessed 14 Aug 2025
168. Smartphone ownership in the US by age 2023. In: Statista. <https://www.statista.com/statistics/489255/percentage-of-us-smartphone-owners-by-age-group/>. Accessed 14 Aug 2025
169. OpenStreetMap. In: OpenStreetMap. <https://www.openstreetmap.org/>. Accessed 19 Aug 2025

-
170. NYC Yellow Taxi Trip Data. <https://www.kaggle.com/datasets/elemento/nyc-yellow-taxi-trip-data>. Accessed 11 Feb 2026
 171. Ustawa z dnia 30 marca 2021 r. o zmianie ustawy - Prawo o ruchu drogowym oraz niektórych innych ustaw.
<https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20210000720>. Accessed 11 Feb 2026
 172. Minneapolis Transportation Action Plan. In: City of Minneapolis Transportation Action Plan. <http://go.minneapolismn.gov>. Accessed 11 Feb 2026
 173. London and Taipei Bike-Share Data, the Basics.
<https://kaggle.com/code/ajohnr/london-and-taipei-bike-share-data-the-basics>. Accessed 11 Feb 2026
 174. 臺北市資料大平臺. <https://data.taipei/#/dataset?topic=topic-transportation>. Accessed 11 Feb 2026
 175. cycling.data.tfl.gov.uk. <https://cycling.data.tfl.gov.uk/>. Accessed 11 Feb 2026

Streszczenie

Autor: mgr inż. Michał Zawodny

Tytuł: Metoda agregacji danych z dużych wolumenów, zróżnicowanego pochodzenia i struktury w celu udoskonalenia i tworzenia modeli transportowych

Słowa kluczowe: modelowanie transportu, symulacja, inżynieria ruchu, big data, łączenie, macierz podróży

Streszczenie:

Tematem niniejszej rozprawy doktorskiej jest wykorzystanie dużej liczby źródeł danych wpisujących się w definicję Big Data w celu implementacji w modelach transportowych. W rozprawie dokonano przeglądu aktualnej literatury w tematyce modelowania podróży, Big Data i symulacji. Na bazie przeglądu stwierdzono lukę badawczą w postaci niskiej integracji przy wzroście liczby źródeł danych. Do realizacji celu badań stworzono dwie metody. Pierwsza metoda polega na łączeniu macierzy podróży ze źródeł Big Data pomimo ich niekompatybilności i zróżnicowanego pochodzenia. W celu realizacji badań w sposób możliwie niezależny od tendencyjności istniejących i znanych aglomeracji stworzono drugą metodę. Polega ona na zasymulowaniu realistycznego ośrodka miejskiego bez niestandardowych właściwości istniejących ośrodków mogących wpływać negatywnie na wyniki. W tym celu stworzono ośrodek miejski generując w nim podróże i rozkładając przestrzennie pomiędzy rejonami transportowymi. W celu wizualizacji i oceny wyników w programie PTV Visum utworzono poglądowy model makroskopowy tej aglomeracji. Ze względu na badania, które w rozprawie ograniczają się do rozkładu przestrzennego i bazują na macierzach podróży źródłowo-docelowych, w modelu poglądowym nie uwzględniono podziału modalnego. Sieć połączeń była uproszczona i ograniczała się do głównych arterii. Na tak stworzonej sieci w menedżerze scenariuszy programu Visum analizowano wyniki głównej metody. Polegała ona na uwzględnieniu macierzy podróży z odpowiednimi wagami i wykluczaniu nierealistycznych wartości na konkretnych relacjach w macierzy. W celu sprawdzenia, czy łączenie dziewięciu źródeł danych jest zasadne, stworzono osiem wariantów „konkurencyjnych” do wykorzystania wszystkich źródeł. Miało to na celu uzyskanie lepszych wyników. Wariant wykorzystujący pełną metodę ze wszystkimi dostępnymi źródłami danych dał najlepsze rezultaty w ogólnej liczbie podróży i w rozłożeniu potoków na uproszczoną sieć transportową symulowanego ośrodka. Sześć źródeł danych pozyskano z literatury, a pozostałe trzy zasymulowano korzystając z opisów literaturowych i map ciepła przedstawiających macierze podróży z tych źródeł danych. Wskazano zalety metody. Zaplanowano kierunki dalszych prac badawczych.

Abstract

Author: MSc Michał Zawodny

Title: Method of aggregating data from large volumes, diverse origins, and structures in order to improve and create transport models

Keywords: transport modelling, simulation, traffic engineering, big data, combining, O-D matrix

Abstract:

The topic of this PhD dissertation is the use of a large number of data sources that fit the definition of Big Data for implementation in transport models. The dissertation reviews the state-of-the-art literature on travel modeling, Big Data, and simulation. Based on the review, a research gap was identified in the form of low integration with an increase in the number of data sources. Two methods were created to achieve the research goal. The first method involves combining travel matrices from Big Data sources despite their incompatibility and diverse origins. A second method was carried out to conduct the research as independently as possible from the biases of existing and known agglomerations. It involves simulating a realistic urban center without non-standard properties of existing centers that may negatively affect the results. For this purpose, an urban center was simulated, with trip generation and spatial distribution among transportation zones. To visualize and evaluate the results, an illustrative macroscopic model of this agglomeration was created using the PTV Visum program. Due to the research in the thesis being limited to spatial distribution and based on O-D matrices, the illustrative model does not include modal division. The transportation network was simplified and limited to the main arteries. The results of the main method were analyzed on the network created in the Visum scenario manager. The method involved taking into account the O-D matrix with appropriate weights and excluding unrealistic values on specific relationships in the matrix. To evaluate whether combining the selected nine data sources is justified, another eight "competitive" variants were created. The reason for the variants was to obtain better results. The variant using the full method with all available data sources provided the best results in the total number of trips and in the distribution of traffic flows over the simplified transport network of the simulated center. Six data sources were obtained from the literature, and the remaining three were simulated using literature descriptions and heat maps representing O-D matrices from these data sources. The advantages of the method were indicated. The direction of further research was planned.

ZAŁĄCZNIKI

Załącznik 1

Pseudokod uzupełniania jednej relacji w macierzy podróży z rejonu i-tego do rejonu j-tego

```
sources = [  
  {sheet: "OD GSM",    value: GSM_ij,    weight_base: W_GSM},  
  {sheet: "OD SocialM", value: SocialM_ij, weight_base: W_SocialM},  
  {sheet: "OD BT",     value: BT_ij,     weight_base: W_BT},  
  {sheet: "OD APC",    value: APC_ij,    weight_base: W_APC},  
  {sheet: "OD Smart Card", value: SmartCard_ij, weight_base: W_SmartCard},  
  {sheet: "OD GPS",    value: GPS_ij,    weight_base: W_GPS},  
  {sheet: "OD ANPR",   value: ANPR_ij,   weight_base: W_ANPR},  
  {sheet: "OD Taxi",   value: Taxi_ij,   weight_base: W_Taxi},  
  {sheet: "OD UTO",    value: UTO_ij,    weight_base: W_UTO}  
]  
  
sum_weighted = 0  
count_sources = 0  
  
for each source in sources:  
  if source.value  $\neq$  0:  
    weight = 1 / source.weight_base  
    sum_weighted = sum_weighted + (source.value * weight)  
    count_sources = count_sources + 1  
  end if  
end for  
  
if count_sources > 0:  
  result = sum_weighted / count_sources  
else:  
  result = 0  
end if
```

Załącznik 2

Formuła uzupełniania jednej relacji w macierzy podróży z rejonu i-tego do rejonu j-tego z programu MS Excel , oraz forma w postaci kodu programowania tej samej formuły

```
=(JEŻELI('OD GSM'!B2<>0;'OD GSM'!B2*(1/'Założenia i wsp'!$B$9);0)+JEŻELI('OD
SocialM'!B2<>0;'OD SocialM'!B2*(1/'Założenia i wsp'!$B$75);0)+JEŻELI('OD
BT'!B2<>0;'OD BT'!B2*(1/'Założenia i wsp'!$B$24);0)+JEŻELI('OD APC'!B2<>0;'OD
APC'!B2*(1/'Założenia i wsp'!$B$29);0)+JEŻELI('OD Smart Card'!B2<>0;'OD Smart
Card'!B2*(1/'Założenia i wsp'!$B$51);0)+JEŻELI('OD GPS'!B2<>0;'OD
GPS'!B2*(1/'Założenia i wsp'!$B$58);0)+JEŻELI('OD ANPR'!B2<>0;'OD
ANPR'!B2*(1/'Założenia i wsp'!$B$61);0)+JEŻELI('OD Taxi'!B2<>0;'OD
Taxi'!B2*(1/'Założenia i wsp'!$B$67);0)+JEŻELI('OD UTO'!B2<>0;'OD
UTO'!B2*(1/'Założenia i wsp'!$B$34);0))/((JEŻELI('OD GSM'!B2<>0;1;0))+JEŻELI('OD
SocialM'!B2<>0;1;0))+JEŻELI('OD BT'!B2<>0;1;0))+JEŻELI('OD
APC'!B2<>0;1;0))+JEŻELI('OD Smart Card'!B2<>0;1;0))+JEŻELI('OD
GPS'!B2<>0;1;0))+JEŻELI('OD ANPR'!B2<>0;1;0))+JEŻELI('OD
Taxi'!B2<>0;1;0))+JEŻELI('OD UTO'!B2<>0;1;0)))

=(
IF('OD GSM'!B2<>0;'OD GSM'!B2*(1/'Założenia i wsp'!$B$9);0)
+IF('OD SocialM'!B2<>0;'OD SocialM'!B2*(1/'Założenia i wsp'!$B$75);0)
+IF('OD BT'!B2<>0;'OD BT'!B2*(1/'Założenia i wsp'!$B$24);0)
+IF('OD APC'!B2<>0;'OD APC'!B2*(1/'Założenia i wsp'!$B$29);0)
+IF('OD Smart Card'!B2<>0;'OD Smart Card'!B2*(1/'Założenia i wsp'!$B$51);0)
+IF('OD GPS'!B2<>0;'OD GPS'!B2*(1/'Założenia i wsp'!$B$58);0)
+IF('OD ANPR'!B2<>0;'OD ANPR'!B2*(1/'Założenia i wsp'!$B$61);0)
+IF('OD Taxi'!B2<>0;'OD Taxi'!B2*(1/'Założenia i wsp'!$B$67);0)
+IF('OD UTO'!B2<>0;'OD UTO'!B2*(1/'Założenia i wsp'!$B$34);0)
)/
(IF('OD GSM'!B2<>0;1;0))
+(IF('OD SocialM'!B2<>0;1;0))
+(IF('OD BT'!B2<>0;1;0))
```

+(IF('OD APC'!B2<>0;1;0))
+(IF('OD Smart Card'!B2<>0;1;0))
+(IF('OD GPS'!B2<>0;1;0))
+(IF('OD ANPR'!B2<>0;1;0))
+(IF('OD Taxi'!B2<>0;1;0))
+(IF('OD UTO'!B2<>0;1;0))
)

Załącznik 3

Macierz podróży symulowanego ośrodka miejskiego – Model Bazowy

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_3_Model_Bazowy



Załącznik 4

Macierz podróży – Wariant 1

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_4_Macierz_Wariant1



Załącznik 5

Macierz podróży – Wariant 2

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_5_Macierz_Wariant2



Załącznik 6

Macierz podróży – Wariant 3

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_6_Macierz_Wariant3



Załącznik 7

Macierz podróży – Wariant 4

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_7_Macierz_Wariant4



Załącznik 8

Macierz podróży – Wariant 5

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_8_Macierz_Wariant5



Załącznik 9

Macierz podróży – Wariant 6

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_9_Macierz_Wariant6



Załącznik 10

Macierz podróży – Wariant 7

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_10_Macierz_Wariant7



Załącznik 11

Macierz podróży – Wariant 8

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_11_Macierz_Wariant8



Załącznik 12

Macierz podróży – Wariant 9

Adres strony: https://www.zits.pwr.edu.pl/zawodny/Zal_12_Macierz_Wariant9

