

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

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

Rijeka, rujan 2019.

Vjera Turk
0069064924

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

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

Mentor: prof.dr.sc. Renato Filjar

Rijeka, rujan 2019.

Vjera Turk
0069064924

Umjesto ove stranice umetnuti zadatak za završni ili diplomski rad

Naslov zadatka: Metodologija za usporedbu kontekstualiziranih polazišno-odredišnih matrica

Thesis title: Methodology for contextualised origin – destination matrices comparison

Polje znanstvenog područja: Računarstvo

Grana znanstvenog područja: Informacijski sustavi

Sadržaj zadatka: Polazišno-odredišna matrica (POM) omogućuje sustavnu statističku procjenu migracija stanovništva u zadanom prostorno-vremenskom okviru. Za razliku od tradicionalnog pristupa brojanja putovanja i putnika, za procjenu POM-e 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. **Pojavljuje se potreba za objektivnom procjenom kvalitete POM-e u odnosu na referentnu (kontrolnu).** U ovom radu potrebno je definirati odnosne parametre kvalitete POM-e te razviti metodologiju usporedbe dviju POM-a dobivenih različitim postupcima procjene i s podacima iz različitih izvora. Usporedbu obaviti korištenjem numeričkog i grafičkog oblika POM-e. Metodologiju je potrebno izvesti u programskom okruženju za statističko računarstvo R te demonstrirati njenu primjenu na slučaju usporedbe dviju POM-a. Komentirati dobivene rezultate sa stajališta apsolutne i relativne točnosti matrica.

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradila ovaj rad.

Rijeka, rujan 2019.

Ime Prezime

Zahvala

Sadržaj

Popis slika	ix
Popis tablica	x
1 Uvod	1
2 Polazišno-Odredišna Matrica	3
2.1 Definicija Polazišno odredišne matrice	3
2.2 Tradicionalni pristupi procjeni Polazišno-Odredišnih Matrica	4
2.2.1 Izvori ulaznih podataka	4
2.2.2 Matematičko modeliranje prometa	5
2.2.3 Pokretna osjetila	6
2.3 Suvremeni pristupi procjeni Polazišno-Odredišnih Matrica	7
2.3.1 Polazišno-Odredišna Matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži	7
2.3.2 Združena očitavanja prijamnika za satelitsku navigaciju	9
2.3.3 Javni prijevoz i <i>pametne kartice</i>	9
2.3.4 WiFi	9

Sadržaj

3	Motivacija	10
3.1	Problematika i ograničenja tradicionalnih pristupa procjene Polazišno-Odredišna Matrica (POM)-e	10
3.2	Formulacija problema	12
3.3	Pregled prethodnih istraživanja	12
3.3.1	Primjeri korištenja vanjskih izvora konteksta u analizi kretanja	12
3.3.2	Grafički oblik i strukturalna sličnost Polazišno-Odredišnih Matrica	13
3.4	Teza	17
4	Metodologija	18
5	Rezultati	19
6	Diskusija	20
7	Zaključak	21
	Bibliografija	22
	Pojmovnik	26
	Sažetak	27
A	Postojeće metrike za validaciju Polazišno-Odredišnih Matrica	28
A.0.1	Metrike za procjenu sličnosti Polazišno-Odredišnih Matrica s referentnom	28
B	Prilagodba prostorne podjele	30
B.1	Grupacija baznih stanica	30
B.2	Interpolacija	31

Sadržaj

B.3 Posljedice agregacije prostornih ćelija	32
---	----

Popis slika

2.1	Polazišno-Odredišna Matrica i pripadajući usmjereni graf *zamjeniti novom kojaima i i j*	4
3.1	Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. Structural Similarity index (SSIM) (i Mean Structural Similarity index (MSSIM)) u odnosu na referentnu POM-u može biti korišten kao dodatna informacija (<i>goodness of fit</i> mjera) prilikom validacije POM-e [1]	16
B.1	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") Vornoi ćelija. Sumom svih rezultata krhotina jedne kvadratne ćelije izračuna se konačni tok. [2]	31
B.2	Distribucija CDR putovanja na području Santiaga, Čile. POM je normalizirana prema redovima (polazištima) (*) [3]	33

Popis tablica

Poglavlje 1

Uvod

Temu rada na temelju dostupne literature smjestiti u šire područje struke. U pravilu Uvod ne bi trebao biti duži od dvije stranice.

POM *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 ras-križjima, 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 nosi informaciju o kontekstu - svrsi kretanja- POM-a specificira socio-ekonomske aktivnosti koje su pokretači putovanja odnosno motivatori individualne i grupne mobilnosti. Suvremeni izvori podataka o položaju kao što su zapisi o telekomunikacijskoj aktivnosti nisu vezani uz prometnu infrastrukturu niti određen način prijevoza.

Nedostaci kao što su mali uzorak stanovništva, neodređenost polazišta i odredišta, korištenje podataka koji nisu ažurni čine tradicionalne metode sve manje kompetentnima u rapidno promjenjivim i razvijajućim sredinama.

Poglavlje 1. Uvod

Suvremeni izvori podataka o položaju i načini procijene POM-a koji uključuju kontekstualizaciju nov su doprinos u području analize kretanja stanovništva. Postojeće metode validacije POM-a definiraju vrijednost nove matrice razmatranjem njene sličnosti s postojećom matricom za isto područje. Postoji potreba da se definira i kvantizira kvaliteta POM-e kroz objektivne parametre. U pojedinom suvremenom pristupu postoji idalje prostor za postizanje različite razine kvalitete. Definiranjem parametara kvalitete i njihovih mogućih vrijednosti te razmatranjem učinka vrijednosti parametara na količinu informacije koju iščitavamo iz POM dan je okvir za donošenje odluka o kvaliteti POM-a. (...)

Poglavlje 2

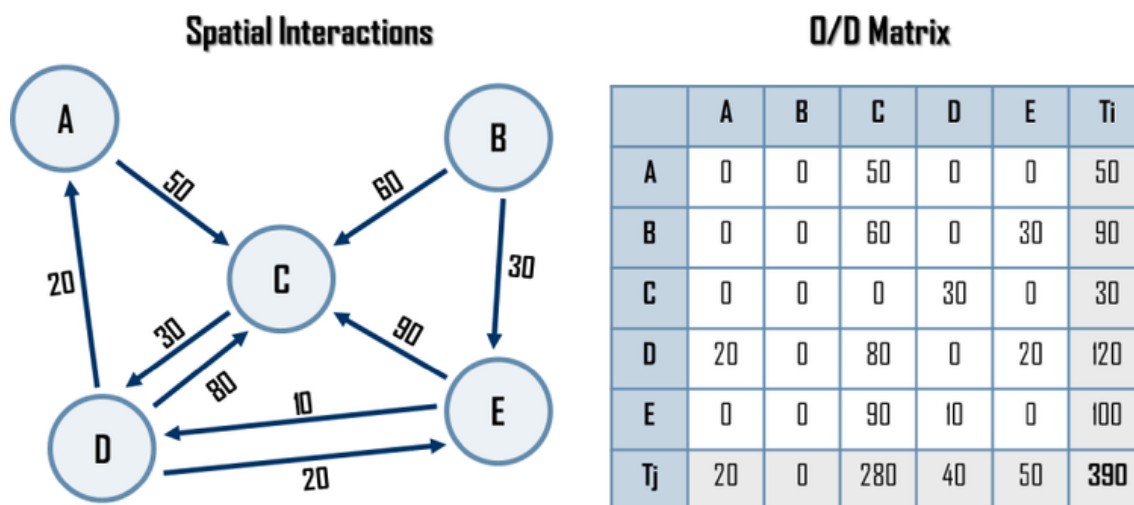
Polazišno-Odredišna Matrica

2.1 Definicija Polazišno odredišne matrice

Definicija 2.1.1. Polazišno-odredišna matrica definirana je tako da svaki njen element predstavlja broj putovanja između odgovarajućeg para zona unutar promatranog područja u promatranom vremenskom razdoblju T , u svakom smjeru zasebno. Redovi u POM-i $i = 1, 2, 3 \dots m$ predstavljaju polazišta (izvore) prometnog toka, a stupci $j = 1, 2, 3 \dots n$ njihova odredišta (ušća). Element na poziciji i, j u POM-i f_{ij} predstavlja broj putovanja iz pripadajuće polazišne zone i u pripadajuću ciljnu zonu j . Broj putovanja naziva se širina prometnog toka (eng. *flow*).

POM-a se može 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.

Postoje brojni postupci procjene POM o kojima je riječ u narednom odjeljku. POM-a može biti procijenjena za sadašnje ili planirano prometno opterećenje. Može biti prikazana ukupno ili po pojedinim načinima prijevoza (modalitetima prometnog sustava - osobna vozila, vozila javnog prijevoza, teretna vozila, pješaci itd.) i/ili različitim svrhama putovanja (putovanja na posao, putovanja kući...) . [4]



Slika 2.1 Polazišno-Odredišna Matrica i pripadajući usmjereni graf *zamjeniti novom kojom i i j*

2.2 Tradicionalni pristupi procjeni Polazišno-Odredišnih Matrica

2.2.1 Izvori ulaznih podataka

Ankete

Tradicionalni pristupi 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 [5]. Ankete su zahtjevne iz perspektive utrošenog vremena, uloženog truda i radne snage 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. [5]

Prebrojavanje vozila

Metode koje uključuju prebrojavanje vozila na ključnim čvorištima prometne infrastrukture manje su zahtjevne za izvođenje u odnosu na ankete i značajno smanjuju vrijeme i troškove prikupljanja podataka. Postupci uključuju: analizu nadzornih snimaka prometa, automatski sustav za prepoznavanje i očitovanje registarskih pločica engl. Automatic Number Plate Recognition (ANPR), osjetila prometnog toka (radarska, magnetna, video-senzori, zvučna..) i ručno prebrojavanje. **Primjenom ovih postupaka nije moguće pouzdano odrediti stvarna polazišta i odredišta putovanja [6], već se ona određuju matematički modeliranjem uz uvažavanje pretpostavki***. Pozicioniranje opreme ili osoba koje rade obzervaciju prometa treba biti plansko, a broj dostupnih uređaja/ljudi ograničavajući je faktor za veličinu područja koje je moguće nadzirati. [7]

2.2.2 Matematičko modeliranje prometa

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]

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 [9], da svi korisnici imaju saznanja o situaciji na cesti [2] te da je mobilnost općenito manja tijekom zime nego tijekom ljeta [6].

2.2.3 Pokretna osjetila

Pokretno osjetilo je uređaj koji je postavljen na vozilo koje se potom kreće prometnom infrastrukturom, postajući tako dijelom prometnog toka, te koji je sposoban opažati vrijednosti definiranih parametara gibanja vozila na koje je postavljeno te ih pohranjivati i prenositi radi naknadnog procesiranja. Podatke koje osjetilo prikuplja su položaj, brzina i smjer kretanja vozila.

Zahtjev koji nužno mora biti ispunjen je poznavanje položaja vozila. Uređaj određuje položaj pomoću Global Navigation Satellite System (GNSS) prijemnika ili se vozilo identificira na određenim točkama u prometnoj mreži, na kojima može ujedno i prenijeti podatke prikupljene od posljednje točke. Anglosaksonski naziv za ovu metodu prikupljanja podataka o stanju u prometu je Floating Car Data (FCD). Danas je FCD metoda zastupljena u ITS. **U prikupljanju podataka sudjeluju uređajima opremljena taksi vozila, službena vozila, vozila za distribuciju robe, i vozila javnog prijevoza.** Kod ITS koji zahtijevaju podatke u realnom vremenu mora biti uspostavljen komunikacijski sustav za učestali prijenos podataka o položaju, a najčešće se odvija preko Global System for Mobile (Communications) (GSM), Wi-fi i Bluetooth komunikacijskih sustava. [4]

Ukoliko pokretno osjetilo nije specijalizirani uređaj već pametni telefon koristi se 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. Položaj pokretne stanice unutar javne pokretne mreže može se odrediti triangulacijom signala koje pokretna stanica hvata iz više baznih stanica ili približno odrediti izjednačavanjem položaja pokretne stanice s položajem (područjem pokrivanja) bazne stanice s kojom je pokretna stanica povezana u promatranom trenutku.

2.3 Suvremeni pristupi procjeni Polazišno-Odredišnih Matrica

Sve brojnije informacijske i telekomunikacijske usluge generiraju masivne skupove podataka vezanih za gibanje u prostoru (mobilnost). Podatci koje generiraju pametni telefoni i pametne kartice omogućuju podatkovnim analitičarima razumjeti ponašanje pojedinaca u brojnim domenama, uključujući i njihov obrazac kretanja. Danas oni postaju suvremeni izvori podataka o dnevnim mobilnosti stanovništva.

Pružatelji usluga zasnovanih na lokaciji i operatori javne pokretne mreže moraju voditi računa o privatnosti pojedinaca kada je riječ o podacima koje oni prikupljaju, uključujući podatke o položaju korisnika. Ukoliko ustupe podatke za razvoj POM-a, takvi podaci moraju biti anonimizirani te ne smije postojati mogućnost povratnog inženjerstva kojim bi se povratio izvorni identitet. Podaci moraju biti anonimizirani prije pristupa da mogu biti korišteni bez razotkrivanja pojedinaca u skladu s postavkama General Data Protection Regulation (GDPR). Važno je osigurati anonimnost i dozvolu pojedinca da se podaci koriste u jasno definirane svrhe. Idealno je da operator ustupi već anonimizirane podatke stranci koja iz njih procjenjuje POM-e. (...)

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

Operatori javnih pokretnih mreža prikupljaju podatke o aktivnostima u javnoj pokretnoj mreži za potrebe pružanja i naplate usluga te za održavanje operativnosti same mreže. Pokretni uređaj (npr. pametni telefon) u javnoj pokretnoj mreži kontinuirano je povezan s mrežom putem odgovarajuće bazne stanice, neovisno o telekomunikacijskoj aktivnosti uređaja. U postupku uspostave poziva, signalizacija uključuje i identifikacijsku oznaku bazne stanice s koje je poziv uspostavljen. Navedena oznaka bilježi se u zapisu o telekomunikacijskoj aktivnosti. Zapisi o telekomunikacijskoj aktivnosti i detaljima telekomunikacijskih transakcija vode se u odgovarajućim komponentama javne pokretne mreže pod imenima: Call Data Records ili Charging Data Records (CDR). CDR obuhvaća transakcijske podatke o

pozivima, razmjeni kratkih poruka (SMS) i prijenosu podataka pokretnim internetom. Transakcijski podatci uključuju, između ostalog: vremenske trenutke početka i završetka telekomunikacijske transakcije (korištenja komunikacijske usluge), vrstu komunikacije, identifikacijske oznake pošiljatelja i primatelja, te identifikacijske oznake baznih stanica na koje su u trenutku započinjanja i završetka transakcije bili spojeni pošiljatelj i primatelj. Operatoru su dodatno poznati podatci o prekapčanju pokretnog uređaja na drugu baznu stanicu (engl. *handover data*) tijekom trajanja telekomunikacijske aktivnosti radi postizanja i održavanja potrebne kvalitete telekomunikacijske usluge, *Location Area Updates* - podatci prikupljeni svakih 20 min do 2 sata o tome na koju skupinu baznih stanica je pokretna stanica spojena (prema [6] svaka LA skupina ima od 150 do 500 baznih stanica), te podatke o tome kada je pokretna stanica uključena u mrežu i isključena iz mreže.

U novije vrijeme su predloženi postupci u kojima se koriste upravo anonimizirane CDR kao izvor podataka za procjenu POM-a.[8] [10] [11] Metodologija korištenja odnosnih elemenata CDR-a za procjenu POM predložena je još 1994. godine.[8] **U ostatku rada, na mjestima gdje nije izričito naveden izvor podataka o kretanju, riječ je o postupku ili POM-i procijenom iz CDR.**

Definicija 2.3.1. Zaustavljanje (engl. *Stop*) je skup minimalno n događaja (telekomunikacijskih aktivnosti) u nizu koji su zabilježeni na istom položaju unutar definiranog minimalnog vremenskog perioda.

Zahtjevom da se radi o više ($n > 1$), a ne jednom ($n = 1$) zabilježenom događaju pokušava se osigurati da se ne radi o tranzitnom već o stvarnom zaustavljanju i suzbija se bilježenje lažnog kretanje kod privremenog spajanja na susjednu baznu stanicu. U slučaju kada je $n > 1$ zaustavljanje ima 2 vremenske oznake, trenutak kada je zabilježen početak i trenutak kada je zabilježen kraj zaustavljanja.

Definicija 2.3.2. Putovanje je skup dva zaustavljanja koja zadovoljavaju unaprijed definirana minimalna i maksimalna ograničenja međusobne vremenske i prostorne udaljenosti. Kraj polazišnog zaustavljanja postaje početak putovanja, a početak odredišnog zaustavljanja kraj putovanja. Niz kontinuiranog zaustavljanja na različitim položajima otkriva putanju kretanja.

Kretanje pojedinca tokom dana najčešće je rascjepkano, odnosno u nizu postoje

”rupe” gdje je informacija o kretanju nepoznata zbog nepostojanja aktivnosti u tom vremenu.

Svako zaustavljanje (2.3.1) ne mora nužno biti (postati) početak ili kraj putovanja. **Postavljanjem uvjeta da zaustavljanje jest polazište ili odredište putovanja samo ako mu se dokaže kontekst od interesa, eliminiraju se usputna zaustavljanja.** Primjerice, ako je kontekst od interesa ograničen na isključivo ”posao” i ”prebivalište” s ciljem promatranja samo komutacijskih putovanja, zaustavljanje na benzinskoj crpki neće se gledati kao kraj putovanja. Komutacijska putovanja se ponavljaju na dnevnoj bazi te su stoga često upravo ona koje se nastoji obuhvatiti POM-ma.

Ograničavanjem na isključivo komutacijska putovanja, velik će dio putovanja *koja su predvidiva na razini toka* biti odbačen. Stoga uvođenjem nekoliko generalnih kontekstnih kategorija ...

2.3.2 Združena očitavanja prijamnika za satelitsku navigaciju

(...)

komercijalna vozila, Taxi [12]

2.3.3 Javni prijevoz i pametne kartice

(..)

82% putovanja javnim prijevozom naprave korisnici javnog prijevoza sa pametnim karticama. [13]

- [14] [15]

2.3.4 WiFi

(...)

- ITrack, CTrack, VTrack

Poglavlje 3

Motivacija

- pojava novih izvora podataka i postupak procjene migracija
- veća preciznost određivanja položaja. polazišta i odredišta
- kontekstualizacija
- točnost opisa putovanja raste
- područje primjene raste
- potencijalno vrlo velik reprezentativni uzorak (CDR)
- jako dobra rezolucija u vremenu
- osigurati anonimnost

3.1 Problematika i ograničenja tradicionalnih pristupa procjene POM-e

Širenje gradova i rast stanovništva rezultirali su rastućim brojem sve ozbiljnijih prometnih zastoja u gradovima diljem svijeta. Prepoznata je potreba za strategijama upravljanja prometom i gradom općenito koje će uvažiti dinamiku razvoja stanovništva i njegove migracije u kontekstu suvremenih društvenih i gospodarskih kretanja. Djelotvorno upravljanje i kontrola prometa doprinose povećanju sigurnosti, kvalitete prometne i kvalitete transportno-logističke usluge, potiče ekonomski rast i smanjuje zagađenje okoliša. Suvremeni odnosi promijenili su prirodu kretanja ljudi

i dobara zbog čega tradicionalni pristup postaje zastario i neprimjeren.

Redefiniranje metodologije koja će pružiti povećanu kvalitetu i transparentnost uz ispunjavanje suvremenih zahtjeva na tehnologije (cijena, zaštita privatnosti i podataka itd.) privlači pažnju istraživača i van područja prometne znanosti. Oni predlažu alternativne izvore podataka o kretanju za promatranje obrazaca putovanja i pristupe procjeni POM-a, no prema [16] nailaze na problem da prometna struka zahtijeva da se njihova vrijednost dokazuje usporedbom s tradicionalnim pristupima, dobro poznatim u praksi. **Istraživači proširuju područje primjene zaključaka analize kretanja stanovništva uzimajući u obzir njihov kontekst [17]. (...)**

Kada je riječ o validaciji POM, pregledom literature utvrđeno je da se jednoznačno koriste pojmovi točnost, pouzdanost i kvaliteta POM-a. Za njihovu procjenu koriste se mjere koje opisuju razinu sličnosti odnosno razlike (greške) s referentnom matricom - a temelj je razlika na razni svakog elementa. **Tradicionalni postupak validacije POM-e svodi se na određivanje njene sličnosti s *grand truth* matricom.** [13] Grand truth matrix je POM-a je koja je uzeta kao ona koja najtočnije opisuje stvarno stanje. U analizi prometa, dobivena je tradicionalnim postupcima (anketiranje i/ili prebrojavanje vozila). Statističke mjere kvantiziraju razliku procijenjenih i “istinitih” vrijednosti. (Vidi Dodatak A - Postojeće metrik za validaciju Polazišno-Odredišnih Matrica). Ankete pokrivaju malen dio populacije te je na taj način dobivena POM-a opterećena neizbježnom pristranošću [10]. Studija iz 2010. [5] pokazala je da se zadovoljavajuća razina kvalitete POM-a procijenjenih isključivo iz anketa u kućanstvima postiže tek ako uzorak obuhvaća 50% populacije.

Novi moderni postupci procjene POM-a postaju sve više razmatrana opcija s obzirom na fleksibilnost, cijenu i ažurnosti koju nude u odnosu na tradicionalne. Prepoznat je potencijal, ne samo za analizu urbanih i državnih razmjera, već i za analizu mobilnosti planetarnog razmjera dijeljenjem i kombiniranjem različitih izvora podataka koji nose podatak o položaju [18][19]. CDR se ističe kao izvor gotovo jednako zastupljen u zemljama u razvoju i razvijenim zemljama, s penetracijom oko 40%-50% stanovništva za pojedine operatore [6]. U Francuskoj 2008. 80% stanovništva starijeg od 12 godina posjedovalo je mobilni uređaj [6]. Istraživanje iz 2016. provedeno na podacima u Čileu iznosi podatak da Čile ima čak 132 mobilne pretplate na 100 stanovnika [3].

(...)

3.2 Formulacija problema

Analiza problema i osnovne zaonitosti vezane uz njegovo rješavanje

Postoji potreba za objektivnom procjenom kvalitete POM-a definiranjem odnosnih parametara kvalitete.

(...)

3.3 Pregled prethodnih istraživanja

3.3.1 Primjeri korištenja vanjskih izvora konteksta u analizi kretanja

Studija A -

U jednom od svojih radova grupa autora kvantizira vezu između ukupno 23 namjenske kategorije Points of Interest (POIs) iz službene gradske baze (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. [10] 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. Njihovi rezultati mogu se primijeniti u planiranju pozicioniranja budućih objekata u gradu.

Studija B - Model Raspodjele Toka

Inovativni pristup kontekstualizaciji toka korištenjem OpenStreetMap baze predstavljen je u radu [20]. 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 vjerovatnosni model koji, na osnovu broja objekata pojedine kategorije u određenoj ćeliji i promatranog vremenskog okvira, ukupni ulazni tok u određenu ćeliju dijeli na 6 tokova usmjerenih prema objektima tih kategorija. Iz jedne POM-e tako se dobije 6 POM-a, po jedna za svaku od kategorija. U demonstraciji modela korištene su POM-e dobivene iz javno dostupnih, anonimiziranih telekomunikacijskih zapisa na području kineskog grada Shenzhena.[21][22] **Definiranjem konteksta putovanja na ovaj način sačuvana je mogućnost postizanja dobre rezolucije u vremenu, a istovremeno je dan kontekst koji uključuje tipične komutacijske *Home* i *Work* kategorije, te umjesto općenite kategorije *Other* precizira 4 kategorije.**

Definiranjem konteksta dobivamo nove informacije o kretanju unutar ćelije, odnosno o konačnoj destinaciji nakon ulaska u ćeliju. U tom smislu poznata je informacija o internom kretanju.

Studija C - Planiranje održivih gradova

(...) [17]

3.3.2 Grafički oblik i strukturalna sličnost Polazišno-Odredišnih Matrica

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 [21], prikazom nijansama jedne boje 0-255 engl. *grayscale* [23], paletom nijansa 2 boje, ili kojom drugom proizvoljnom paletom s više boja (vidi paletu boja na slici ??).

Poglavlje 3. Motivacija

Nekolicina autora ističe važnost strukturalne sličnosti s referentnom matricom kao važnu mjeru kvalitete POM-e jer visoka razina strukturalne sličnosti može biti prisutna i kod POM-a s manjom razinom sličnosti prema statističkim mjerama. Također, strukturalna sličnost je (vizualno) vidljiva u grafičkom obliku POM-e. **Do-bro odgovara ljudskoj vizualnoj percepciji sličnosti slike.**

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 POM-a prvi se puta spominje i demonstrira na simuliranim POM-ma dobivenim iz referentne POM-e dodavanjem šuma. [1]

Informacija o strukturi slike definira se kao atributi 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.

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}

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

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

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

Poglavlje 3. Motivacija

$l(x, y)$ 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

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

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

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

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

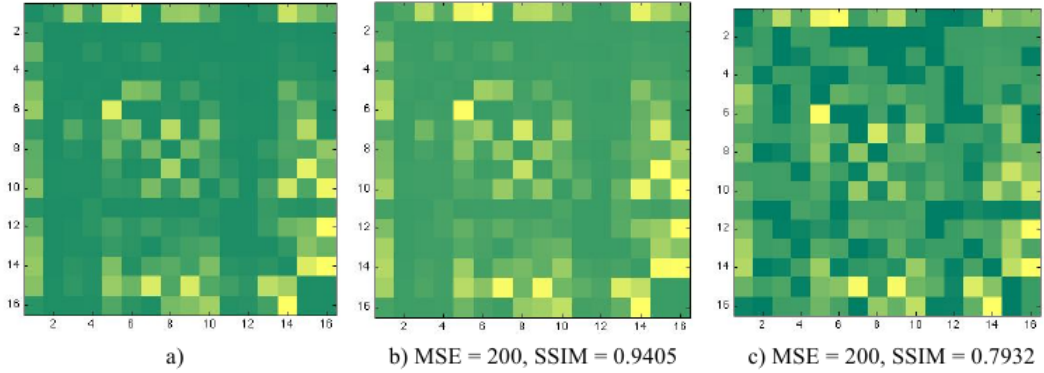
$$SSIM(x, y) = 1 \iff x = y \quad (3.5c)$$

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

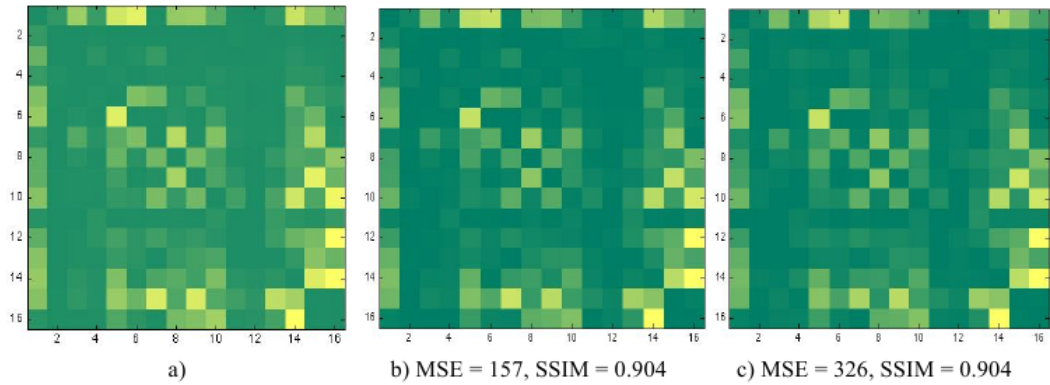
Poboljšani

Nekoliko godina nakon prvog spominjanja $MSSIM$ kao metrike usporedbe POM-a [23] se dotiče 3 problema postavljajući pitanja : Koliko treba biti veliki blok? Kako usporediti "guste" i "rijetke" POM-e? Koja je prihvatljiva vrijednost $MSSIM$? Autori definiraju poboljšani model koji nazivaju $4D-MSSIM$ gdje u izračun dodaju stvarne euklidske udaljenosti prostornih zona.

Poglavlje 3. Motivacija



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



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

Slika 3.1 Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE. SSIM (i MSSIM) u odnosu na referentnu POM-u može biti korišten kao dodatna informacija (*goodness of fit* mjera) prilikom validacije POM-e [1]

3.4 Teza

- Statistička analiza podataka iz suvremenih informacijskih i komunikacijskih sustava sve više se koristi za procjenu POM-a čime je omogućeno poboljšanje kvalitete procjene preslikavanjem na kontekst.

Određivanje konteksta proširuje područje pokrivanja u fizičkom i informacijskom smislu te povećava točnost pome unosom relevantnih atributa vezanih za način i svrhu putovanja - čime je putovanje točnije opisano

- izbor metode procjene definira preciznost određivanja konteksta

Poglavlje 4

Metodologija

opis sklopa/programske podrške. Opis rješavanja zadatka i odgovarajući proračuni

Poglavlje 5

Rezultati

iznošenje rezultata i analiza rezultata do kojih je došlo rješavanjem problematike zadatka, rezultati mjerenja

Poglavlje 6

Diskusija

Poglavlje 7

Zaključak

Bibliografija

- [1] T. Djukić, “Reliability assessment of dynamic OD estimation methods based on structural similarity index,” 92nd meeting of the Transportation Research Board, 2013.
- [2] J. Goulding, Best Practices and Methodology for OD Matrix Creation from CDR data, N/LAB, University of Nottingham, 2016.
- [3] E. Graells-Garrido and D. Saez-Trumper, “A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow,” 2016.
- [4] N. Jelusic, “Telematicka sucelja (nastavni tekst),” 2016.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] D. Gundlegard, C. Rydergren, N. Breyer, and B. Rajna, “Travel demand estimation and network assignment based on cellular network data,” COMPUTER COMMUNICATIONS, 2016.
- [9] A. Peterson, “The Origin-Destination Matrix Estimation Problem- Analysis and Computations,” Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.

Bibliografija

- [10] 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.
- [11] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, “Estimation of urban commuting patterns using cellphone network data,” 2012.
- [12] C. Krause and L. Zhang, “Short-term travel behaviour predictions with gps, land use and point of interest data,” Elsevier, 2018.
- [13] 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.
- [14] W. Kuhlman, “The construct of purpose specific od matrices using public transport smart card data,” Tech. Rep., 2015.
- [15] N. S. Aslam, T. Cheng, and J. Cheshire, “A data precision heuristic model to detect home and work location from smart card data,” Geo-spatial Information Science, 2018.
- [16] 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.
- [17] E. Barbour, C. C. Davila, S. Gupta, C. Reinhart, J. Kaur, and M. C. González, “Planning for sustainable cities by estimation building occupancy with mobile phones,” Nature Communications, 2019.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] M. Filić, R. Filjar, and K. Vidović, “Graphical Representation of Origin-Destination Matrix in R Statistical Environment,” 36. skup o

Bibliografija

- prometnim sustavima s međunarodnim sudjelovanjem AUTOMATIZACIJA U PROMETU, November 2016.
- [22] (03.06.2019) Data description for urban cps. , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>
- [23] T. van Vuren, “256 shades of gray - comparing OD matrices using image quality assessment techniques,” Scottish Transport Applications Research (STAR), 2015.
- [24] E. Mellegard, “Obtaining origin-destination- matrices from cellular network data,” Master’s thesis, Chalmers University of Technology, Department of Mathematical Sciences, 2011.
- [25] M. Coscia and R. Hausmann, “Evidence that Calls-based Mobility Networks are Isomorphic,” Working Papers - Center for International Development at Harvard University, 2015.
- [26] 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.
- [27] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evskoff, and M. C. González, “The path most traveled: Travel demand estimation using big data resources,” Transport Research Part C, 2015.

Pojmovnik

ANPR Automatic Number Plate Recognition. 4

CDR Call Data Records ili Charging Data Records. x, 6–11, 13, 16–18, 21, 23, 26, 33–38, 40

D4D Data for Development. 18, 21

FCD Floating Car Data. 5

FPD Floating Phone Data (Floating Cellular Data). 6

GDPR General Data Protection Regulation. 8

GIS Geoprostorni Informacijski Sustav. 3, 11

GNSS Global Navigation Satellite System. 5–7

GPS Global Positioning System. 11, 16

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

HW Home-Work. 9, 26, 33

ITS Inteligentnim Transportnim Sustavima (engl. *Intelligent Transportation Systems*). 4, 5, 20

MSSIM Mean Structural Similarity index. x, 29, 31

ODM Origin-destination Martix. 1

OSM OpenStreetMap. 11, 12, 15

POIs Points of Interest. 10, 13

POM Polazišno-Odredišna Matrica. vi, vii, 1–10, 13, 16, 20, 21, 23, 25–43

RSI Road Side Interview. 3, 37

SSIM Structural Similarity index. x, 29, 31

TAZ Traffic Analysis Zone. ix, 13–15

WH Work-Home. 9, 26, 33

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

Postojeće metrike za validaciju Polazišno-Odredišnih Matrica

A.0.1 Metrike za procjenu sličnosti Polazišno-Odredišnih Matrica s referentnom

Za procjenu kvalitete POM-a dobivenih isključivo anketranjem u radu [5] korištena je mjera Mean Absolute Percentage Error (MAPE), te je prikazano da se zadovoljavajuća razina kvalitete takvih POM-a postiže tek ako uzorak obuhvaća 50% populacije. Istaknuta je važnost korištenja dodatnih izvora za izradu POM-a.

U radu [7] 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).[1]

U [11] korišten je *Pearsonov koeficijent korelacije* - r da bi se utvrdila **sličnost svakog retka POM-e** dobivene iz CDR s **retkom referentne** (ukupni izlazni tok iz svake polazišne ćelije). Isti postupak korišten je za kontekstualizirane Home-Work (HW) i Work-Home (WH) POM-e dobivene iz CDR u usporedbi s referentnim POM-ma dobivenim anketiranjem.

U svome radu [13] navodi se 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 POM-e 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 [3] za procjenu sličnosti POM-a dobivenih iz CDR sa tada aktualnim POM-ma dobivenim anketiranjem.

Dodatak B

Prilagodba prostorne podjele

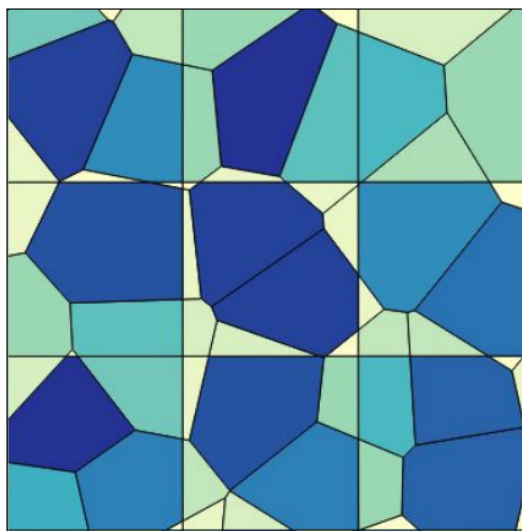
Ako je točnost matrice definirana isključivo kao sličnost svakog njenog elementa s ekvivalentnim u elementom u *grand truth* matrici nužno je osigurati da prostorna podjela matrice odgovara prostornoj podjeli referentne ili je potrebno obje matrice svesti na zajedničku prostornu podjelu.

B.1 Grupacija baznih stanica

Kod analize na području cijele države [2] koristi *k-sredina* algoritam za grupiranje baznih stanica u broj grupa jednak broju Traffic Analysis Zone (TAZ) na području Senegala. [10] na području grada Riaydha agregira bazne stanice u 1492 TAZ-e. [24] ne definira unaprijed broj grupa već za grupaciju koristi *mean-shift* algoritam koji sam određuje broj grupa na području Švedske, potom koristeći OpenStreetMap (OSM) grupe veže uz imena gradova i mjesta. U istraživanju [25] bazne stanice na području Kolumbije grupiraju po jedinicama samouprave (eng. *municipalities*). Isti postupak je korišten i u istraživanju na području Madrida, kako bi POM-e u konačnici bile usporedive s onima dobivenim iz anketa [11].

B.2 Interpolacija

Postupak konverzije POM-a iz prostorne podjele na Voronoi ćelije u drugu prostornu podjelu, npr. uniformnu kvadratnu mrežu (ćelije 1km^2), opisan u priručniku [2]. 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 B.1 je prikazan treći oblik interpolacije gdje su korišteni podaci o ukupnoj kvadratnoj površini zgrada.



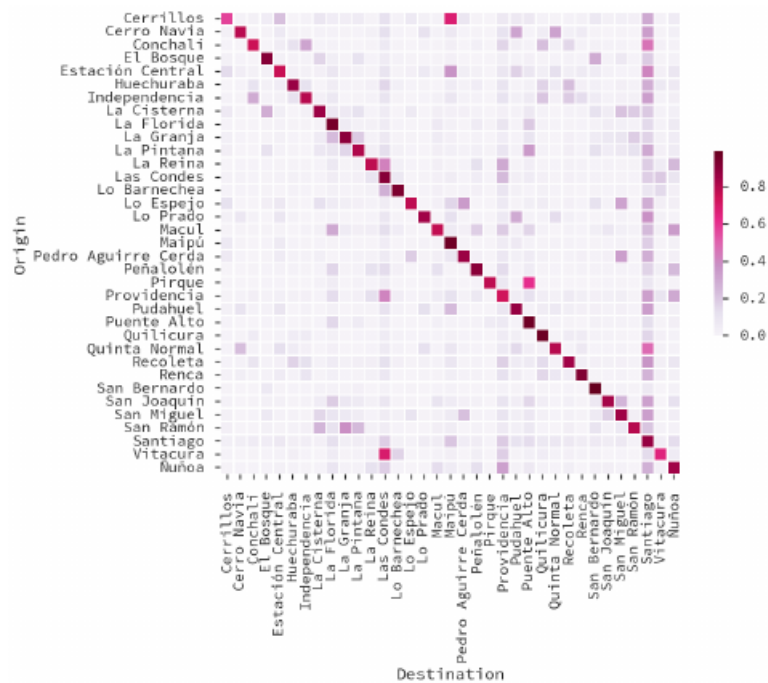
Slika B.1 Primjer mozaika *krhotina* koje dozvoljavaju interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u *krhotini* u odnosu na cijelu Voronoi ćeliju kojoj pripada. Rezultat se dobiva na osnovu tih *težinskih faktora* polazišne i odredišne *krhotine* te pripadajućih ("roditeljskih") Voronoi ćelija. Sumom svih rezultata *krhotina* jedne kvadratne ćelije izračuna se konačni tok. [2]

B.3 Posljedice agregacije prostornih ćelija

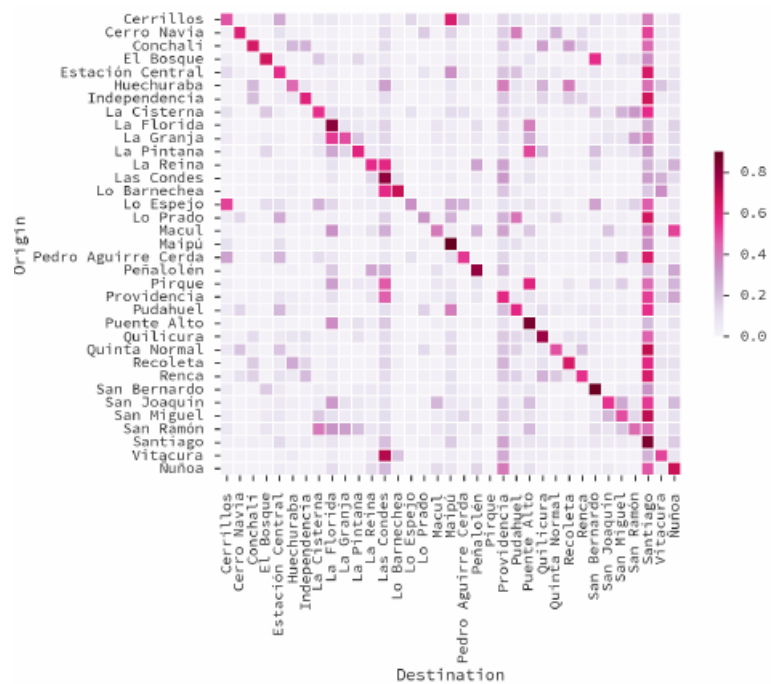
Voronoi ćelije također je moguće ugnijezditi (što se često radi kod usporedbe s POM-ma dobivenim drugim postupkom), a pritom se može u matricu bilježiti i aproksimacija internog toka, odnosno broja putovanja unutar nove ćelije (dijagonala matrice). Takva putovanja nazivaju se internim putovanjima.

Postotak eksternih putovanja koji se gubi u procesu agregacije Voronoi ćelija analiziran je za POM-e 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.** [26]. Sličnu situaciju opisuje i [3] (Vidi sliku B.2)

Na temelju usporedbi POM-a kretanja u 4 različita grada, dobivenih iz CDR i iz anketa, gdje je svaki grad u anketnim POM-ma imao svoju prostornu podjelu odnosno rezoluciju, Studija [27] je pokazala da sa **smanjenjem rezolucije (visokim stupnjem agregacije Voronoi ćelija) stupanj korelacije s POM-ma iz ankete (za isti grad) raste.**



(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika B.2 Distribucija CDR putovanja na području Santiago, Čile. POM je normalizirana prema redovima (polazištima) (*) [3]