

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

**Metodologija za usporedbu
kontekstualiziranih
polazišno-odredišnih matrica**

Rijeka, srpanj 2019.

Vjera Turk
0069064924

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

**Metodologija za usporedbu
kontekstualiziranih
polazišno-odredišnih matrica**

Mentor: izv.prof.dr.sc. Renato Filjar

Rijeka, srpanj 2019.

Vjera Turk
0069064924

Umjesto ove stranice umetnuti zadatak
za završni ili diplomski rad

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradio ovaj rad.

Rijeka, srpanj 2019.

Ime Prezime

Zahvala

Sadržaj

Popis slika	ix
Popis tablica	xi
1 Uvod	1
2 Polazišno-odredišna matrica	2
2.1 Definicija	2
2.2 Generiranje Polazišno-Odredišne Matrice	3
2.2.1 Ankete	3
2.2.2 Prebrojavanje vozila	3
2.2.3 Modeliranje prometa	4
2.2.4 Problematika i ograničenja tradicionalnih načina	4
2.2.5 Pokretna osjetila	5
2.3 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži	6
2.3.1 Anonimizacija podataka	7
2.3.2 Putovanje	7
2.3.3 Kontekst iz izvora podataka o kretanju	8
2.3.4 Sredstva kretanja	9

Sadržaj

2.4	Kontekst izveden iz vanjskih izvora	9
2.5	Primjeri iskorištavanja konteksta izvedenog iz vanjskih izvora	10
2.5.1	OpenStreetMap	10
2.5.2	<i>Points of Interest</i>	11
2.6	Geometrija prostorne podjele	12
2.7	Vremenska interpretacija Polazišno-Odredišnih Matrica	19
2.7.1	Vremenska rezolucija	19
2.8	19
2.9	Alternative izvora prostornih podataka za razvoj Polazišno-Odredišne Matrice	20
2.9.1	Združena očitavanja prijamnika za satelitsku navigaciju	20
2.9.2	Javni prijevoz i <i>pametne kartice</i> (<i>Smart Card</i> sustavi)	20
2.9.3	WiFi	21
3	Validacija Polazišno-Odredišne Matrice	22
3.1	Točnost Polazišno-Odredišne Matrice	22
3.1.1	Metrike za procjenu točnosti Polazišno-Odredišnih Matrica	23
3.2	Grafički prikaz (oblik) Polazišno-Odredišne Matrice	24
3.3	Strukturalna sličnost Polazišno-Odredišnih Matrica	25
3.3.1	MSSIM	26
4	Usporedba dviju Polazišno-Odredišnih Matrica	29
4.1	Uvjeti usporedbe	29
4.2	Kriteriji usporedbe	29
4.2.1	Definicija putovanja	30
4.2.2	Prostorna razlučivost (rezolucija)	30
4.2.3	Vremenska različivost	32

Sadržaj

4.2.4	Širina toka	34
4.2.5	Gustoća informacija - kontekst	35
4.3	Diskusija	40
5	Zaključak	41
	Bibliografija	42
	Pojmovnik	47
	Sažetak	48
A	Naslov priloga	49
A.1	Naslov sekcije	49
A.2	Naslov sekcije	49

Popis slika

2.1	Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [1] . . .	13
2.2	Jedinice lokalne samouprave za grad Shenzhen [2]	14
2.3	Voronoi dijagram 1090 ćelija baznih stanica - Shenzhen	16
2.4	Podjela na šesterokute Shenzhen - grupacija u različite boje na osnovu širine toka [3]	16
2.5	Primjer mozaika <i>krhotina</i> koje dozvoljavaju interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u <i>krhotini</i> u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih <i>težinskih faktora</i> polazišne i odredišne <i>krhotine</i> te pripadajućih ("roditeljskih") Voronoi ćelija. Sumom svih rezultata <i>krhotina</i> jedne kvadratne ćelije izračuna se konačni tok. [4]	17
2.6	Prijedlog prilagodbe administrativnih jedinica mreži kretanja - bojom su označene grupacije nastale na temelju matrice kretanja, crnom su označene postojeće granice administrativnih jedinica [5]	18
3.1	<i>Spaghetti-effect</i> - problem koji se može javiti kod grafičkog prikaza tokova matrice visoke prostorne i niske vremenske rezolucije. Na slici su prikazani jednodnevni tokovi. [6]	25

3.2	Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. Mean Structural Similarity index (MSSIM) u odnosu na referentnu Polazišno-Odredišna Matrica (POM)-u može biti korišten kao dodatna informacija (<i>goodness of fit</i> mjera) prilikom validacije POM-e [7]	28
4.1	Dvije matrice - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [8]	36
4.2	Distribucija CDR putovanja na području Santiaga, Čile. Matrica je normalizirana prema redovima (polazištima) (*) [9]	37
4.3	Prostorna podjela na anketne jedinice u gradu Santiagu. Gustoća tornjeva baznih stanica po anketnim jedinicama. Gustoća varira od 1 do 250 tornjeva po anketnoj jedinici. [9]	38
4.4	Gustoća tornjeva baznih stanica po kvadratnom kilometru za pojedine države i grad-državu Singapur. [10]	39

Popis tablica

Poglavlje 1

Uvod

Izvorišno-Odredišna ili Polazišno-Odredišna matrica *eng. Origin-destination Matrix (ODM)* ili *Trip Table* alat je koji omogućuje opis i sustavnu statističku procjenu migracija stanovništva na nekom području u zadanom prostorno-vremenskom okviru. Služi za opis grupne mobilnosti i mjerenje socio-ekonomske aktivnosti u nekoj regiji, a najčešće se koristi u prometnoj znanosti za analizu i strateško planiranje prometnog opterećenja i prometne infrastrukture.

Za razliku od tradicionalnog pristupa prebrojavanja putovanja i putnika na licu mjesta, anketiranjem ili pomoću različitih strateški postavljenih osjetila, za procjenu POM-a danas se sve više koristi statistička analiza podataka iz suvremenih informacijskih i komunikacijskih sustava (zapisi o aktivnostima u javnoj pokretnoj mreži, združena očitavanja prijamnika za satelitsku navigaciju i sl.), čime je omogućeno poboljšanje kvalitete procjene preslikavanjem POM-e na kontekst. Kada daju informaciju o kontekstu- svrsi kretanja- POM-e su dodatno obogaćene i postaju izvor informacija za primjenu i van konteksta prometa.

(...)

Poglavlje 2

Polazišno-odredišna matrica

2.1 Definicija

Polazišno-odredišna matrica sadrži broj putovanja između svakog para zona unutar nekog područja za određeni vremenski okvir (u svakom smjeru zasebno). Redovi u POM-i predstavljaju polazišta (izvore prometnog toka), a stupci odredišta. Svaki element u POM-i predstavlja broj putovanja iz pripadajuće polazišne u pripadajuću ciljnu zonu. Može se smatrati matričnim zapisom težinskog usmjerenog grafa, gdje su težinski faktori usmjerenih veza širine prometnih tokova između čvorova mreže koji predstavljaju zone na koje je promatrano područje podijeljeno.

U prometnoj znanosti postoje brojne metode estimacije POM-a. I dok se još uvijek ulaže u razvoj matematičkih modela koji imaju za cilj povećati kvalitetu POM-a generiranih tradicionalnim metodama, u novije vrijeme razvijaju se nove, moderne metode, koje imaju veliki potencijal, zbog svoje jednostavnosti u izvedbi, ažurnosti podataka i velikog uzorka stanovništva koji obuhvate. Kod pojedinih operatera radi se o oko čak 40-50% svih stanovnika neke države [11]. (...)

POM-a može biti generirana za sadašnje ili planirano prometno opterećenje. Može biti prikazana ukupno ili po pojedinim sredstvima kretanja (modulima prometnog sustava - osobna vozila, vozila javnog prijevoza, teretna vozila, pješaka itd.) i/ili različitim svrhama putovanja (putovanja na posao, putovanja kući...).

Određivanje konteksta putovanja dodatno obogaćuje POM-e jer specificira socio-

ekonomske aktivnosti koje su pokretači putovanja odnosno razlozi kretanja stanovništva. Mogućnost preciznog određivanja konteksta zavisi o metodi estimacije POM-e.

Kada kontekst (svrha) putovanja nije inicijalno poznat iz prikupljenih podataka o putovanjima, razvijene su metode koje pomoću vanjskih izvora o sadržaju procjenjuju kontekst putovanja.

2.2 Generiranje Polazišno-Odredišne Matrice

2.2.1 Ankete

Tradicionalne metode uključuju metode provođenja anketiranja. Postoji nekoliko vrsta anketa: u kućanstvima, presretanje vozila, anketiranje prijevoznih kompanija i tvrtki, tranzit na granici, turisti u hotelima i parking [12]. Ovakve metode su prije svega resursno zahtjevne stoga se ne provode često. Anketiranje putovanja u kućanstvima provodi se na područjima brojnih svjetskih metropola tek jednom u deset godina. U većini svjetskih metropola na provođenje anketa odlazi najveći postotak budžeta planiranja, u procjeni oko 7.4 milijuna dolara godišnje izdvaja se za ankete po kućanstvima. [12]

Anketiranje vozača presretanjem vozila eng. Road Side Interview (RSI) u prosjeku košta čak 10 eura po jednom zabilježenom putovanju i ukupno obuhvati 10% prometa. Ova vrsta najčešće se provodi na autocestama. [13]

Provođenje ovih metoda zahtjeva mnogo vremena i ponekad su netočne. Pokrivaju malen dio populacije te iz tog razloga mogu biti pristrane.[14]

2.2.2 Prebrojavanje vozila

Metode koje uključuju prebrojavanje vozila na ključnim čvorištima prometne infrastrukture manje su zahtjevne u odnosu na ankete i značajno smanjuju vrijeme i troškove prikupljanja podataka. Radi se o analizi nadzornih snimaka prometa, automatskom sustav za prepoznavanje i očitovanje registarskih pločica eng. Automatic

Number Plate Recognition (ANPR), osjetilima prometnog toka (radarski, magnetni, video-senzori, zvučni...) i ručnom prebrojavanju. Ove metode direktno ne govore o stvarnim polazištima i odredištima putovanja [11].

2.2.3 Modeliranje prometa

Matematičko modeliranje prometa zahtjeva veliku količinu podataka - informacija o prometnoj mreži i prometnoj potražnji. Točnost modelirane (planirane) prometne situacije ovisi o kvaliteti dostupnih informacija te kako su podaci kombinirani, koji težinski faktori su primijenjeni za različite izvore. Prometna potražnja ključna je komponenta te stoga gotovo svaki prometni model zahtjeva POM-u koja specificira prometnu potražnju između zona u prometnoj mreži. Točna POM-a osnova je za odluke u mnogim Inteligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*) (ITS) koji su zauzvrat ključni za izbor ruta u različitim sustavima navigacijskog navođenja. Neke od pretpostavki kojima se vode u modeliranju su primjerice da je ukupna dnevna potražnja podjednaka u oba smjera kod svakog para zona na području grada [15], da svi korisnici imaju saznanja o situaciji na cesti [4] te da je mobilnost općenito manja tijekom zime nego tijekom ljeta [11].

Klasičan bazni model za procjenu prometne potražnje je Gravitacijski model (eng. *Gravity model*) koji kod proračuna "privlačnosti" 2 zone uzima u obzir gustoću populacije, udaljenost i troškove putovanja. Izvedene verzije integriraju i neke druge (socio-ekonomske) faktore. [8]

2.2.4 Problematika i ograničenja tradicionalnih načina

Širenje gradova i rast stanovništva rezultirali su rastućim brojem sve ozbiljnijih prometnih zastoja u velikim gradovima diljem svijeta. Prepoznata je potreba za opsežnim strategijama upravljanja prometom kako bi se suočili s izazovima koje donosi brzo razvijajuća okolina i demografija populacije. Efektivno upravljanje i kontrola prometa mogu povećati sigurnost, kvalitetu usluge, poticati ekonomski rast i smanjiti zagađenje zraka. Zbog dinamike kojom se gradovi mijenjaju razumno je pitati se ima li smisla koristiti POM-e stare više godina za modeliranje današnjeg

prometa. Dobivanje što točnijih i jeftinijih POM-a privlači pažnju znanstvenika i van područja prometne znanosti. Alternativni izvori koje oni predlažu za promatranje obrazaca putovanja su značajno jeftiniji, ali nailaze na problem da profesija zahtijeva da se njihova vrijednost dokazuje usporedbom s tradicionalnim metodama. [13]

2.2.5 Pokretna osjetila

Osnovna ideja pokretnih osjetila je da vozila koja se kreću po prometnoj mreži i sastavni su dio prometnog toka prikupljaju i šalju podatke. Nužan zahtjev koji treba biti ispunjen kod pokretnih osjetila je poznavanje pozicije vozila u određenim vremenskim trenucima. Za to se koristi Global Navigation Satellite System (GNSS) ili se vozila identificiraju na određenim točkama. Engleski naziv za ovu metodu je Floating Car Data (FCD). Danas je FCD izvor vrijednih podataka i postaje najvažnija tehnologija prikupljanja podataka u ITS. U prikupljanju podataka sudjeluju “namjenska” vozila (taksi, distribucija robe, javni prijevoz, službena vozila i dr.) i mora biti uspostavljen način učestalog slanja podataka o položaju, a najčešće se odvija preko Global System for Mobile (Communications) (GSM), Wi-fi pa i Bluetooth tehnologije. [16]

Ukoliko pokretno osjetilo nije vozilo već pametni telefon tada se često koristi naziv Floating Phone Data (Floating Cellular Data) (FPD). FPD se odnosi na prikupljanje podataka pomoću mobilnih (pametnih) telefona neovisno o načinu određivanja položaja - korištenjem GNSS ili pomoću javne pokretne mreže.

Zaseban pristup unutar FPD je određivanje položaja unutar javne pokretne mreže (mreže baznih stanica). Kod ovog načina položaj pokretne stanice može se odrediti triangulacijom (pomoću signala iz više baznih stanica čije signal hvata pokretna stanica) ili se aproksimira područjem pokrivanja bazne stanice na koju je u trenutku očitavanja položaja pokretna stanica spojena.

2.3 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži

Kako bi pružali i naplaćivali usluge, mobilni operateri moraju prikupljati podatke o pokretnim stanicama korisnika. Pojednostavljeno, da bi se uspostavio poziv od/do odgovarajućeg uređaja, mora se biti poznato s koje bazne stanice odaslati signal. Naziv za podatke koje operateri prikupljaju kako bi naplaćivali usluge glasi Call Data Records ili Charging Data Records (CDR). Radi se o zapisima aktivnosti u mobilnoj mreži - pozivima, porukama (engl. Short Messaging Service - SMS) te započetom prijenosu mobilnih podataka. Tipičan zapis između ostalog sastoji se od: identifikacijske oznake pristupne točke radijske mreže (bazne stanice), vremenskog žiga početka telekomunikacijske aktivnosti i trajanja telekomunikacijske aktivnosti. Istraživanje veza između kretanja i društvenih veza (poznanstava) [17] od operatera je dobilo podatke i na koju baznu stanicu je poziv usmjeren.

Dodatni podatci koje operateri prate prikupljeni su pomoću sustava za ispitivanje kvalitete mreže (*Signaling data*). Uključuju *handover* podatke - podatke o transferu na drugu baznu stanicu tijekom trajanja telekomunikacijske aktivnosti, *Location Area Updates* - podatke prikupljane svakih 20 min do 2 sata o tome na koju skupinu baznih stanica je pokretna stanica spojena (primjerice na području Ile-de-France svaka LA skupina ima od oko 150 - 500 stanica), te informacije o tome kada je pokretna stanica uključena u mrežu i isključena iz mreže. Pri izradi matrica, ako su dostupni, svi ovi podatci mogu se uzeti u obzir i znatno poboljšati model.

U novije vrijeme znanstvenici predlažu metode u kojima koriste upravo anonimizirane CDR kao izvor podataka za generiranje POM-a. [8] [14] [18] iako se sama ideja spominje u istraživanjima još iz 1994., gotovo od kad je javna pokretna mreža postala široko dostupna.[8] **U ostatku rada, na mjestima gdje nije naveden izvor matrica, radi se o postupku ili matricama generiranim iz CDR.**

Zanimljivo je ovdje spomenuti europsku inicijativu *eCall* za brzu pomoć stradalima u prometnim nesrećama bilo gdje u Europskoj Uniji. *eCall* zahtjeva da u svakom novom motornom vozilu bude ugrađen uređaj koji u slučaju prometne ne-

sreće automatski obavijesti centar 112 te mu šalje informacije o aktivaciji zračnih jastuka, prikupljene podatke osjetila za procjenu jačine sudara i u dakako položaj vozila u trenutku prometne nesreće. Također treba omogućiti uspostavu poziva (GSM). Prognoza je da će se korištenjem tog sustava vrijeme odziva (dolaska službi na mjesto prometne nesreće) smanjiti za između 40% i 50%. Od travnja 2018. svi novi automobili u prodaji u Europskoj Uniji imaju ugrađen ovaj uređaj. Nakon uspješnog uspostavljanja sustava očekuje se da bi se uređaj mogao početi koristiti i u druge svrhe. [19]

2.3.1 Anonimizacija podataka

Danas uz sveprisutnost i prodornost tehnologije, podatci koje generiraju pametni telefoni omogućuju podatkovnim analitičarima razumjeti ponašanje pojedinaca u brojnim domenama, uključujući i njihov obrazac kretanja. Kao što je već spomenuto matrice je moguće generirati iz CDR koje operateri prikupljaju u svrhu pružanja i naplaćivanja usluge. Za ovu metodu stoga ne postoji potreba za dodatnim uređajima, sustavima ili protokolima centralnog prikupljanja položaja kao što je to u slučaju kada se položaj određuje pomoću GNSS prijamnika.

Situacija je takva da telekomunikacije kompanije, ali i ostali tehnološki subjekti, moraju voditi računa o privatnosti pojedinaca i ukoliko neki podaci jesu korišteni za razvoj POM-a, takvi podaci najprije moraju biti striktno anonimizirani te ne smije postojati mogućnost reengineeringa i povratka na izvorni identitet. U pravilu to znači povezivanje sa trećim stranama u poslovnom procesu, koje iz nekog razloga dozvoljavaju pristup podacima. Podaci prije pristupa moraju biti anonimizirani tako da oni mogu biti korišteni za druge svrhe bez razotkrivanja pojedinaca, u skladu s postavkama General Data Protection Regulation (GDPR)

2.3.2 Putovanje

Priroda ovog pristupa rudarenju podataka za estimaciju POM-a zahtjeva da prvo definiramo što točno smatramo putovanjem, odnosno polazištem i odredištem putovanja. Zaustavljanje se definira kao skup minimalno n događaja (aktivnosti) koji

su zabilježeni na istoj poziciji unutar definiranog minimalnog vremenskog perioda. Time se osigurava da se ne radi o tranzitnoj već o namjerenoj poziciji zaustavljanja (i suzbija se bilježenje lažnog kretanje kod privremenog spajanja na susjednu baznu stanicu). Algoritam u nizu aktivnosti svakog pojedinca traži takva zaustavljanja. Dobiveni niz kontinuiranog zaustavljanja na različitim položajima otkriva putanju kretanja. Algoritam izvlači uzastopne parove zaustavljanja te, prema dodatno definiranim kriterijima dozvoljenog minimalnog i maksimalnog vremenskog razmaka između polaska i dolaska na odredište, ranija ćelija zaustavljanja postaje polazište, a kasnija odredište putovanja.

Kod izrade tzv. tranzitnih matrica, u kojima je naglasak osim na broj putovanja stavljen i na vremensku domenu, skupovi svih putovanja u pripadajućem vremenskom okviru sačinjavaju matricu za taj vremenski okvir, te je rezultat generiranja skup matrica.

2.3.3 Kontekst iz izvora podataka o kretanju

Kada su dostupni zapisi prikupljeni dulje vremensko razdoblje, tjednima ili mjesecima često se putovanja odnosno matrice mogu definirati na drugačiji način. Pojednostavljeno, analizom zapisa može se utvrditi na kojem položaju (u kojoj ćeliji) neki pojedinci pretežito borave u tipičnom radnom vremenu te u kojoj tokom noći. Prva ćelija za njihova putovanja postaje *Work* destinacija, druga *Home*, a sve preostale destinacije svrstaju se u zasebnu kategoriju *Other*. Potom se generiraju matrice koje uključuju samo putovanja iz jednog tipa destinacije u drugi npr. Home-Work (HW) i Work-Home (WH).[4][9][6][20][21]. Jedan od razloga je da su takve matrice usporedive s onima dobivenim iz anketa stanovništva s istom podjelom putovanja.

CDR zapisi u trajanju od 6 mjeseci bili su dovoljni za konačnu točnu procjenu položaja čak 98% kućanstava nešto više od 39 tisuća najaktivnijih korisnika na području Pariza i Chicaga, izdvojenih iz ukupnog skupa korisnika (milijuni korisnika). U podacima se ispitalo gdje svaki od korisnika provodi najviše vremena između 0:00 i 6:00 sati.[22] Dvije godine kasnije bitno drugačiju definiciju daje [21] gdje definira kućanstvo kao mjesto gdje korisnik provodi najviše vremena između 20:00 i 7:00 na radne dane, a radno mjesto kao mjesto gdje korisnik provodi najviše vre-

mena, barem 3 puta tjedno, između 7:00 i 20:00 dok [8] koristi intervale od 22:00 do 7:00 te od 9:00 do 16:00. Ostale lokacije svrstava u kategoriju *Ostalo eng. Other*. Bitno je naglasiti da kod ovih vrsta matrica najčešće naglasak nije stavljen na dobru vremensku rezoluciju kao što je to slučaj kod tranzitnih. [8] Duža razdoblja prikupljanja podataka omogućuju uočavanje uzoraka kretanja i motiva (teorija mreža). [22]

2.3.4 Sredstva kretanja

Isključivo CDR zapisi ne osiguravaju rezoluciju pogodnu za određivanje sredstva kretanja iako su razvijene metodologije [21]. Iznimno se mogu izdvojiti bazen stanice dedicerane pružanju usluge isključivo u podzemnoj željeznici.

Teralytics je jedna od mladih kompanija koja nudi uslugu modeliranja prometa iz telekomunikacijskih podataka. Između ostalog prema brzini kojom pokretna stanica (ili skupina njih istovremeno) mijenja položaj donose zaključak o dominantnom sredstvu kretanja (pješač, automobil, javni prijevoz, avion). Tajna njihovog uspjeha je da osim CDR koriste i ranije spomenute dodatne podatke koji se prikupljaju. [23]

Za razliku od tradicionalnih modela prebrojavanja vozila, telekomunikacijski zapisi obuhvaćaju sva sredstva kretanja kao i svu prometnu infrastrukturu (kao i područja bez razvijene prometne infrastrukture), ne samo onu gdje se provode intervjui ili gdje su postavljena osjetila za prebrojavanje prometa. Uklanja potrebu za različitim pristup prebrojavanja za pojedina sredstva kretanja.

2.4 Kontekst izveden iz vanjskih izvora

Neka istraživanja spojila su ljudsku mobilnost s namjenom zemljišta (engl. *land use*) ili točkama interesa engl. *Points of Interest (POIs)* kako bi grupirala urbane predjele prema njihovoj funkciji ili upotrebi. Izvori podataka o kretanju u tim istraživanjima variraju od Global Positioning System (GPS) putanja, položaja ukrcaja i iskrcaja taxi vozila, zapisa koje generiraju pametne kartice u javnom prijevozu, podatci o dijeljenim lokacijama na društvenim mrežama (engl. *check-in*) i zapisi o

telekomunikacijskim aktivnostima (CDR)[14].

Izvori podataka o kontekstu također variraju od istraživanja do istraživanja, a radi o opisima objekata u prostoru i relacija među tim objektima. Korišteni podaci su primjerice broj objekata, površina koju oni zauzimaju te, za kontekst najbitnija, socio-ekonomska svrha (funkcija) pojedinog objekta.

2.5 Primjeri iskorištavanja konteksta izvedenog iz vanjskih izvora

2.5.1 OpenStreetMap

OpenStreetMap (OSM) je svjetski raširen projekt koji kreira i pruža slobodne zemljopisne podatke (zemljovide gradova i naselja) temeljen na volonterskom doprinosu zajednice. Pruža detaljne i ažurne digitalne zemljovide kompatibilne s Geographic Information System (GIS) aplikacijama. Započet je prije 15 godina, u Ujedinjenom Kraljevstvu, kao odgovor na tehnička ili pravna ograničenja postojećih "slobodnih" baza prostornih podataka kao što je GoogleMaps. Zajednički doprinos i međusobna kontrola unosa održavaju kartu na visokoj razini kvalitete i točnosti.

Struktura osnovnih elemenata na toj rasterskoj karti je hijerarhijska. Element u hijerarhiji može biti: čvor (engl. *node*), put (engl. *way*) ili relacija (engl. *relation*). Čvor je jedinstvena točka u prostoru (sa identifikacijskom oznakom, zemljopisnom širinom i dužinom) koja najčešće predstavlja fizički objekt (zgrada, dio ceste...) te sadrži jedan ili više *ključ=vrijednost* žigova (engl. *key=value tag*) koji definiraju razne značajke objekta. Put je uređena lista 2 do 2,000 čvorova, koja također može biti opisana žigovima. Relacija je uređena lista čvorova, puteva i/ili relacija. Definira logičku ili geografsku povezanost članova. Opisni element nazivamo žigom (engl. *tag*), a njegova funkcija je opisati značajke elementa uz koji je vezan. Primjeri žigova vezanih za čvorove: *office=company*, *building=residential*, *building=hotel*, *building=church*, *leisure=sports_centre*, *amenity=school*. Primjer žiga vezanog uz put *highway=residential*.

Model raspodjele toka

Inovativni pristup kontekstualizaciji toka korištenjem OpenStreetMap baze predstavljen je u radu [24]. Na osnovu pripadajućih opisnih podataka, prostorni objekti na području interesa kategorizirani su prema tipu socio-ekonomske aktivnosti u koju su uključeni. Definirano je ukupno 6 kategorija: Dom (Home), Posao (Work), Zdravlje (Health), Edukacija (Education), Zabava (Leisure) i Ostalo (Other). Razvijen je vjerojatnosni model koji, na osnovu broja objekata pojedine kategorije u odredišnoj ćeliji i promatranog vremenskog okvira, ukupni ulazni tok u odredišnu ćeliju dijeli na 6 tokova usmjerenih prema objektima tih kategorija. Iz jedne matrice tako se dobije 6 matrica, po jedna za svaku od kategorija. Korištene su matrice dobivene iz javno dostupnih, anonimiziranih telekomunikacijskih zapisa na području kineskog grada Shenzhena.[25][26] **Definiranjem konteksta putovanja na ovaj način sačuvana je mogućnost postizanja dobre rezolucije u vremenu, a istovremeno je dan kontekst koji uključuje tipične *Home* i *Work* kategorije, te umjesto općenite kategorije *Other* precizira 4 kategorije.**

** Daje podatak i o internim putovanjima (Petra) **

Izvor infrastrukture

Jedan od načina na koji OSM doprinosi istraživanjima u ovom području je kao izvor prometne infrastrukture i pripadajućih meta podataka (kolnik, mostovi, ograničenja brzine...). Raspodjela toka po prometnoj infrastrukturi nadilazi područje ovog rada no jest dio problematike modeliranja prometa. [21] Još jedna moguća uloga spominje u potpoglavlju 2.6 koje opisuje interpolaciju Voronoi ćelija.

2.5.2 *Points of Interest*

U jednom od svojih radova grupa autora kvantizira vezu između ukupno 23 namjenske kategorije POIs (npr. tvornice, bolnice, javne škole, religijski objekti, hoteli, knjižnice, sportski objekti...) i onime što nazivaju 3 tipa privlačenja. Dolaze do zaključka da postoje 3 tipa privlačenja te da svaki ima karakterističan tok (gustoću ukupnog toka, udaljenost i disperziju polazišta tokova usmjerenih prema objektu tog

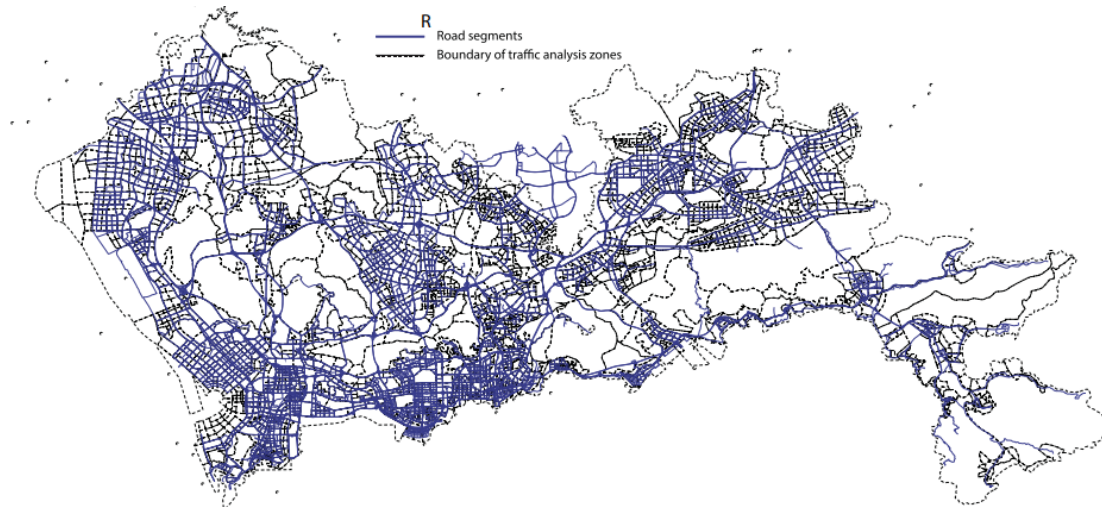
tipa privlačenja). Svaku od kategorija karakterizira jedan od ta 3 tipa privlačenja. [14] U radu koriste podatke iz službene baze s 12,000 POIs i CDR zapisima iz razdoblja od mjesec dana. Cilj projekta bio je ispitati mobilnost na području grada Riyadha, u Saudijskoj Arabiji za planiranje izgradnje podzemne željeznice. Također njihovi rezultati mogu se primijeniti u planiranju pozicioniranja novih objekata u gradu.

2.6 Geometrija prostorne podjele

Konačna geometrija prostorne podjele alternativnih izvora uglavnom ovisi o prostornoj podjeli *grand truth* matrice, matrice koja se uzima kao najvjerniji prikaz stvarnog kretanja te u odnosu na koju se procjenjuje točnost procijenjene matrice. Obično se radi o jedinici prostorne podjele naziva Traffic Analysis Zone (TAZ) kod analize prometa, o anketnim jedinicama ili jedinicama samouprave. Neprihvaćenija prostorna podjela kod matrica izvedenih iz CDR je ona na Voronoi ćelije.

Traffic Analysis Zone (TAZ)

Traffic Analysis Zone (TAZ) je strani naziv za jedinicu prostorne podjele (ćeliju) kod analize prometa. Kod programskih podrška za analizu prometa veličina jedinica varira i nije uniformna. Podjela se može temeljiti na populaciji ili broju zaposlenih na nekom prostoru, primjerice svaka ćelija pokriva prostor na kojem obitava otprilike tri tisuće ljudi (propisanih preporučenih vrijednosti nema), ili podjela u kojoj ćelije generiraju podjednak broj izlaznih putovanja (jednaka širina izlaznih tokova) [4]. U hijerarhijskoj podjeli kod ankete stanovništva općenito postoje manje i veće jedinice podjele, a nema podjele koja bi precizno odgovarala TAZ. Do 2000. godine zastupljen je bio jedan model predviđanja prometa, a veličina TAZ bila je uvijek približno jednaka anketnom bloku od 600 do 3000 ljudi. Pojavom kompleksnijih modela predviđanja prometa (i modela upotrebe zemljišta) počinju se koristiti sve manje prostorne jedinice jer se povećava i razina detalja koju noviji modeli mogu ostvariti. Jedinica veća od TAZ uvedena 2010. naziva se Traffic Analysis District (TAD) i obuhvaća oko 20,000 stanovnika.[27]



Slika 2.1 Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [1]

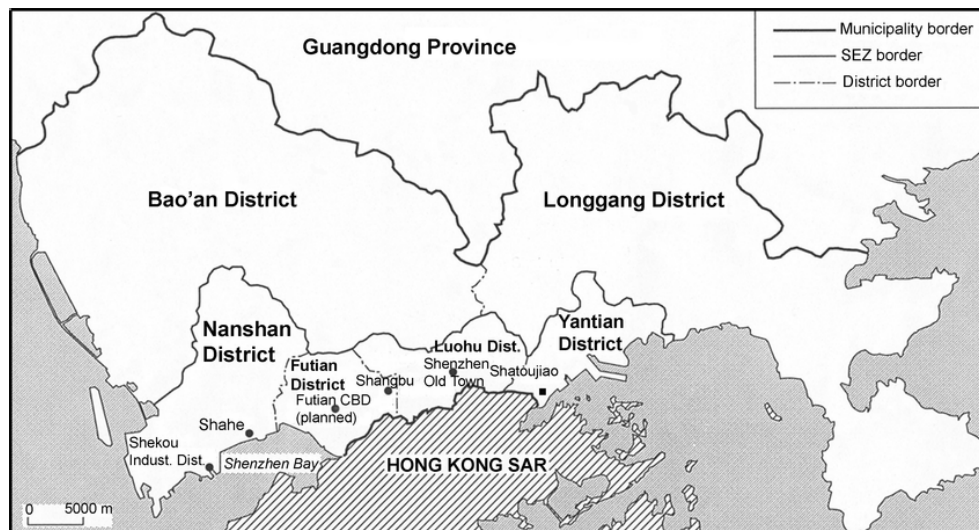
Prema podacima korištenim u istraživanju iz 2018. [28] Shenzhen je podijeljen na 996 TAZ od kojih 844 zone imaju barem jednu autobusnu ili stanicu podzemne željeznice. Godinu dana ranije objavljeno istraživanje spominje više zona (Vidi sliku 2.1).

Kod analize na području cijele države [4] koristi *k-sredina* algoritam za grupiranje baznih stanica u broj grupa jednak broju TAZ na području Senegala. [14] na području grada Riaydha agregira bazne stanice u 1492 TAZ-e.

Jedinice samouprave i anketne jedinice

[29] ne definira unaprijed broj grupa već za grupaciju koristi *mean-shift* algoritam koji sam određuje broj grupa na području Švedske, potom koristeći OSM grupe veže uz imena gradova i mjesta. U istraživanju [17] bazne stanice na području Kolumbije grupiraju po jedinicama samouprave (eng. *municipalities*). Isti postupak je korišten i u istraživanju na području Madrida, kako bi matrice u konačnici bile usporedive s onima dobivenim iz anketa [18]. Prikaz jedinica samouprave za grad Shenzhen možemo vidjeti na slici 2.2. Podjela na gradske četvrti obično je relativno uniformna

[14].



Slika 2.2 Jedinice lokalne samouprave za grad Shenzhen [2]

Voronoi teselacija

Najfinija rezolucija koju je moguće postići kada je riječ o matricama dobivenim iz CDR postiže se podjelom prema Voronoievoj teselaciji.

Općenito, u ravnini, za svaku točku iz skupa točaka postoji odgovarajuća regija koju čine sve točke ravnine bliže toj točki negoli ikojoj drugoj iz skupa. Te regije se nazivaju Voronoievim ćelijama. Svaka takva ćelija dobiva se presjecištem pola prostora između susjednih točaka skupa, pa stoga čini konveksni poligon. Dijelovi Voronoiova dijagrama su sve točke u ravnini koje su ekvidistantne prema dva najbliža položaja (granice poligona). Voronoiovi vrhovi su točke ekvidistantne tri (ili više) položaja.

Točke skupa su položaji baznih stanice, a ćelije oko njih aproksimacija površine koju ta bazna stanica pokriva. U stvarnosti se te površine prekrivanja preklapaju jer područje prekrivanja nije sasvim fiksno i ovisi o trenutnoj aktivnosti svake bazne stanice [11]. (Vidi sliku 2.3)

Poglavlje 2. Polazišno-odredišna matrica

Ponekad operateri u CDR zapisima ne otkrivaju položaje baznih stanica već samo njihov ID. Postoji nekoliko javno dostupnih baza kao što je OpenCellID [30] nastalih volonterskim doprinosom koje pohranjuju podatke o procjenama položaja baznih stanica, snazi signala i sl. Položaji se prikupljaju pomoću aplikacije instalirane na mobilnom uređaju, registriranjem GPS pozicije i identifikacijske oznake bazne stanice na koju je uređaj trenutno spojen. Dakle, ove baze ne sadrže "najtočnije" položaje baznih stanica već usrednjene položaje mobilnih uređaja koji su u nekom trenutku bili spojeni na tu baznu stanicu. Može se reći da taj procijenjeni položaj za potrebe izrada matrica čak i više odgovara od stvarnog položaja. [29]

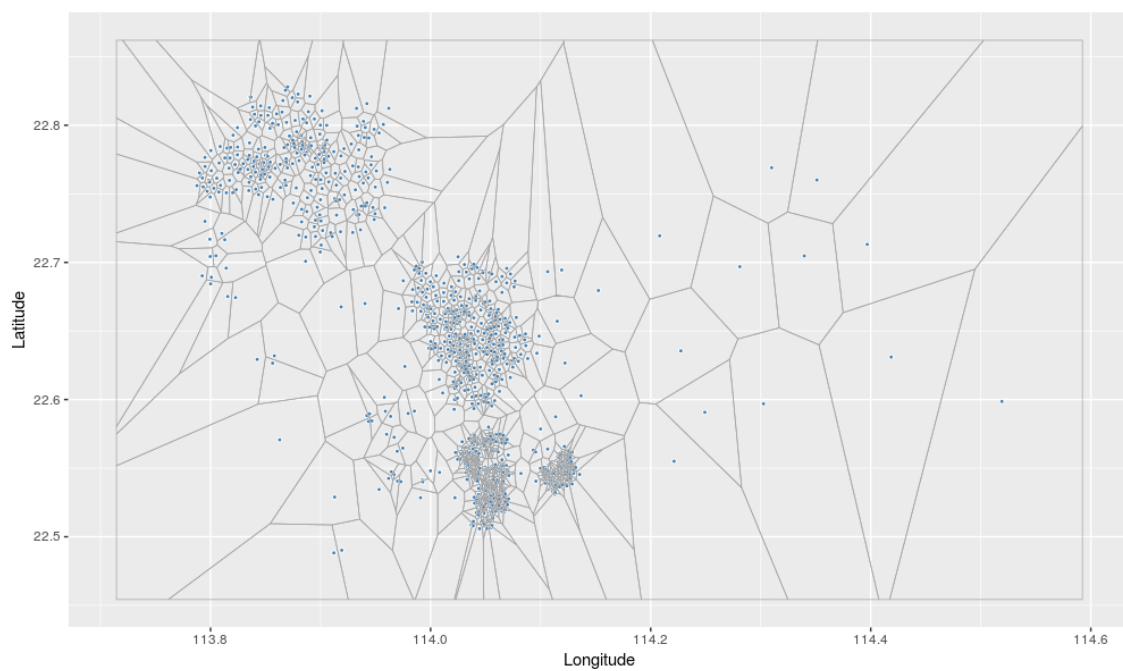
Gustoća Voronoi ćelija snažno je korelirana s gustoćom stanovništva te stoga bolju prostornu rezoluciju ima na područjima veće gustoće populacije. Mjerna nesigurnost određivanja položaja tako iznosi od približno stotinu metara u gusto naseljenom, do nekoliko kilometara u ruralnom području.[20] No Gundlegard [8] smatra da ne-uniformnost ćelija nije problem, već preklapanje područja pokrivanja u stvarnosti, posljedica čega su stalne oscilacije - uređaj mijenja baznu stanicu na koju je spojen iako svo vrijeme zadržava isti položaj. Jedan od načina na koji se može ublažiti efekt koji oscilacija ima na definiranje putovanja je odrediti minimalnu udaljenost putovanja ili čak potpuno zanemariti putovanja u susjedne ćelije. Važno je kod definiranja putovanja uzeti u obzir postojanje oscilacija.

Kvadratna mreža i mreža šesterokuta

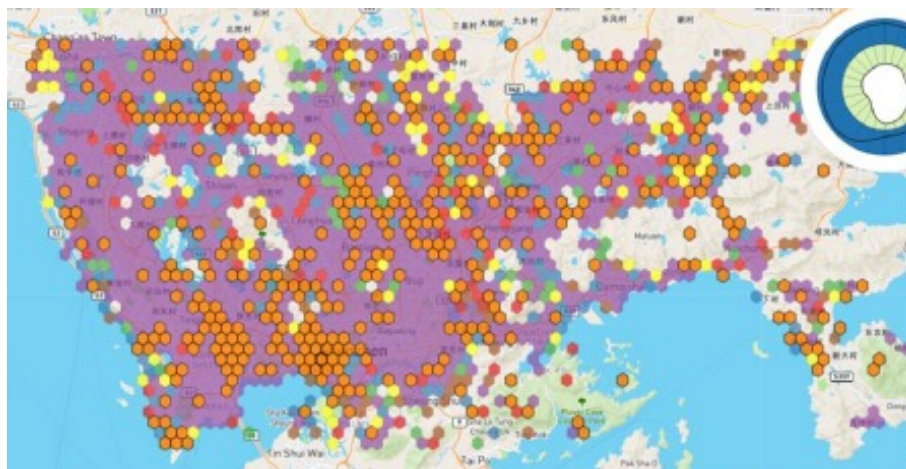
Uniformna podjela na kvadratne blokove od 1 kvadratni kilometar spominje se kao dobra praksa u jednom britanskom priručniku za izradu matrica iz CDR [4].

Područje pokrivanja bazne stanice na grafikama se standardno prikazuje kao šesterokut (heksagon). Terralytics u praksi [23] između ostalih podjela koristi i šesterokute koje po potrebi ugnježđuje i tako mijenja rezoluciju u kojoj promatra gibanja. Tvrde da su se u nekim segmentima pokazali i 13% bolji od kvadratne mreže. Prikaz šesterokuta za grad Shenzhen možemo vidjeti na slici 2.4

Poglavlje 2. Polazišno-odredišna matrica



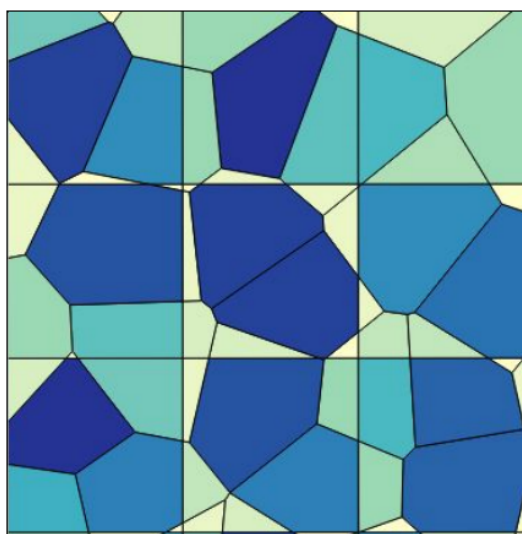
Slika 2.3 Voronoi dijagram 1090 ćelija baznih stanica - Shenzhen



Slika 2.4 Podjela na šesterokute Shenzhen - grupacija u različite boje na osnovu širine toka [3]

Interpolacija i usporedba

Postupak konverzije broja putovanja iz prostorne podjele na Voronoi ćelije u drugu prostornu podjelu, npr. uniformnu kvadratnu mrežu (ćelije $1km^2$), opisan u [4]. U postupku određivanja postotka ukupnog toka Voronoi ćelije koji će se dodijeliti novoj kvadratnoj ćeliji predlaže se uzeti u obzir: površinu preklapanja tih ćelija, broj zgrada ili ukupnu površinu zgrada (uključujući katove) na području preklapanja tih ćelija. Na slici 2.5 je prikazan treći oblik interpolacije gdje su korišteni podaci o ukupnoj kvadratnoj površini zgrada.

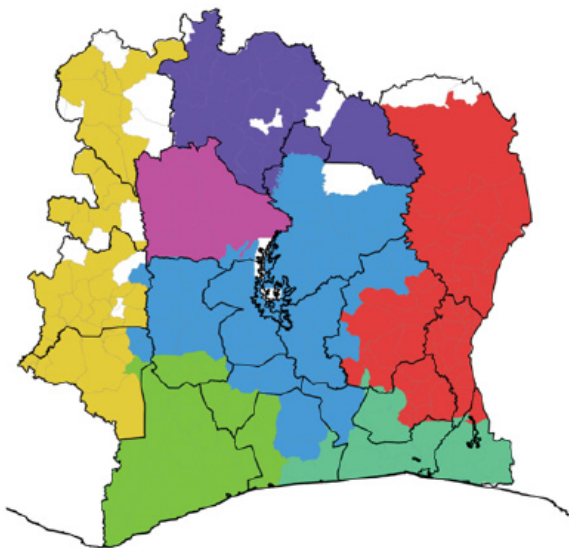


Slika 2.5 Primjer mozaika *krhotina* koje dozvoljavaju interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u *krhotini* u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih *težinskih faktora* polazišne i odredišne *krhotine* te pripadajućih ("roditeljskih") Voronoi ćelija. Sumom svih rezultata *krhotina* jedne kvadratne ćelije izračuna se konačni tok. [4]

(...)[21] [6]

Redefiniranje prostorne podjele na temelju matrica

U sklopu projekta Data for Development (D4D) francuski operater *Orange* ustupio je na korištenje CDR višemjesečne zapise s područja Obale Bjelokosti. Različita istraživanja u ovom području rađena su upravo na tim podacima. Jedno od njih [17] ističe da je dobar broj istraživanja dokazao da administrativne regije u zemlji faktor koji definira ljudsku mobilnost i interakcije na području države. Da bi testirali tu tvrdnju oni, nakon što generiraju matricu kretanja, ćelije prostorne podjele grupiraju prema *community detection* algoritmu. Iako se dio grupa vrlo lijepo preklapa sa postojećim administrativnim jedinicama, one administrativne jedinice u kojima se nalaze veliki gradovi, grupacije nastale na temelju mobilnosti bitno su drugačije. Zaključak je da potencijalno postoji potreba da se granice administrativnih jedinica izmjene (Slika 2.6)



Slika 2.6 Prijedlog prilagodbe administrativnih jedinica mreži kretanja - bojom su označene grupacije nastale na temelju matrice kretanja, crnom su označene postojeće granice administrativnih jedinica [5]

****Ideja: Dinamička prostorna podjela, drugačija prostorna podjela za doba dana/godine****

[22] -to remove towers which are only used during travel, all towers which are less

frequently visited than a certain threshold are ignored; in this study, less than 0.5 per cent during the entire observational period;

— the day, starting at 03.00, is divided into 48, 30-min slots for each of the 154 days;

to eliminate signal transitions between neighbouring towers, these towers are merged **for one day**, if more than three back and forward transitions between them are recorded during a single day;

to identify an activity location, only the most frequently observed location during each time slot is assigned as an activity location for this time slot;

2.7 Vremenska interpretacija Polazišno-Odredišnih Matrica

2.7.1 Vremenska rezolucija

(...)

-tranzitna matrica vs. matrica bez vremenske dimenzije (odnosno ima samo jedan vremenski okvir)

-radni dan, vikend, petak drugačiji → sposobnost CDR da uhvate razlike

-sposobnost da uhvati sezonske promjene

2.8

former section **Dobre prakse u generiranju Polazišno-Odredišnih Matrica iz Charging Data Records**

-anonimizacija i GDPR

- onemogućiti reengineering: k-anonymization 5 - 10 [23] i in rural area [4]

-Grad, Država

A potential is also recognized for a planet-scale mobility measurement through opening to researchers and combining the different big dataset carrying location infor-

mation. [5] [31]

-CDR- POM u zemljama u razvoju (NE)
-Penetracija od 82%, 9X% → pronaći radove
2008. Francuska 80% starijih od 12 godina [11]

-Mobilni uređaji iz godine u godinu zastupljeni su kod sve većeg postotka stanovništvu, ne samo u razvijenim zemljama već i u zemljama u razvoju.

** slobodno spomenem da se negdje ne smiju koristiti mobiteli u autima **
netko:negdje je rekao da bi imalo smisla pome iz CDR da se smiju koristiti mobiteli
prilikom vožnje

2.9 Alternative izvora prostornih podataka za razvoj Polazišno-Odredišne Matrice

2.9.1 Združena očitavanja prijamnika za satelitsku navigaciju

(...)

komercijalna vozila, Taxi

Jedno istraživanje matrica iz CDR je eliminiralo zapise onih korisnika koju su radili dulja "neprekidna putovanja" pod napomenom da se radi o taksistima i sl. te da ih neće uzeti u obzir.

2.9.2 Javni prijevoz i *pametne kartice (Smart Card sustavi)*

(..)

82% putovanja javnim prijevozom naprave korisnici javnog prijevoza sa pametnim karticama. [32]

- [33]

2.9.3 WiFi

(...)

- ITrack.CTrack,VTrack

Poglavlje 3

Validacija Polazišno-Odredišne Matrice

Postupak validacije je objektivna procjena kvalitete, vrijednosti tog objekta koji se validira. Validacija POM-e je proces koji ima za cilj procjenu obilježja kvalitete te POM-e i znači identifikaciju obilježja kvalitete POM-e, njenog sadržaja, njene smislenosti a izražava se slijedećim/ indikatorima kao što su: točnost, prostorno obuhvaćanje, zrnatost.

Pojedinačno navesti *metrike* za određivanje vrijednosti svakog od tih indikatora

3.1 Točnost Polazišno-Odredišne Matrice

Točnost procijenjenih matrica gotovo uvijek se definira u odnosu na referentnu matricu (eng. *grand truth matrix*) koja je dobivena tradicionalnim postupcima (anketiranje i/ili prebrojavanje vozila). Statističke mjere kvantiziraju razliku procijenjenih i “istinitih” vrijednosti, ako su nam one poznate.

Često se u literaturi (jednoznačno) koriste pojmovi *točnost*, *pouzdanost* i *kvaliteta*. Gotovo uvijek radi se o mjerama koje opisuju razinu sličnosti odnosno razlike (greška) s referentnom matricom. **Tradicionalni način validacije POM-e svodi se na određivanje njene točnosti u odnosu na *grand truth* matricu.**

3.1.1 Metrike za procjenu točnosti Polazišno-Odredišnih Matrica

Za procjenu kvalitete matrica dobivenih isključivo anketranjem u radu [12] korištena je mjera Mean Absolute Percentage Error (MAPE), te je prikazano da se zadovoljavajuća razina kvalitete takvih matrica postiže tek ako uzorak obuhvaća 50% populacije. Istaknuta je važnost korištenja dodatnih izvora za izradu matrica.

U radu [34] navedene su statističke mjere Relative Error (RE), Total Demand Deviation (TDD), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) te Maximum Possible Relative Error (MPRE) i Travel Demand Scale (TDS) koji procjenjuju kvalitetu neovisno o referentnoj matrici (no MPRE ne dopušta pogreške u prebrojavanju prometa, dok TDS ovisi o topologiji mreže i odabiru ruta). [7]

U [18] korišten je *Pearsonov koeficijent korelacije* - r da bi se utvrdila sličnost svakog retka matrice dobivene iz CDR s retkom referentne (izlazni tok iz svake polazišne ćelije). Isti postupak korišten je za kontekstualizirane HW i WH matrice dobivene iz CDR u usporedbi s referentnim matricama dobivenim anketiranjem.

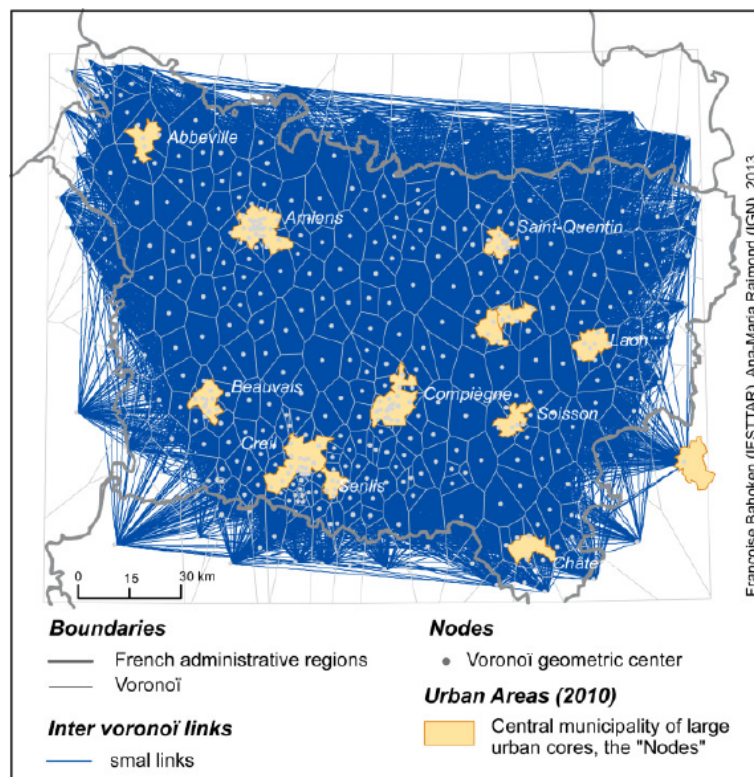
Travassoli u svome radu [32] navodi nekoliko uobičajeno korištenih mjera - R^2 , Geoffrey E. Havers statistics (GEH), Root Mean Squared Error percentage %RMSE te uvodi novu mjeru Eigenvalue-based measure (EBM) (temeljenu na svojstvenim vrijednostima matrica) i procjenjuje pouzdanost matrice dobivene iz sustava automatskog prikupljanja podataka u javnom prijevozu (autobus, vlak i trajekt). Spominje i Wasserstein metric, mjeru koja se razlikuje po tome da ne uspoređuje samo vrijednosti parova istih ćelija (elementwise).

Spearmanov koeficijent korelacije ranga korišten je u [9] za procjenu sličnosti matrica dobivenih iz CDR sa tada aktualnim matricama dobivenim anketiranjem.

3.2 Grafički prikaz (oblik) Polazišno-Odredišne Matrice

Na slici 3.1 može se vidjeti jedan oblik grafičkog prikaza matrice, odnosno njenih tokova. Svaki par centara ćelija spojen je ravnom linijom čija debljina odgovara širini toka između te dvije linije. Međutim, riječ je o primjeru kada ovaj oblik prikaza nikako nije primjeren, zbog visoke rezolucije (velikog broja ćelija) kao i velikog broja zabilježenih tokova. Dodatno, da bi se u ovom obliku prikazali svi podaci iz matrice, potrebno je prikazati jednu sliku za ulazne te jednu za sve izlazne tokove.

Postoji i još jedan često korišten grafički oblik matrice. Vrijednosti u matrici mogu se normalizirati na opseg vrijednosti koji je moguće prikazati u obliku slike tako da svaka vrijednost predstavlja vrijednost piksela. Koriste se različite palete, od raspona 0-100 koji predlažu [25], prikazom nijansama jedne boje 0-255 engl. *grayscale* [35], paletom nijansa 2 boje, ili kojom drugom proizvoljnom paletom s više boja (vidi matrice na slici 4.1).



Slika 3.1 *Spaghetti-effect* - problem koji se može javiti kod grafičkog prikaza tokova matrice visoke prostorne i niske vremenske rezolucije. Na slici su prikazani jednodnevni tokovi. [6]

3.3 Strukturalna sličnost Polazišno-Odredišnih Matrica

Dosada spomenute mjere neće uhvatiti strukturalnu sličnost matrica. Nekolicina autora ističe važnost strukturalne sličnosti s referentnom matricom kao važnu mjeru kvalitete matrice jer visoka razina strukturalne sličnosti može biti prisutna i kod matrica s manjom razinom sličnosti prema statističkim mjerama. Također, strukturalna sličnost je (vizualno) vidljiva u grafičkom obliku matrice. **Dobro odgovara ljudskoj vizualnoj percepciji sličnosti slike.**

3.3.1 MSSIM

MSSIM dolazi iz područja računalne obrade slike i koristi se kao mjera usporedbe digitalnih slika (*eng. measure of comparison*). U prometu ideja o korištenju MSSIM za mjerenje sličnosti matrica se prvi puta spominje i demonstrira na simuliranim matricama dobivenim iz referentne matrice dodavanjem šuma. [7]

Informaciju o strukturi slike definiramo kao attribute slike koji predstavljaju strukturu objekata na sceni, i neovisni su o prosječnom osvjetljenju i kontrastu. Jer osvjetljenje i kontrast mogu znatno varirati na sceni, moraju se u obzir uzeti samo njihove lokalne vrijednosti.

Structural Similarity index (SSIM) bazira se na degradaciji strukturalnih informacija na jednoj slici u usporedbi s drugom (referentnom) slikom. SSIM se računa za svaki kvadratni blok veličine N elemenata na način da se jezgra (da bi obuhvatila novi blok) pomiče ćeliju po ćeliju dok ne prođe preko cijele slike. MSSIM je srednja vrijednost svih SSIM.

Osnovni

Neka su X i Y matrice koje uspoređujemo a $x = \{x_n | x = 1, 2, \dots, N\}$ i $y = \{y_n | y = 1, 2, \dots, N\}$ parovi vrijednosti kvadratnih prozora veličine jezgre na istim pozicijama u X i Y ; $SSIM$ je određen prosječnim vrijednostima μ_x i μ_y sa standardnim devijacijama σ_x i σ_y i kovarijancom σ_{xy}

$$l(x, y) = (2\mu_x\mu_y + C1)/(\mu_x^2 + \mu_y^2 + C1) \quad (3.1)$$

$$c(x, y) = (2\sigma_x\sigma_y + C2)/(\sigma_x^2 + \sigma_y^2 + C2) \quad (3.2)$$

$$s(x, y) = (\sigma_{xy} + C3)/(\sigma_x\sigma_y + C3) \quad (3.3)$$

$l(x, y)$ opisuje razliku u osvjetljenju, $c(x, y)$ razliku u kontrastu, a $s(x, y)$ razliku u strukturi između x i y . $C1$, $C2$ i $C3$ su konstante uvedene da se izbjegne "nestabilnost" kada su nazivnici bliski 0. Opća forma $SSIM$ definira se kao

$$SSIM(x, y) = [l(x, y)^\alpha][c(x, y)^\beta][s(x, y)^\gamma] \quad (3.4)$$

Poglavlje 3. Validacija Polazišno-Odredišne Matrice

gdje su α , β i γ parametri relativne važnosti svake komponente. Za $SSIM$ vrijedi slijedeće:

$$SSIM(x, y) \leq 1 \quad (3.5a)$$

$$SSIM(x, y) = SSIM(y, x) \quad (3.5b)$$

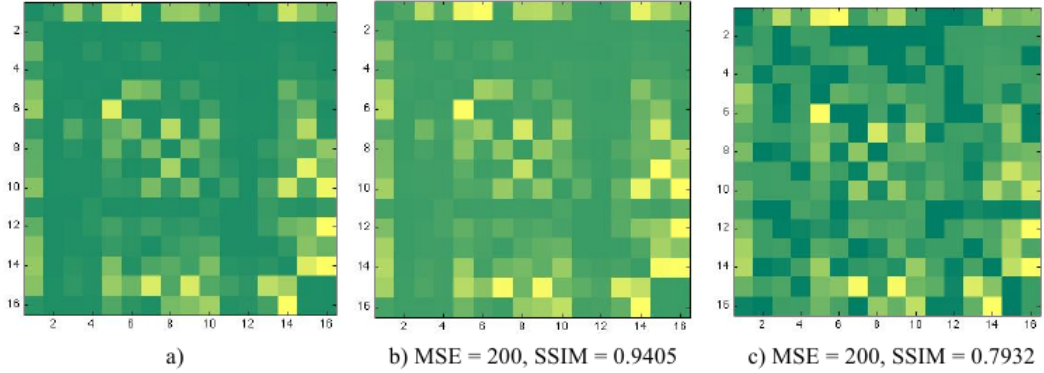
$$SSIM(x, y) = 1 \iff x = y \quad (3.5c)$$

$$MSSIM(X, Y) = \frac{1}{M} \sum_{m=1}^M SSIM(x_m, y_m) \quad (3.6)$$

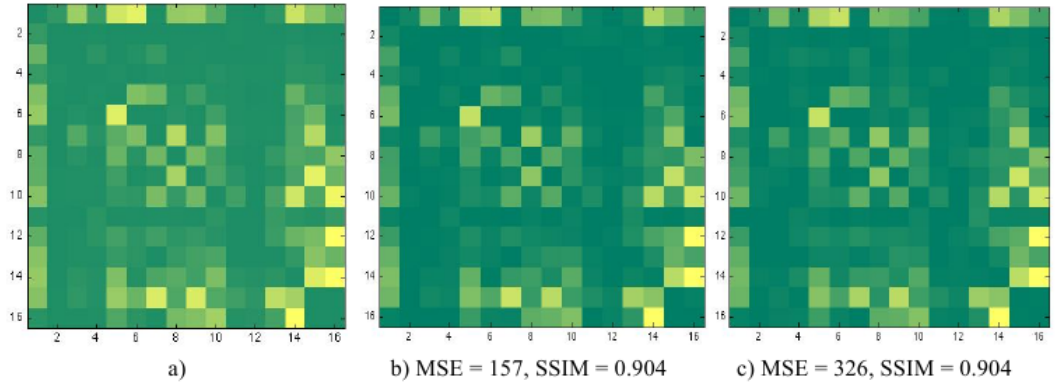
Poboljšani

Nekoliko godina nakon prvog spominjanja $MSSIM$ kao metrike usporedbe POM-a [35] se dotiče 3 problema postavljajući pitanja : Koliko treba biti veliki blok? Kako usporediti "guste" i "rijetke" matrice? Koja je prihvatljiva vrijednost $MSSIM$? Autori definiraju poboljšani model koji nazivaju $4D-MSSIM$ gdje u izračun dodaju stvarne euklidske udaljenosti prostornih ćelija matrice.

Poglavlje 3. Validacija Polazišno-Odredišne Matrice



- (a) U odnosu na referentnu *grand truth* POM-u a), vrijednost metrike točnosti MSE jednak je za obje matrice b) i c), dok se indeks strukturalne sličnosti znatno razlikuje, gdje POM-a b) ima veću strukturalnu sličnost s a) nego što ima c) (3.5a). Lako je na matrici c) uočiti narušenost strukturalne sličnosti.



- (b) U odnosu na referentnu POM-u a), vrijednost metrike strukturalne sličnosti iznosi 0.904 za obje matrice b) i c), dok se MSE razlikuje (veći MSE znači veću razliku). Obje matrice dobivene su dodavanjem Gaussianovog šuma na referentnu matricu.

Slika 3.2 Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. MSSIM u odnosu na referentnu POM-u može biti korišten kao dodatna informacija (*goodness of fit* mjera) prilikom validacije POM-e [7]

Poglavlje 4

Usporedba dviju Polazišno-Odredišnih Matrica

4.1 Uvjeti usporedbe

- Isti grad
- Isto doba godine
- Isto vremensko razdoblje
- Week day, work day
- Ista ili slična definicija putovanja (tranzitna ili commuting matrica)
- Isto pridodijeljena putovanja periodu
- Ima li matrica diagonalu?
- obje POM-e napravljene objektivno na cijeloj populaciji 45:38 (...)
- normalizirana

4.2 Kriteriji usporedbe

*Iz teksta zadatka: ...U ovom radu potrebno je definirati **odnosne parametre kvalitete POM-e** ...*

4.2.1 Definicija putovanja

Kod generiranja matrica treba definirati hoće li se putovanja koja se protežu kroz više perioda dodijeliti vremenskom okviru u kojem započinju ili u kojem završavaju. Iako ponekad nije specificirano, autori [34] [25] [8] sva putovanja dodjeljuju intervalu u kojem je putovanje započeto. Teralytics u praksi generira obje vrste matrica [23]. Autori [8] kako bi uvrstili što više putovanja, ne zanemaruju one zapise iz kojih nije jasno vidljiv početak putovanja već takva putovanja dodjeljuje na osnovu vjerojatnosti da su započela u određenom periodu. Sličan postupak spominje se i u [21] gdje za izračun vjerojatnosti posežu za nacionalnim anketama ili frekvencijom poziva.

Usporedba matrica iz različitih izvora (ankete i CDR) s 2 različite definicije putovanja (samo putovanja HW i WH te sva putovanja) pokazala se kao loš pristup, razumljivo su matrice imale mali stupanj sličnosti ne sasvim očekivano je i strukturalna sličnost bila vrlo mala. [11] Isti zaključak kasnije iznosi i [8] čije grafičke prikaze matrica (toplinske matrice engl. (*heatmap* [25])) možemo vidjeti na slici 4.1. O definiciji putovanja dakako ovisit će njihov konačan broj, odnosno širina toka.

(...)

** Kategorične vrijednosti definicija putovanja npr.

- bez putovanja u susjedne ćelije
- bez internih putovanja (unutar iste ćelije)

**

[8] The home and work location are identified as locations with a minimum distance of three kilometers, and not belong to neighboring antennas in the Voronoi graph.

4.2.2 Prostorna razlučivost (rezolucija)

Prema hrvatskoj enciklopediji definicija razlučivosti (rezolucije) glasi: mjera za razaznavanje sitnih pojedinosti na nekom prikazu (npr. televizijskoj slici). U računalstvu se odnosi na finoću rasterske slike iskazanu ukupnim brojem slikovnih elemenata (relativna razlučivost) ili brojem slikovnih elemenata po inču (stvarna razlučivost).

(...)

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

Kod POM-a razvijenih iz CDR rezolucija će ovisiti o gustoći tornjeva bazni stanica. Gustoća tornjeva snažno je korelirana s gustoćom populacije. [8] Distribucija tornjeva nije homogena, niti na području grada, niti na bilo kojem drugom području. Kod izrazito visoke gustoće [4] predlaže grupiranje aktivnosti tornjeva koji se nalaze na udaljenosti manjoj od 75 m (dodjeljivanje aktivnosti najaktivnijem od tornjeva) kako bi se izbjegla lažna kretanja (*). (...)

Rezolucija će ovisiti o točnosti položaja! ...

Također dobro spomenuti kod poglavlja dinamička podjela → From a telecom operator perspective, the knowledge about the call density and specifically the correlation of the call density to the geography is an important call pattern. This call pattern helps the telecom operator to tune the cellular network [36]

From Figure 3 observe that the antenna density is more towards the south east of Ivory Coast. Correlating this data with Figure 5 and Figure 6 one would expect a higher call density, but a combination of light and dark shades indicates that this is not the case. Our experiments with the CDRs show a consistent low call density in the region. It cannot be conclusively said whether this observation is true all year long since the CDRs are only for five months. However, based on the available data, it can be inferred that some amount of network planning to optimise antenna density to match call density is advisable. [36]

Voronoi ćelije također je moguće ugnijezditi (što se često radi kod usporedbe s matricama dobivenim drugim postupkom), a pritom se može u matricu zabilježiti i aproksimacija internog toka, odnosno broja putovanja unutar nove ćelije (popuniti dijagonalu). Postotak eksternih putovanja koji se gubi u procesu agregacije Voronoi ćelija analiziran je za matrice iz regije Picardie u Francuskoj. **Agregacijom Voronoi ćelija na razinu *Urban Areas* (područja oko gradova) 85% svih putovanja postaje internim putovanjima, a na razini *Urban Cores* (područja oko većih gradova) čak 97% početno zabilježenih putovanja je interno, te ostaje samo 3% eksternih putovanja.** [6]. Sličnu situaciju opisuje i [9] (Vidi sliku 4.2)

Na temelju usporedbi matrica kretanja u 4 različita grada, dobivenih iz CDR i iz anketa, gdje je svaki grad u anketnim matricama imao svoju prostornu podjelu

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

odnosno rezoluciju, Toole u svom radu [21] (očekivano) zaključuje da sa **smanjenjem rezolucije (visokim stupnjem agregacije Voronoi ćelija) stupanj korelacije s matricama iz ankete (za isti grad) raste.**

**

Finally, we also recommend as best practice the merging of BTS activity in high density areas (in Annex 1 we use a 75m cordon, below which any adjacent BTS are merged). This is due to a deleterious reduction in confidence in high density areas, which are subject to greater load balancing, that the BTS's associated voronoi region accurately reflects a user's true position. [4]

all towers within 75m of each other was merged with that of the most active tower in the set.

[6] na razini pokrajine, regije : The median value of the Vorono? links' length is 75 km, the minimum is 0.6 km and the maximum is 208 km.

- gustoća baznih stanica, cca 150m → ako ignoriramo sva putovanja u susjedne ćelije

**

4.2.3 Vremenska različivost

Neki autori predlažu da kraj odnosno početak jednog dana bude u 3:00 kako bi putovanja započeta prethodnog dana, a koja prelaze u novi dan, u analizi spadala u raniji dan, odnosno kako dobar dio putovanja ne bi bio izostavljen samo zato jer se proteže kroz 2 dana [22][21] .

** brzina kretanja, primjer, matrice prozora od 5 min, ferrari vs renault **

Doba dana

(...)

- morning peak

Dani u tjednu i praznici

Istraživanja modernih pristupa generiranju matrica ukazuju na smanjenu mobilnost za neradnih dana u tjednu, najviše u nedjelju.[20] Bitno je napomenuti da dio istraživanja ističe da je vikendima također zabilježen i manji broj telekomunikacijske aktivnosti. [6]

Osim očekivane bitne razlike u radnim i neradnim danima, uočeno je da je petak, kao zadnji radni dan u tjednu, bitno drugačiji od ostalih radnih dana. [32] [20] [5] Uočena je pojačana mobilnost petkom na razini regije (Boston Metropolitan Area) [20] i na razini cijele države (Obale Bjelokosti) s povećanjem od 35% u odnosu na nedjelju [5] na temelju matrica izrađenih iz CDR, no zanimljivo učena je smanjena mobilnost vezana za javni prijevoz (izvor za generiranje matrica *Pametne kartice za javni prijevoz*) na razini regije (Southeast Queensland, Australija). Osim prosječnog broja dnevnih putovanja, petak ima i nešto drugačiji dnevni uzorak po satima (rast prometa prema kraju radnog vremena započinje ranije). Ponovno, petkom je također uočena i najviša razina telekomunikacijske aktivnosti [6] [11] što ukazuje na moguću pristranost kada su u pitanju matrice generirane iz CDR. Rezultati nacionalne ankete u Sjedinjenim Američkim Državama pokazuju da je prosječan broj dnevnih putovanja po stanovniku 4.18 tijekom radnog dana te 3.86 tijekom neradnih dana u tjednu, što jest u skladu sa zaključcima modernih metoda.

(...)

- praznici i izvanredni događaji

Privremene turističke atrakcije i sportski događaji mogu bitno utjecati na izmjenu uobičajenog uzorka kretanja. [11]

- Odbrojavanje na novu godinu na New Yourk times square, - On an average the call density on festival days is about 10-15% higher than normal. [36]

Sezonske razlike

- [33]

- [20]

From Figure 8 if we compare the call density across different months, and days of the month, it is interesting to note that the call density towards the middle

of the month is lower in January than in other months. On the same lines note that call density in the second week of December is higher than other months, but reduces in the following weeks. Based on 2These are not recommendations, but only observations based on visual inspection this observation, we are attempting to find the mobility pattern of mobile phone users and to determine whether the mobility pattern influences this call density pattern. [36]

4.2.4 Širina toka

Ukupan broj odlazaka/dolazaka po vremenskom okviru za cijelu matricu. Suma svih njenih elemenata.

(...)

- Obuhvaća pješački promet
- veličina uzorka (postotak stanovništva)
- k-anonimizacija
- dakako za pojedinu matricu ovisi o prostornoj i vremenskoj rezoluciji
- najviše ipak vjerojatno ovisi o definiciji putovanja
- prosječan broj putovanja po stanovniku/ispitaniku Netko:nekad nešto tipa 3 i 4.5, ugl. nešto veći broj za CDR
- o tome će ovisiti i Zero-cells (Nul-ćelije)
- samo najaktivniji korisnici [22] [21] (pristranost?)-i jedno istraživanje podijelilo je korisnike u 3 skupine po količini podataka koju generiraju te su dokazali da frekvencija korištenja nije uvela pristranost. [11]
- za razliku od izvlačenja najaktivnijih korisnika -[20] uzima samo slučajno odabranih 25% korisnika (njih 1 milijun) da pojednostavi analizu. ...Dobro šta, da može pješke na prste računati?!
- broj aktivnih korisnika unutar jednog tjedna jako varira iz dana u dan. [11]

[9]

- Pristranost količine generiranih podataka s dobi [11]-i još uvijek su mobiteli slabije zastupljeni kod djece i starijih ali to su također i skupine sa smanjenom mobilnosti.

zasebni kriterij - dodatak*

- takve matrice su subjektivne, "nisu objektivne"

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

- faktor skaliranja spominje se u najmanje 2 istraživanja, u jednom se radi o $\times 10$
- noramlizacija matrica? koja je praksa ?

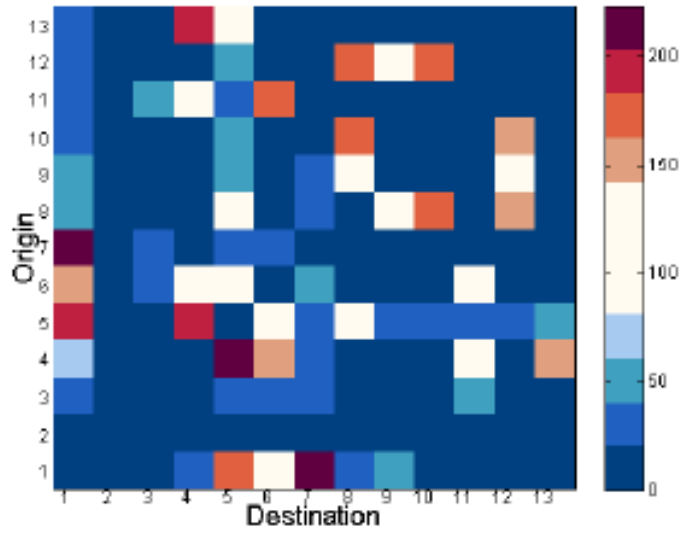
Infrastruktura ?

Obuhvaća pješački promet

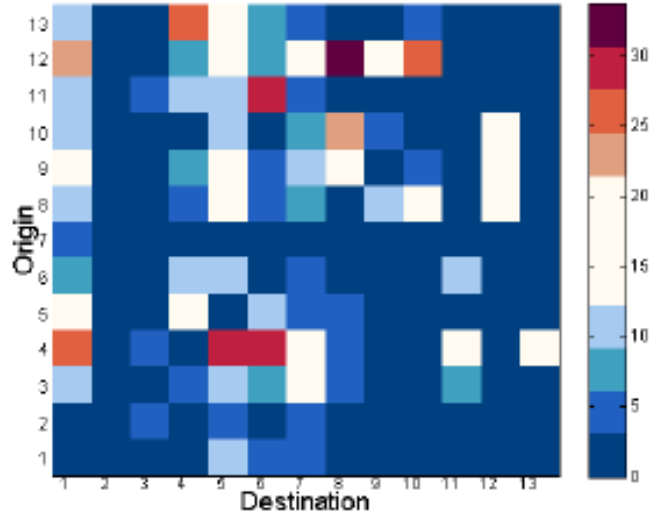
4.2.5 Gustoća informacija - kontekst

(...)

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica



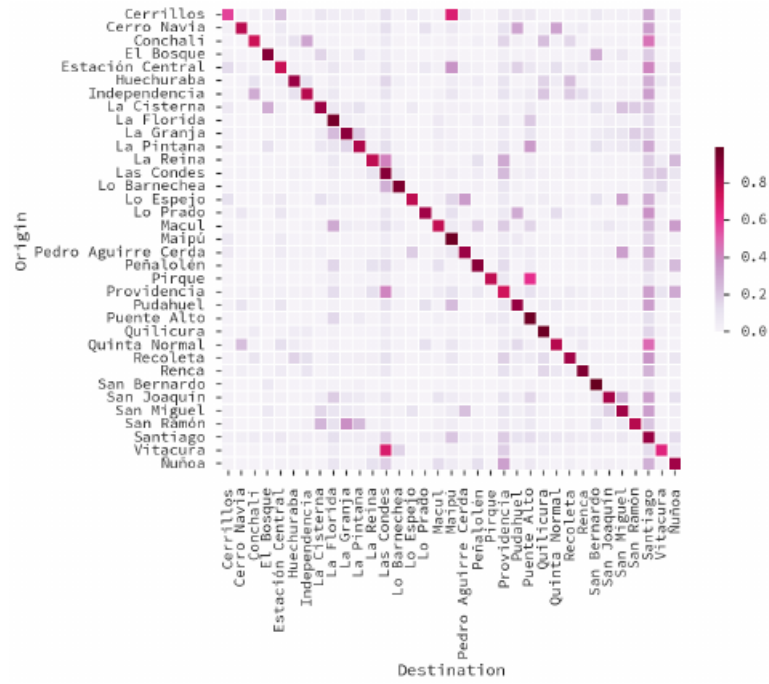
(a) Matrica u kojoj je putovanje definirano kao zabilježeni pomak između 2 antene - *kratka putovanja*



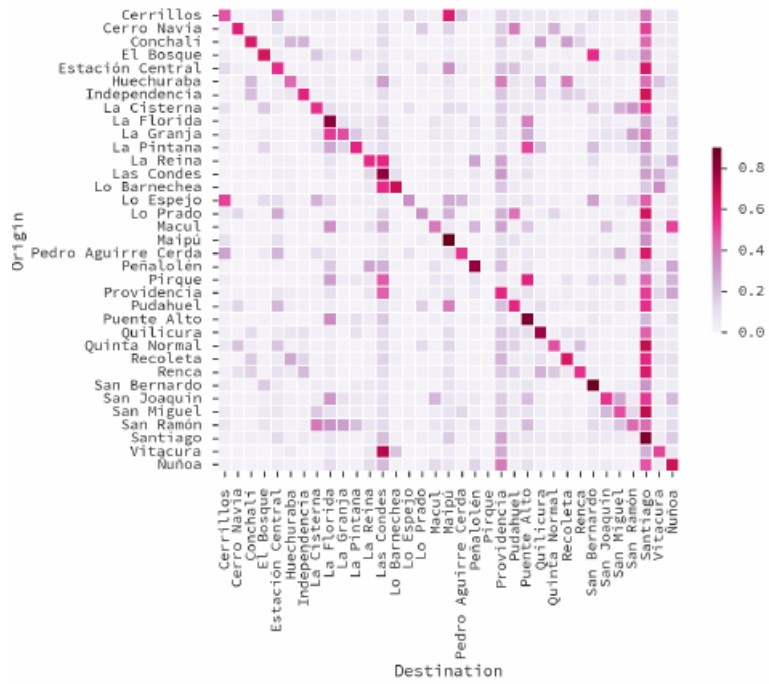
(b) Matrica u kojoj je putovanje definirano kao promjena pozicije između 2 prethodno definirana perioda, odnosno između položaja na kojima korisnik pretežito boravi u ta 2 perioda - definiranjem tih perioda pokušavaju se opisati putovanja od prebivališta na posao i s posla nazad na prebivalište engl. *commuting trips*

Slika 4.1 Dvije matrice - produkti različitih definicija putovanja. Dobivene su iz potpuno istog izvora CDR zapisa. [8]

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

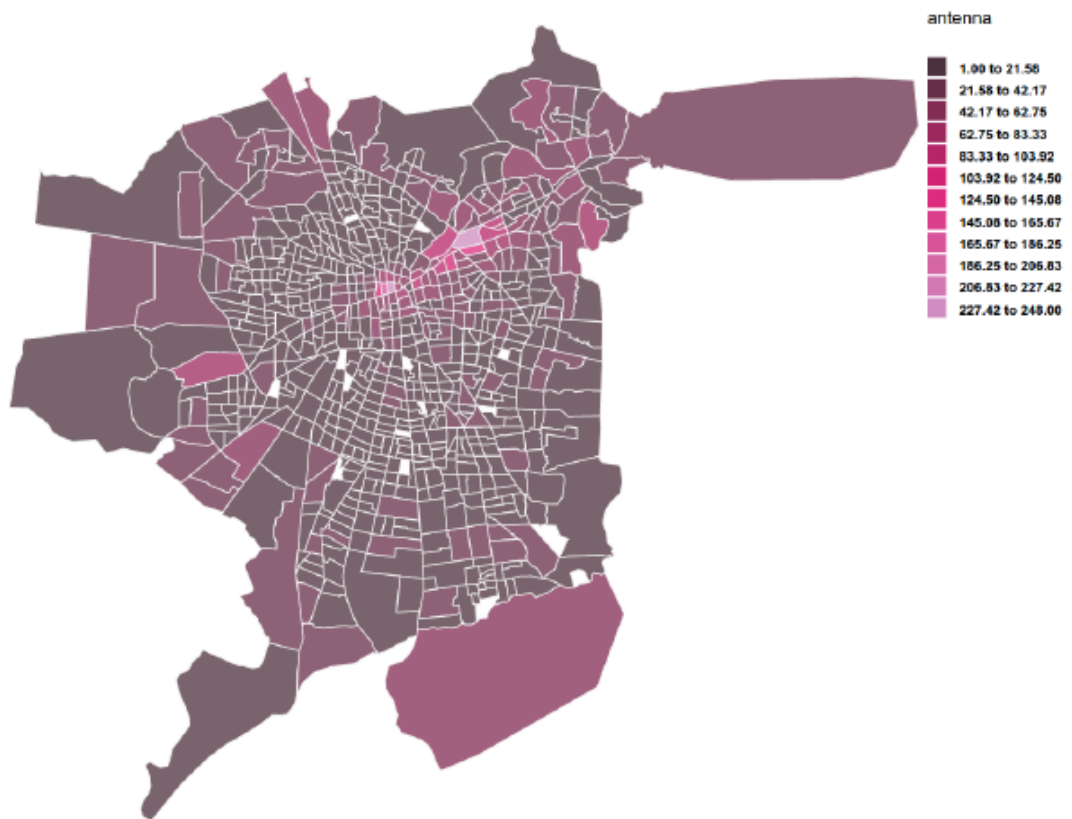


(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika 4.2 Distribucija CDR putovanja na području Santiago, Čile. Matrica je normalizirana prema redovima (polazištima) (*) [9]



Slika 4.3 Prostorna podjela na anketne jedinice u gradu Santiagu. Gustoća tornjeva baznih stanica po anketnim jedinicama. Gustoća varira od 1 do 250 tornjeva po anketnoj jedinici. [9]

Poglavlje 4. Usporedba dviju Polazišno-Odredišnih Matrica

Country		Area(/km ²)	No. of BSs	Density of BSs(1/km ²)
Europe	Poland	312685	136905	0.4378
	UK	244100	160827	0.6589
	Germany	357376	276408	0.7734
	France	553965	273279	0.4933
	Italy	301338	180867	0.6002
	Netherlands	41864	83891	2.0040
Asia	Singapore	719.1	13396	18.6288
	South Korea	100210	38845	0.3876
	Japan	377972	316977	0.8386
	China	9634057	193764	0.0201
	Tailand	513120	62410	0.1216
	India	2980000	178834	0.0600

Slika 4.4 Gustoća tornjeva baznih stanica po kvadratnom kilometru za pojedine države i grad-državu Singapur. [10]

4.3 Diskusija

(...)

- Ukoliko je rezolucija mala (velike ćelije) nema potrebe za preciznim definiranjem kraja putovanja
- O definiciji putovanja ovisit će i zero-cells ? - [8] str. 9 definicija putovanja i prostorna i vremenska rezolucija, DEFINICIJA PUTOVANJA JE KLJUČNA
- Za analizu Peak hour - intervali od 5 minuta! [8] str. 10
- 30 min vremenski okvir[22] str. 6
- short trips are less likely to use public transport[9]

stvar koje stvaraju neodređenost jer ne mogu biti objektivno definirane. Sve navedite tu što nije jednoznačno određeno, ima nekakvih preklapanja, što onda izaziva neodređenost.

Poglavlje 5

Zaključak

Bibliografija

- [1] J. Xie, Y. Nie, and X. Liu, “Testing the proportionality condition with taxi trajectory data,” *Transportation Research Part B Methodological*, May 2017.
- [2] M. K. Ng and W. S. Tang, “The Role of Planning in the Development of Shenzhen, China: Rhetoric and Realities,” *Eurasian Geography and Economics*, July 2014.
- [3] Z. Zhou, J. Yu, Z. Guo, and Y. Liu, “Visual exploration of urban functions via spatio-temporal taxi OD data,” *Journal of Visual Languages & Computing*, vol. 48, pp. 169–177, October 2018.
- [4] J. Goulding, *Best Practices and Methodology for OD Matrix Creation from CDR data*, N/LAB, University of Nottingham, 2016.
- [5] S. Scepanovic, I. Mishovski, P. Hui, J. K. Nurminen, and A. Yla-Jaaski, “Mobile Phone Call Data as Regional Socio-Economic Proxy Indicator,” *PLoS ONE*, 2015.
- [6] F. Bahoken and A.-M. Olteanu-Raimond, “Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths: The effect of spatiotemporal filtering on flow measurement,” in *ICC 13 - 26th International Cartographic Conference, Aug 2013, DRESDEN, Germany. ICC 13 - 26th International Cartographic Conference 15p*, 2013., hal-01011987v2.
- [7] T. Djukić, “Reliability assessment of dynamic OD estimation methods based on structural similarity index,” *92nd meeting of the Transportation Research Board*, 2013.
- [8] D. Gundlegard, C. Rydergren, N. Breyer, and B. Rajna, “Travel demand estimation and network assignment based on cellular network data,” *COMPUTER COMMUNICATIONS*, 2016.

Bibliografija

- [9] E. Graells-Garrido and D. Saez-Trumper, “A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow,” 2016.
- [10] Y. Chen, R. Li, Z. Zhao, and H. Zhang, “Study on Base Station Topology in Cellular Networks: Take Advantage of Alpha Shapes, Betti Numbers, and Euler Characteristics,” 2018.
- [11] P. Bonnel, E. Hombourger, A. maria Olteanu-Raimond, and Z. Smoreda, “Passive mobile phone data to construct origin-destination matrix: Potentials and limitations,” *Transportation Research Procedia, Elsevier*, no. 11, pp. 381.–398., 2015.
- [12] M. Cools, E. Moons, and G. Wets, “Assessing the Quality of Origin-Destination Matrices Derived from Activity Travel Surveys,” *Transportation Research Record: Journal of the Transportation Research Board*, 2010.
- [13] T. Pollard, N. Taylor, T. van Vuren, and M. MacDonald, “Comparing the Quality of OD Matrices in Time and Between Data Sources,” *European Transport Conference*, 2013.
- [14] M. Alhazzani, F. Alhasoun, Z. Alawwad, and M. C. González, “Urban Attractors: Discovering Patterns in Regions of Attraction in Cities,” *Public Library of Science*, 2016.
- [15] A. Peterson, “The Origin-Destination Matrix Estimation Problem- Analysis and Computations,” Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.
- [16] N. Jelusic, “Telematicka sucelja (nastavni tekst),” 2016.
- [17] M. Coscia and R. Hausmann, “Evidence that Calls-based Mobility Networks are Isomorphic,” *Working Papers - Center for International Development at Harvard University*, 2015.
- [18] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, “Estimation of urban commuting patterns using cellphone network data,” 2012.
- [19] (06.05.2019) ecall. , s Interneta, <https://en.wikipedia.org/wiki/ECall>
- [20] F. Calabrese, G. D. Lorenzo, L. Liu, and C. Ratti, “Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area,” *IEEE Pervasive Computing* 10, no. 4, pp. 36–44., April, 2011.

Bibliografija

- [21] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evskoff, and M. C. González, “The path most traveled: Travel demand estimaton using big data resources,” *Transport Research Part C*, 2015.
- [22] C. M. Schneider, V. Beik, T. Couroneé, Z. Smoreda, and M. C. González, “Unraveling daily human motifs,” *J. R. Soc. Interface*, no. 10, May 2013.
- [23] (03.06.2019) Teralytics. , s Interneta, <https://thenewstack.io/teralytics-takes-big-data-approach-human-movement/>
- [24] I. Stupar, P. Martinjak, V. Turk, and R. Filjar, “Socio-Economic Origin-Destination Matrix Derivation Through Contextualization of Material World,” *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018.
- [25] M. Filić, R. Filjar, and K. Vidović, “Graphical Representation of Origin-Destination Matrix in R Statistical Environment,” *36. skup o prometnim sustavima s međunarodnim sudjelovanjem AUTOMATIZACIJA U PROMETU*, November 2016.
- [26] (03.06.2019) Data description for urbancps. , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>
- [27] H. Zhao and Y. Zhao, “Traffic analysis zones - how do we move forward?” *AASHTO CTPP and TRB Joint Conference, Applying Census Data for Transportation*, 2017.
- [28] Z. Zheng, Z. Haung, F. Zhang, and P. Wang, “Understand coupling dynamic of public transport network,” *EPJ Data Science*, 2018.
- [29] E. Mellegard, “Obtaining origin-destination- matrices from cellular network data,” Master’s thesis, Chalmers University of Technology, Department of Mathematical Sciences, 2011.
- [30] (21.06.2019) Opencellid. , s Interneta, <https://www.opencellid.org/>
- [31] P. Hui, R. Mortier, M. Piorkowski, T. Henderson, and J. Crowcroft, “Planet-scale human mobility measurement,” in *2ND ACM International Workshop on Hot Topics in Planet-scale Measurement. Hot Planet*, 2010.
- [32] A. Travassoli, A. Alsger, M. Hickman, and M. Meshbah, “How close the models are to the reality? Comparison of Transit Origin-Destination Estimates with Automatic Fare Collection Data,” *Australian Transport Research Forum*, 2016.

Bibliografija

- [33] W. Kuhlman, “The construct of purpose specific od matrices using public transport smart card data,” Tech. Rep., 2015.
- [34] S. Bera and K. V. K. Rao, “Estimation of origin-destination matrix from traffic counts: the state of the art,” *European Transport Trasporti Europei n. 49*, 2011.
- [35] T. van Vuren, “256 shades of gray - comparing OD matrices using image quality assessment techniques,” *Scottish Transport Applications Research (STAR)*, 2015.
- [36] J. Varun, H. Sunil, R. Vasisht, S. Saragur, and A. Lele, “On the Analysis and Visualisation of Anonymised Call Detail Records,” 2013.

Pojmovnik

ANPR Automatic Number Plate Recognition. 3

CDR Call Data Records ili Charging Data Records. ix, x, 6–10, 13–15, 17, 20, 22, 27–29, 31, 32

D4D Data for Development. 17

FCD Floating Car Data. 6

FPD Floating Phone Data (Floating Cellular Data). 6

GIS Geographic Information System. 10

GNSS Global Navigation Satellite System. 6, 8

GPS Global Positioning System. 10, 14

GSM Global System for Mobile (Communications). 6, 7

HW Home-Work. 9, 22, 27

ITS Inteligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*). 4, 6

MSSIM Mean Structural Similarity index. 24

ODM Origin-destination Martix. 1

OSM OpenStreetMap. 10, 11, 13

POIs Points of Interest. 10

POM Polazišno-Odredišna Matrica. vi, vii, 3, 4, 7, 25–33

RSI Road Side Interview. 3

SSIM Structural Similarity index. 24

TAZ Traffic Analysis Zone. ix, 12, 13

WH Work-Home. 9, 22, 27

Sažetak

Ovo je tekst u kojem se opiše sažetak vašega rada. Tekst treba imati duh rekapitulacije što je prikazano u radu, nakon čega slijedi 3-5 ključnih riječi (zamijenite dolje postavljene općenite predloške riječi nekim suvislim vlastitim ključnim riječima).

Ključne riječi — Polazišno-odredišna matrica, parametri kvalitete, usporedba

Abstract

This is a text where a brief summary of your work is outlined. The text should have a sense of recap of what was presented in the thesis, followed by 3-5 keywords (replace the general keyword templates below with some meaningful keywords of your own) .

Keywords —Origin-Destination Matrix, quality parameters, keyword 3

Dodatak A

Naslov priloga

A.1 Naslov sekcije

A.2 Naslov sekcije