

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

**Izazovi privatnosti na društvenim mrežama:
anonimizacija korisničkih podataka
(engl. Social data privacy)**

Dino Fuks

Voditelj: *doc. dr. sc. Vedran Podobnik*

Zagreb, svibanj, 2015

Sadržaj

1.	Uvod	1
2.	Metode anonimizacije podataka	3
2.1	K-anonimnost.....	3
2.2	SNAP - primjer anonimizacije.....	5
2.3	Anonimizacija geolokacijskih podataka	8
3.	Usporedba predloženih postupaka za anonimizaciju geolokacijskih podataka ..	14
4.	Anonimizacija geolokacijskog područja odabranim postupkom	16
5.	Poboljšanje odabranog postupka za anonimizaciju geolokacijskih podataka.....	22
6.	Zaključak.....	27
7.	Literatura.....	28
8.	Sažetak.....	29

1. Uvod

Društvena mreža predstavlja skup entiteta i njihovu međusobnu povezanost. Entiteti predstavljaju stvarne osobe, koje su društveno povezane. Društveno povezivanje označava povezivanje individualaca u specifične grupe, poput povezivanja na temelju [1]:

- mjesta u kojem ljudi žive,
- škole ili fakulteta koji pohađaju,
- radnog mjeseca,
- zajedničkih interesa grupe ljudi.

Stoga, kada govorimo o analiziranju društvene mreže, zapravo govorimo o proučavanju načina na koji su entiteti koji čine tu mrežu međusobno povezani.

Tehnološki razvoj omogućava lakše prikupljanje podataka o korisnicima društvenih mreža, a samim time povećavaju se mogućnosti analize istih. Razne agencije i istraživački timovi koji se bave prikupljanjem ranije spomenutih podataka, mogu:

- zadržati prikupljene podatke,
- javno objaviti podatke.

Ukoliko se prikupljeni podaci zadrže samo za vlastita istraživanja, biti će puno teže ili vrlo često i nemoguće provesti detaljniju analizu istih. Javno objavljivanje prikupljenih podataka omogućuje izvršavanje analize nad njima od strane većeg broja timova, a samim time analiza će biti detaljnija i kvalitetnija. Međutim, javno objavljivanje takvih podataka uzrokuje razne povrede privatnosti korisnika čiji su podaci objavljeni.

Raznim znanstvenim istraživanjima ustanovljeno je da objavljivanje podataka koji su neadekvatno anonimizirani, na način da je samo uklonjen identifikator korisnika, može dovesti do ponovne identifikacije čvorova tako anonimiziranih podataka. Ponovna identifikacija korisnika je moguća zbog toga što je promatrani (naivno anonimizirani) čvor u mreži zapravo određen vezama s drugim čvorovima, odnosno prijateljima i osobinama koje posjeduje na društvenoj mreži, dok je identifikator običan (jedinstveni) broj pridružen tom čvoru [2]. Ranije navedeno je

dovelo do razvoja brojnih algoritama i metoda koje mogu vrlo lako re-identificirati loše anonimizirani čvor u nekoj društvenoj mreži.

Geolokacijski podaci o korisniku još su osjetljivije prirode, jer se osim identiteta korisnika može otkriti njegovo kretanje, odnosno razotkrivanje njegove lokacije. Provedenim istraživanjima ustanovljeno je da se preko 80% ljudi može jednoznačno identificirati pomoću 3 lokacija na kojima on najčešće boravi [3]. Stoga je prilikom objavljuvanja jednog takvog skupa podataka koji će poslužiti za razne analize, od presudne važnosti da su ti podaci anonimizirani na ispravan način.

Pravilno anonimizirani skup podataka mora onemogućiti ponovnu identifikaciju čvorova koji predstavljaju korisnike društvene mreže, ali istovremeno mora osigurati da su elementi tog skupa podataka anonimizirani na način da se oni mogu uspoređivati odnosno omogućiti vršenje različitih analiza.

U nastavku seminarског rada najprije će biti opisani osnovni pojmovi i definicije vezane uz anonimizaciju podataka, a nakon toga će biti prikazan jedan od načina anonimizacije podataka koje je predložilo Sveučilište Stanford. Nakon spomenutog primjera slijedi opis problematike pri anonimizaciji geolokacijskih podataka, te prijedlog anonimizacije istih, koji zahtijeva poseban pristup. Konačno upoznavši se s prethodnim izazovima i zahtjevima biti će dan prijedlog anonimizacije geolokacijskog skupa podataka jednom od predloženih metoda.

2. Metode anonimizacije podataka

U nastavku teksta nalazi se opisana metoda na kojoj se zasniva postupak *K-anonimizacije* uz primjer korištenja, nakon čega slijedi analiza skupa podataka koje je objavilo Sveučilište Stanford, a poglavje završava opisom izazova koji se javljaju pri anonimizaciji geolokacijskih podataka te prijedlogom metoda za rješavanje istih.

2.1 K-anonimnost

K-anonimnost ili *K-anonimizacija* predstavlja anonimizaciju u kojoj jedan anonimizirani parametar (odnosno zapis u skupu podataka) predstavlja skup sličnih parametara koji su pridruženi najmanje *k* osobama [4]. Na taj način se ne može direktno neku osobu povezati s anonimiziranim parametrom ukoliko i dođe do otkrivanja skupa parametara koji je anonimiziran, jer svaka osoba s jednakom vjerojatnošću posjeduje bilo koji od tih parametara.

Primjer gdje bi ovakav način anonimizacije bio izuzetno pogodan je anonimiziranje podataka o studentima nekog fakulteta. U prvom primjeru pretpostavimo da imamo 5 studenata raznih fakulteta. Podaci o studentima prikazani su u tablici Tabela 1.

Tabela 1 podaci o studentima

Ime studenta	JMBAG	Ime fakulteta
Student 1	0036457801	Fakultet elektrotehnike i računarstva
Student 2	0036457802	Fakultet elektrotehnike i računarstva
Student 3	0036468811	Fakultet elektrotehnike i računarstva
Student 4	0036468812	Fakultet elektrotehnike i računarstva
Student 5	0016110001	Fakultet organizacije i informatike
Student 6	0016110002	Fakultet organizacije i informatike

Ukoliko podatke o studentima anonimizirati na način da prilikom usporedbe podataka želimo znati s kojeg fakulteta promatrani student dolazi, tada je potrebno prve 4 znamenke JMBAG-a koje predstavljaju identifikator pojedinog fakulteta ostaviti

netaknute dok je ostale znamenke potrebno anonimizirati. Na ovaj način osigurali smo da se za sve studente može saznati s kojeg fakulteta oni dolaze, ali istodobno ne znamo nikakve druge detalje o tom studentu. Navedena anonimizacija posjeduje svojstvo *K-anonimnosti* jer ukoliko i uspijemo otkriti koji se sve skup podataka (JMBAG-a) krije iza prametra (npr. 0036xxxxxx) još uvijek ne znamo koji točno JMBAG pripada kojem studentu, već svaki student može s jednakom vjerojatnošću posjedovati bilo koji JMBAG. Primjer ovakve anonimizacije prikazan je u tablici Tabela 2.

Tabela 2. anonimizirani JMBAG iz kojeg je moguće saznati ime fakulteta ali ne i studenta

Ime studenta	JMBAG
Student 1	0036xxxxxx
Student 2	0036xxxxxx
Student 3	0036xxxxxx
Student 4	0036xxxxxx
Student 5	0016xxxxxx
Student 6	0016xxxxxx

Želimo li u navedeni skup podataka anonimizirati na način da se iz njega može zaključiti koje godine je pojedini student upisao studij, tada je potrebno anonimizirati samo zadnje 3 znamenke studentovog JMBAG-a. Opisani način ponovo će zadovoljiti uvijete *K-anonimnosti*. Naime, ponovo je kao i u prethodnom primjeru osigurano da se iz anonimiziranog podatka ne može zaključiti o kojem se točno studentu radi čak niti ukoliko se otkrije koji se skup podataka krije iza anonimiziranog parametra već svaki student s jednakom vjerojatnošću ponovo posjeduje bilo koji parametar iz obuhvaćenog pri anonimizaciji. S druge strane, ovako anonimizirani parametar u odnosu na prethodni primjer dodatno omogućava da se zaključi koje godine je student upisao studij. Opisani primjer prikazan je u tablici Tabela 3.

Tabela 3. anonimizirani JMBAG iz kojeg je moguće saznati ime fakulteta i godinu upisa studija

Ime studenta	JMBAG
Student 1	0036457xxx
Student 2	0036457xxx
Student 3	0036468xxx
Student 4	0036468xxx
Student 5	0016110xxx
Student 6	0016110xxx

2.2 SNAP - primjer anonimizacije

Sveučilište Stanford u sklopu svog projekta SNAP (*Stanford Network Analysis Project*) objavili su anonimizirani skup podataka, te će spomenuti primjer biti analiziran u nastavku. Način na koji je Sveučilište Stanford vršilo anonimizaciju je zamjena identifikacijskog broja (ID) korisnika društvene mreže Facebook s proizvoljnom jedinstvenom vrijednosti. Također osobine (engl. *features*) koje korisnici posjeduju su anonimizirane na način da se parametar koji predstavlja, npr. političko opredjeljenje korisnika (npr. political = Democratic Party) zamijeni s political=*anonymized feature 1* [5]. Tako anonimizirane osobine zapisane su u posebnu datoteku nazvanu *featnames*, a dio njenog sadržaja je prikazan na slici Slika 1. Sa slike vidimo kako je svaka anonimizirana osobina numerirana što služi kako bi se kasnije pojedina osobina mogla pridružiti nekom korisniku. U navedenoj datoteci nalaze se sve anonimizirane osobine skupa podataka koji se želi anonimizirati.

```

1 0 birthday;anonymized feature 0
2 1 birthday;anonymized feature 1
3 2 birthday;anonymized feature 2
4 3 education;classes;id;anonymized feature 3
5 4 education;classes;id;anonymized feature 4
6 5 education;classes;id;anonymized feature 5
7 6 education;classes;id;anonymized feature 6
8 7 education;classes;id;anonymized feature 7
9 8 education;concentration;id;anonymized feature 8
10 9 education;concentration;id;anonymized feature 9
11 10 education;concentration;id;anonymized feature 10
12 11 education;concentration;id;anonymized feature 11
13 12 education;degree;id;anonymized feature 12
14 13 education;degree;id;anonymized feature 13
15 14 education;school;id;anonymized feature 14

```

Slika 1. dio sadržaja datoteke *featnames*

Sljedeća datoteka koju Sveučilište Stanford koristi je datoteka koja nosi naziv *feat*. U navedenoj datoteci nalaze se ranije spomenuti proizvoljni jedinstveni identifikatori koji predstavljaju korisnika neke društvene mreže (npr. *Facebook*). Slika Slika 2 prikazuje dio sadržaja datoteke *feat*, te kao što možemo vidjeti sa slike identifikator svih promatranih korisnika društvene mreže zamijenjen je brojevima koji počinju od 0, tako da nema nikakve povezanosti s identifikatorom tog korisnika na društvenoj mreži. Sa slike također vidimo kako je svakom identifikatoru pridružen brojevni niz koji se sastoji samo od brojeva 0 i 1. Taj niz nula i jedinica označava koje osobine korisnik predstavljen anonimiziranim identifikatorom posjeduje iz datoteke *featnames*. Točnije, ukoliko korisnik predstavljen anonimiziranim identifikatorom posjeduje osobinu koja se u datoteci *featnames* prikazanoj prethodnom slikom nalazi na prvom mjestu, tada će i brojevni niz pridružen tom korisniku početi brojem 1, dok će u suprotnom početi brojem 0. Dakle redni broj u nizu povezan je s rednim brojem osobine u datoteci *featnames*, a taj broj ima vrijednost 1 ukoliko korisnik posjeduje osobinu koja se nalazi na tom rednom broju ili 0 ukoliko ju ne posjeduje.

1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	3	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	5	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	6	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	7	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Slika 2. dio sadržaja datoteke *feat*

Posljednja datoteka koja se koristi pri anonimizaciji zove se *edges*. Navedena datoteka, odnosno njen dio, prikazan je slikom Slika 3. Vidimo da svaki redak sadrži zapis o tome koja dva korisnika su povezana na društvenoj mreži.

1	1	3
2	1	5
3	1	6
4	2	3
5	2	7
6	4	5
7	4	6
8	5	7

Slika 3. dio sadržaja datoteke *edges*

Ukoliko želimo ranije navedene podatke o studentima anonimizirati poput Sveučilišta Stanford, tada možemo razlike među osobinama studenata različitih fakulteta anonimizirati tako da se ne zna o koji je točno fakultet pridružen pojedinom studentu. Tada je imena fakulteta potrebno anonimizirati na način da svaki fakultet predstavlja jedan anonimizirani parametar (npr FER = *Anonimizirana osobina 1*, FOI = *Anonimizirana osobina 2*). Provedbom navedenog postupka, tablica zapisa o studentima prikazana je u tablici Tabela 4. anonimizirani podaci o studentima.

Tabela 4. anonimizirani podaci o studentima

Ime studenta	Ime fakulteta
Student 1	Anonimizirana osobina 1
Student 2	Anonimizirana osobina 1
Student 3	Anonimizirana osobina 1
Student 4	Anonimizirana osobina 1
Student 5	Anonimizirana osobina 2
Student 6	Anonimizirana osobina 2

Podatke iz tablice zatim treba razdvojiti u dvije datoteke, od kojih bi u prvoj bili zapisi o osobinama studenata (odnosno anonimizirana imena fakulteta), a u drugoj bi jedinstvenim identifikatorima koji predstavljaju studente bio pridružen niz brojeva kojim bi se oni povezali s osobinama iz druge datoteke, na način koji je opisan ranije.

2.3 Anonimizacija geolokacijskih podataka

Kao što je rečeno u uvodu, anonimizacija geolokacijskog položaja zahtijeva poseban pristup, jer je više od 80% ljudi moguće je jedinstveno identificirati pomoću njegove 3 lokacije na kojima najčešće boravi. Još jedan problem koji je potrebno riješiti je problem pravilne anonimizacije dvije geolokacijski bliske pozicije kako bi se one mogle kasnije uspoređivati i omogućiti vršenje analize nad takvim podacima. Točnije, iz usporedbe više anonimiziranih geolokacijskih podataka mora se moći zaključiti odnos udaljenosti među njima, odnosno detektirati koji anonimizirani podaci su bliži jedni drugima a koji udaljeniji.

U literaturi se spominju dva načina geolokacijske anonimizacije, a to su [3]:

- Top Regions Model TRm (regije),
- Top Venues Model TVm (mjesta).

Međutim ovi primjeri samo su teoretski opisani, ali su bili dovoljni da mogu donijeti vlastite ideje za predlaganje anonimizacije navedenih geolokacijskih podataka.

U nastavku su opisani prijedlozi anonimizacije geolokacijskih podataka.

Svi predloženi postupci temelje se na korištenju API-ja koji nudi Google, točnije GoogleMaps kako bi bilo moguće odrediti koordinate navedenih lokacija a samim time i provesti opisane postupke.

Prijedlozi anonimizacije geolokacijskih podataka:

1. Prvi prijedlog temelji se na tome da se za svaku lokaciju korisnika odrede njene koordinate. Lokacije se zatim anonimiziraju te se pomoću koordinata odrede udaljenosti između svih anonimiziranih lokacija. Na kraju je potrebno anonimizirane parametre svrstatи u intervale koji predstavljaju udaljenost između neke dvije lokacije. Posljednji korak algoritma potreban je kako se ne bi znala točna udaljenost između neke dvije lokacije, već samo

interval unutar kojeg se nalazi vrijednost koja predstavlja udaljenost. Prikazivanje točne udaljenosti dvije lokacije bi znatno olakšalo otkrivanje o kojim lokacijama se radi.

Prethodni algoritam pokazat ćeemo na primjeru. Uzmimo npr. 3 grada Republike Hrvatske, neka to budu Zagreb, Varaždin i Split. Za svaki grad potrebno je najprije odrediti njegove koordinate koje će kasnije služiti za određivanje međusobne udaljenosti. Tablica Tabela 5 prikazuje gradove i njima pridružene koordinatne točke.

Tabela 5. pridruživanje koordinatnih točaka svakoj lokaciji

lokacija	Sjever (engl. North)	Istok (engl. East)
Zagreb	45°48'43"	15°58'52"
Varaždin	46°18'29"	16°20'33"
Split	43°30'29"	16°26'40"

Nazive lokacija (gradova) sada je potrebno anonimizirati, na način da svaki grad predstavlja jedan anonimizirani parametar (npr. Zagreb = *Anonimizirana lokacija 1*). U tablici Tabela 6 prikazana je provedba opisanog koraka algoritma.

Tabela 6. anonimizacija imena lokacije

lokacija	Sjever (engl. North)	Istok (engl. East)
Anonimizirana lokacija 1	45°48'43"	15°58'52"
Anonimizirana lokacija 2	46°18'29"	16°20'33"
Anonimizirana lokacija 3	43°30'29"	16°26'40"

Uz pomoću koordinatnih točaka svake lokacije možemo izračunati njihove međusobne udaljenosti, kako je prikazano u tablici Tabela 7.

Tabela 7. međusobne udaljenosti među lokacijama

lokacija	Anonimizirana lokacija 1	Anonimizirana lokacija 2	Anonimizirana lokacija 3
Anonimizirana lokacija 1	0	61 km	260 km
Anonimizirana lokacija 2	61 km	0	310 km
Anonimizirana lokacija 3	260 km	310 km	0

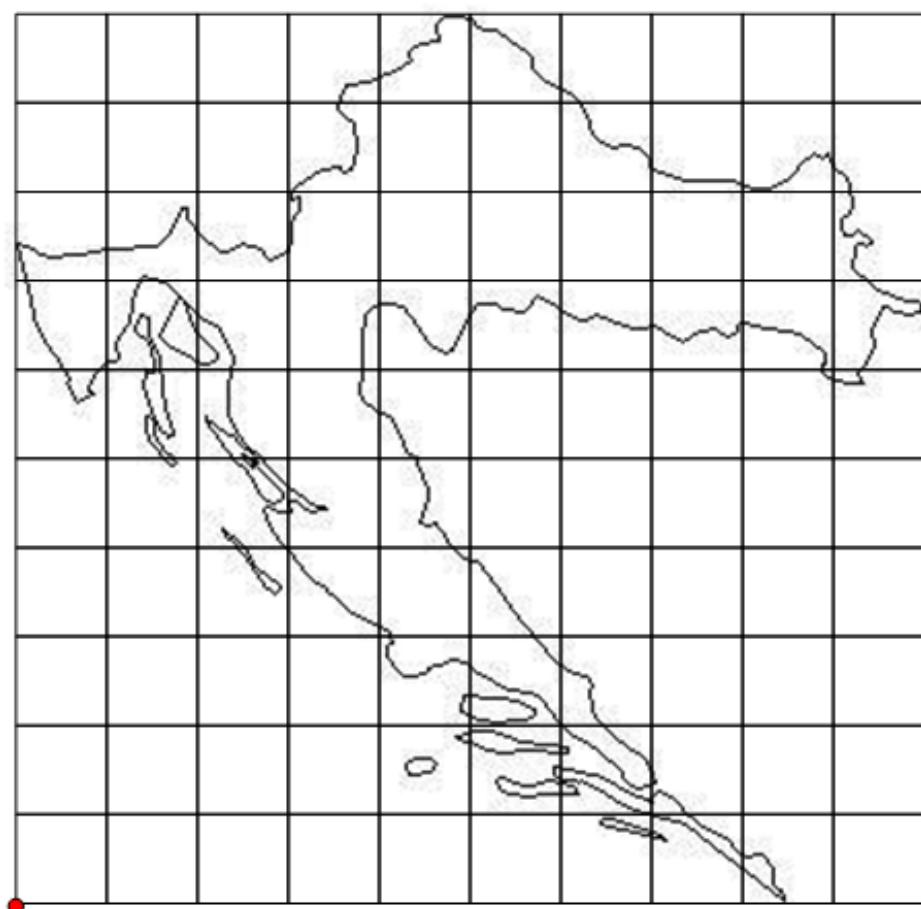
Za kraj je potrebno udaljenosti među gradovima predstaviti intervalima te svrstati koji grad se nalazi u kojem intervalu, kako je prikazano u tablici Tabela 8.

Tabela 8. prvi prijedlog anonimizacije geolokacijskog položaja

lokacija	<50 km	<100 km	<500 km
Anonimizirana lokacija 1	-	Anonimizirana lokacija 2	Anonimizirana lokacija 3
Anonimizirana lokacija 2	-	Anonimizirana lokacija 1	Anonimizirana lokacija 3
Anonimizirana lokacija 3	-	-	Anonimizirana Lokacija1, Anonimizirana lokacija 2

2. Geografski prostor koji promatramo možemo podijeliti na mrežu koja se sastoji od kvadrata željenih dimenzija. Na ovaj način možemo odrediti međusobne položaje tih kvadrata, a dva grada koja su unutar istog kvadrata anonimizirani su istim parametrom što nam odgovara jer ti gradovi i jesu međusobno vrlo malo udaljeni i moguće je vršiti ispravnu analizu nad podacima. Još jedna prednost je da ovaj način zadovoljava načela *Kanonimizacije*, jer unutar jednog takvog kvadrata može biti više gradova ili mjesta koja predstavljaju lokaciju, te ukoliko se i sazna koje točne geografske granice predstavlja anonimizirani kvadrat nije moguće točno reći o kojem se gradu iz tog kvadrata radi. Slika 4 prikazuje geografski prostor Republike Hrvatske podijeljen na mrežu kvadrata. Za podjelu geografskog prostora na mrežu potrebno je odrediti najjužniju i najzapadniju točku države koju želimo podijeliti na mrežu, a zatim sjecište geografske širine koja prolazi najzapadnjom i dužine koja prolazi najjužnjom koordinatnom točkom postaviti za ishodišnu točku promatranog geografskog područja. Spomenuta točka označena je crveno na slici Slika 4. Kada imamo definiranu ishodišnu točku, potrebno je odrediti duljinu stranice kvadrata, te će o tome ovisiti i ukupni broj kvadrata na koji je podijeljen geografski prostor. Koordinate korisnikove

lokacije određuju se na način opisan u prvom prijedlogu, a poznavajući ishodišnu točku i dimenziju kvadrata za svaku lokaciju možemo izračunati u kojem kvadratu se ona nalazi. Zadnji korak koji je potrebno napraviti je izračunati koliko su međusobno udaljeni kvadri, promatrajući međusobnu udaljenost koordinatnih točaka koje predstavljaju donji lijevi kut kvadrata, a potom te kvadrate anonimizirati i prikazati njihov međusobni odnos u tablici, na način sličan prvom opisanom prijedlogu.



Slika 4. promatrano područje podijeljeno na mrežu kvadrata

3. Državu čiji geografski prostor se analizira moguće je podijeliti na županije i općine, odnosno na administrativna područja raznih razina (administrative_area_level_x). Administrativno područje za neku lokaciju moguće je dohvatiti pomoću korištenja API-ja koje nudi Google (GoogleMaps). Treća ideja stoga je svaku lokaciju smjestiti u pripadnu županiju/administrativno područje, a zatim to područje anonimizirati. Administrativna područja Republike Hrvatske prikazana su na slici Slika 5.



Slika 5. administrativna područja Republike Hrvatske [6].

Kao u prethodnom primjeru ova metoda također zadovoljava uvijete *k-anonimizacije* jer se više korisničkih lokacija može nalaziti unutar istog administrativnog područja. U tablici

Tabela 9 na prikazan je postupak anonimizacije 3 odabrana administrativna područja na kojima će biti opisan daljnji postupak algoritma anonimizacije.

Tabela 9. anonimizacija administrativnih područja

Oznaka administrativnog područja - Županije	Ime administrativnog područje - Županije	Anonimizirano administrativno područje
XXI.	Grad Zagreb	Anonimizirano područje 1
I.	Zagrebačka	Anonimizirano područje 2
IV.	Karlovačka	Anonimizirano područje 3

Za anonimizirane podatke iz gornje tablice, nadalje je potrebno odrediti međusobni odnos lokacija koje oni predstavljaju. Točnije, potrebno je odrediti koja administrativna područja su direktno međusobno povezana. Navedeni odnosi prikazani su tablicom Tabela 10.

Tabela 10. međusobni odnos anonimiziranih administrativnih područja

Anonimizirano administrativno područje	Susjedno administrativno područje
Anonimizirano područje 1	Anonimizirano područje 2
Anonimizirano područje 2	Anonimizirano područje 1, Anonimizirano područje 3
Anonimizirano područje 3	Anonimizirano područje 2

Analizom podataka prikazanih u prethodnoj tablici možemo odrediti odnos između administrativnih područja, odnosno pogledamo li drugi redak tablice, vidimo da je *Anonimizirano područje 1* direktno povezano s *Anonimiziranim područjem 2*, a kako iz sljedećeg retka saznajemo da je *Anonimizirano područje 2* direktno povezano još i sa *Anonimiziranim područjem 3*, tada vrlo lako zaključujemo da je iz *Anonimiziranog područja 1* u *Anonimizirano područje 3* povezano *Anonimiziranim područjem 2*.

3. Usporedba predloženih postupaka za anonimizaciju geolokacijskih podataka

Kako bi mogli odabratи najprikladniju metodu, od prethodno ponuđenih, za provedbu anonimizacije nad odabranim skupom podataka potrebno je navedene najprije usporediti te utvrditi koje su prednosti i nedostaci svake od metoda.

Glavni izazov koji je postavljen pred prethodno opisane metode je bio anonimizirati određene lokacije na način da je moguće usporediti njihovu međusobnu udaljenost, a spomenuti izazov sve 3 metode uspješno svladavaju.

Objasnili smo koja je važnost da skup anonimiziranih podataka zadovoljava načela *K-anonimizacije*, a od predložene 3 metode, samo prva ne zadovoljava to načelo. Naime u prvoj metodi se svaka lokacija zasebno anonimizira, dok se u preostale dvije metode anonimizira veće područje (pravokutnik u mreži ili administrativno područje neke države).

Usporedimo li kompleksnost same izvedbe predložena 3 postupka, zaključujemo sljedeće:

- prva predložena metoda posjeduje potencijalni problem. Naime, kako se u navedenoj za svaki par anonimiziranih lokacija računa njihova udaljenost, tada se problem zapravo svodi na računanje duljine svih dijagonala nekog n -terokuta (gdje je n broj lokacija). Broj dijagonala mnogokuta (n -terokuta) predstavlja broj svih međusobnih udaljenosti među lokacijama, a računa se formulom:

$$N = \frac{1}{2}n(n - 3)$$

Možemo uočiti da je složenost računanja broja dijagonala jednak $O(n^2)$ što predstavlja potencijalni problem ukoliko je potrebno anonimizirati veliki broj lokacija, zbog toga što broj udaljenosti među lokacijama koje je potrebno izračunati raste s kvadratnom potencijom u odnosu na sam broj lokacija.

- drugi prijedlog unutar jednog kvadrata smješta više lokacija. Stoga je najprije potrebno izračunati granice kvadrata u mreži, a zatim izračunati međusobne udaljenosti između svih kvadrata. Navedenim postupkom postiže se

smanjenje ukupnog broja udaljenosti koje je potrebno izračunati, ali složenost računanja i dalje ostaje $O(n^2)$.

- zadnji prijedlog također unutar područja koje anonimizira smješta više lokacija. Kako treći prijedlog kao područje koje anonimizira koristi županije (administrativna područja) čije su granice jasno definirane na razini svake države, tada nije potrebno provoditi nikakva dodatna računanja vezana uz određivanje tog područja, već je za svako područje korištenjem API-ja koje nudi Google moguće odrediti u koje administrativno područje ono pripada. Također u navedenom algoritmu nije potrebno računati koliko su područja koja se anonimiziraju međusobno udaljena, jedino što je potrebno jest odrediti koja područja su međusobno susjedna.

Tablica Tabela 11 prikazuje usporedbu predloženih metoda po ranije opisanim kriterijima. Kao što se iz tablice može vidjeti, treći prijedlog je najbolje zadovoljava sve postavljene kriterije, te će navedeni biti odabran za izvedbu anonimizacije nad odabranim skupom podataka.

Tabela 11. usporedba predloženih algoritama

Metoda	Usporedba udaljenosti	K-anonimizacija	Optimalna kompleksnost određivanja odnosa među lokacijama
Prijedlog broj 1	+	-	-
Prijedlog broj 2	+	+	-
Prijedlog broj 3	+	+	+

4. Anonimizacija geolokacijskog područja odabranim postupkom

Usporedbom predloženih metoda u prethodnom poglavlju prijedlog broj 3 pokazao se najpogodniji za provedbu anonimizacije na odabranom skupu podataka, te će izvedba istog biti prikazana u nastavku.

Za provedbu prijedloga broj 3 najprije je potrebno za svakom administrativnom području neke države (Slika 5. iz poglavlja 2.3) definirati anonimizirani naziv, kao što je prikazano u tablici Tabela 12. Poželjno je da se broj anonimiziranog područja (npr. Anonimizirano područje 1 iz tablice) ne podudara a oznakom administrativnog područja (npr I. iz tablice), međutim, kako bi bilo lakše pratiti postupak u nastavku, ti brojevi će se podudarati.

Tabela 12. anonimizirani nazivi pridruženi administrativnim područjima

Oznaka administrativnog područja - Županije	Ime administrativnog područje - Županije	Anonimizirano administrativno područje
I.	Zagrebačka	Anonimizirano područje 1
II.	Krapinsko-zagorska	Anonimizirano područje 2
III.	Sisačko-moslavačka	Anonimizirano područje 3
IV.	Karlovačka	Anonimizirano područje 4
V.	Varaždinska	Anonimizirano područje 5
VI.	Koprivničko-križevačka	Anonimizirano područje 6
VII.	Bjelovarsko-bilogorska	Anonimizirano područje 7
VIII.	Primorsko-goranska	Anonimizirano područje 8
IX.	Ličko-senjska	Anonimizirano područje 9
X.	Virovitičko-podravska	Anonimizirano područje 10
XI.	Požeško-slavonska	Anonimizirano područje 11
XII.	Brodsko-posavska	Anonimizirano područje 12
XIII.	Zadarska	Anonimizirano područje 13
XIV.	Osječko-baranjska	Anonimizirano područje 14
XV.	Šibensko-kninska	Anonimizirano područje 15
XVI.	Vukovarsko-srijemska	Anonimizirano područje 16
XVII.	Splitsko-dalmatinska	Anonimizirano područje 17
XVIII.	Istarska	Anonimizirano područje 18
IXX.	Dubrovačko-neretvanska	Anonimizirano područje 19
XX.	Međimurska	Anonimizirano područje 20
XXI.	Grad Zagreb	Anonimizirano područje 21

Nadalje, potrebno je administrativna područja grupirati na način da se u prvom stupcu prikažu brojevne oznake administrativnog područja, a u drugom stupcu brojevne oznake administrativnih područja koje neposredno povezanih s administrativnim područjem iz prvog stupca. Rezultat opisanog postupka za administrativna područja Republike Hrvatske prikazan je u tablici Tabela 13.

Tabela 13. povezivanje susjednih administrativnih područja

Oznaka administrativnog područja - Županije	Oznake susjednih administrativnih područja
I.	II, III, IV, V, VI, VII, XXI
II.	I, V, XXI
III.	I, IV, VII, XI, XII
IV.	I, III, VIII, IX
V.	I, II, VI, XX
VI.	I, V, VII, X, XX
VII.	I, III, VI, X, XI
VIII.	IV, IX, XVIII
IX.	IV, VIII, XIII
X.	VI, VII, XI, XIV
XI.	III, VII, X, XII, XIV
XII.	III, XI, XIV, XVI
XIII.	IX, XV
XIV.	X, XI, XII, XVI
XV.	XIII, XVII
XVI.	XII, XIV
XVII.	XV, XIX
XVIII.	VIII
IXX.	XVII
XX.	V, VI
XXI.	I, II

Prethodnu tablicu potrebno je prikazati na anonimizirani način, tako de se svaka oznaka administrativnog područja prikaže njenim pridruženim anonimiziranim nazivom koji smo definirali ranije, u tablici Tabela 12. Rezultat provedene anonimizacije je tablica Tabela 14.

Tabela 14. anonimizacija grupiranih administrativnih područja

Anonimizirano administrativno područje	Susjedno administrativno područje
Anonimizirano područje 1	Anonimizirano područje 2, Anonimizirano područje 3, Anonimizirano područje 4, Anonimizirano područje 5, Anonimizirano područje 6, Anonimizirano područje 7, Anonimizirano područje 21
Anonimizirano područje 2	Anonimizirano područje 1, Anonimizirano područje 5, Anonimizirano područje 21
Anonimizirano područje 3	Anonimizirano područje 1, Anonimizirano područje 4, Anonimizirano područje 7, Anonimizirano područje 11, Anonimizirano područje 12
Anonimizirano područje 4	Anonimizirano područje 1, Anonimizirano područje 3, Anonimizirano područje 8, Anonimizirano područje 9
Anonimizirano područje 5	Anonimizirano područje 1, Anonimizirano područje 2, Anonimizirano područje 6, Anonimizirano područje 20
Anonimizirano područje 6	Anonimizirano područje 1, Anonimizirano područje 5, Anonimizirano područje 7, Anonimizirano područje 10, Anonimizirano područje 20
Anonimizirano područje 7	Anonimizirano područje 1, Anonimizirano područje 3, Anonimizirano područje 6, Anonimizirano područje 10, Anonimizirano područje 11
Anonimizirano područje 8	Anonimizirano područje 4, Anonimizirano područje 9, Anonimizirano područje 18
Anonimizirano područje 9	Anonimizirano područje 4, Anonimizirano područje 8, Anonimizirano područje 13
Anonimizirano područje 10	Anonimizirano područje 6, Anonimizirano područje 7, Anonimizirano područje 11, Anonimizirano područje 14
Anonimizirano područje 11	Anonimizirano područje 3, Anonimizirano područje 7, Anonimizirano područje 10, Anonimizirano područje 12, Anonimizirano područje 14
Anonimizirano područje 12	Anonimizirano područje 3, Anonimizirano područje 11, Anonimizirano područje 14, Anonimizirano područje 16
Anonimizirano područje 13	Anonimizirano područje 9, Anonimizirano područje 15
Anonimizirano područje 14	Anonimizirano područje 10, Anonimizirano područje 11, Anonimizirano područje 12, Anonimizirano područje 16
Anonimizirano područje 15	Anonimizirano područje 13, Anonimizirano područje 17
Anonimizirano područje 16	Anonimizirano područje 12, Anonimizirano područje 14
Anonimizirano područje 17	Anonimizirano područje 15, Anonimizirano područje 19
Anonimizirano područje 18	Anonimizirano područje 8
Anonimizirano područje 19	Anonimizirano područje 17
Anonimizirano područje 20	Anonimizirano područje 5, Anonimizirano područje 6
Anonimizirano područje 21	Anonimizirano područje 1, Anonimizirano područje 2

Anonimizirana i grupirana administrativna područja iz Tabela 14 zatim je potrebno zapisati u datoteku, koja nosi ime *geolokacija.txt*. Sa slike Slika 6 vidimo da svaki redak datoteke sadrži anonimiziranu oznaku administrativnog područja, te brojevni niz načinjen od brojeva 0 i 1. Za promatrano administrativno područje, redni broj u znamenke u nizu povezan je s rednim brojem anonimiziranog administrativnog područja, te ukoliko je npr peta znamenka u nizu jednaka 1, to označava da je promatrano administrativno područje direktno povezano s područjem koje je anonimizirano parametrom iz retka s rednim brojem 5.

Slika 6. prikaz sadržaja datoteke *geolokacija.txt*

Uz pomoć datoteke *geolokacija.txt*, programskog koda koji koristi Google API, i uz pomoć tablice Tabela 12 u možemo izvršiti konačan postupak anonimizacije na konkretnom primjeru.

Prepostavimo da imamo 3 osobe i pridruženi im skup podataka koji ih opisuje, te on sadrži barem jedan zapis o lokaciji te osobe kao npr:

- Osoba 1 - Varaždin,
 - Osoba 2 - Koprivnica,
 - Osoba 3 - Zagreb.

Spomenutim programskim kodom za svaku korisničku lokaciju možemo dohvatiti administrativno područje u kojem se ta lokacija nalazi. Rezultat izvršavanja programskog koda na primjeru Varaždina kao korisnikove lokacije dobiven je u JSON (*JavaScript Object Notation*) i prikazan je na slici Slika 7.

```
"address_components" : [
  {
    "long_name" : "Varaždin",
    "short_name" : "Varaždin",
    "types" : [ "locality", "political" ]
  },
  {
    "long_name" : "Općina Varaždin",
    "short_name" : "Općina Varaždin",
    "types" : [ "administrative_area_level_2", "political" ]
  },
  {
    "long_name" : "Varaždin County",
    "short_name" : "Varaždin County",
    "types" : [ "administrative_area_level_1", "political" ]
  }
]
```

Slika 7. rezultat pretrage lokacije u JSON formatu

Vidimo da tip podataka `administrative_area_level_1` ima definiran naziv (`long_name`), a njegova vrijednost je: "Varaždin County" što predstavlja Varaždinsku županiju. Iz tablice Tabela 12 možemo uočiti kako je navedenoj županiji pridružen anonimizirani naziv "Anonimizirano područje 5", a samim time zaključujemo da Osoba 1 ima pridružen navedeni parametar za svoju lokaciju. Sličnim postupkom možemo doći do podatka da Osoba 2 ima pridružen parametar "Anonimizirano područje 6", a Osoba 3 "Anonimizirano područje 21".

Koristeći datoteku `geolokacija.txt` navedene parametre možemo uspoređivati, naime sa slike Slika 8 vidimo da kod "Anonimiziranog područja 5" je šesta znamenka u nizu postavljena na 1 što znači da je ono direktno povezano s "Anonimiziranim područjem 6" ali ne i sa " Anonimiziranim područjem 21" jer je posljednja znamenka niza postavljena na 0. Zaključujemo da su Osoba 1 i Osoba 2 lokacijski bliže nego Osoba 1 i Osoba 3.

4	Anonimizirano područje	4 1 0 1 0 0 0 0 1 1 0
5	Anonimizirano područje	5 1 1 0 0 0 1 0 1 0
6	Anonimizirano područje	6 1 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
7	Anonimizirano područje	7 1 0 1 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8	Anonimizirano područje	8 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0

Slika 8. određivanje susjednih lokacija

Za određivanje koliko su udaljene lokacije Osobe 1 i Osobe 3 potrebno je izvršiti detaljniju analizu datoteke *geolokacija.txt*. Naime analiza se svodi na određivanje preko kojih susjednih područja se može doći do područja u kojem se nalazi Osoba 3. Analizom bi se moglo zaključiti da je lokacija Osobe 1 direktno povezana s "Anonimiziranim područjem 1", a navedeno područje je direktno povezano s "Anonimiziranim područjem 21", kao što je prikazano slikom Slika 9.

1	Anonimizirano područje	1 0 1 1 1 1 1 1 0 1
2	Anonimizirano područje	2 1 0 0 0 1 0 1
3	Anonimizirano područje	3 1 0 0 1 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4	Anonimizirano područje	4 1 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5	Anonimizirano područje	5 1 0 0 0 1 0 1 0
6	Anonimizirano područje	6 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0

Slika 9. određivanje udaljenosti među lokacijama koje nisu direktno povezane

Opisani algoritam uspješno anonimizira stvarnu korisničku lokaciju, ali istovremeno pruža mogućnost da se nad tako anonimiziranim skupom podataka vrše razne analize na temelju njihove udaljenosti.

5. Poboljšanje odabranog postupka za anonimizaciju geolokacijskih podataka

Metoda opisana u prethodnom poglavlju ima nekoliko nedostataka odnosno propusta. Naime, pogledom u tablicu Tabela 15, moguće je uočiti da samo jedna županija ima 7 susjednih županija, stoga ukoliko znamo da se radi o područjima Republike Hrvatske, prethodnom metodom anonimizacije može se odmah zaključiti da to područje predstavlja Zagrebačku županiju. Također, dekodiranje pojedinog područja se može odrediti prema broju stanovnika koji je pridružen nekom anonimiziranim području. Pogledamo li ponovo spomenutu tablicu vidimo da Grad Zagreb jako odskače od ostalih županija po broju stanovnika [7].

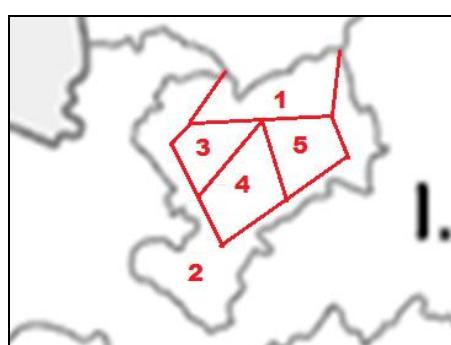
Tabela 15. broj stanovnika i broj susjednih županija

Ime administrativnog područja - Županije	Broj stanovnika	Broj susjednih područja
Grad Zagreb	790 017	2
Splitsko-dalmatinska	454 798	2
Zagrebačka	317 606	7
Osječko-baranjska	305 032	4
Primorsko-goranska	296 195	3
Istarska	208 055	1
Vukovarsko-srijemska	179 521	2
Varaždinska	175 951	4
Sisačko-moslavačka	172 439	5
Zadarska	170 017	2
Brodsko-posavska	158 575	4
Krapinsko-zagorska	132 892	3
Karlovačka	128 899	4
Dubrovačko-neretvanska	122 568	1
Bjelovarsko-bilogorska	119 764	5
Koprivničko-križevačka	115 584	5
Međimurska	113 804	2
Šibensko-kninska	109 375	2
Virovitičko-podravska	84 836	4
Požeško-slavonska	78 034	5
Ličko-senjska	50 927	3

Rješenje ranije spomenutog problema je provedba "zamišljene" podjele pojedinih županija, time bi se Grad Zagreb mogao podijeliti na više "zamišljenih" područja od kojih svako područje predstavlja Grad Zagreb. Spomenutim bi se početni broj stanovnika jednoliko podijelio na više "zamišljenih" područja. Također pravilnom manipulacijom podjele moguće je ujednačiti broj susjednih područja neke županije i time bi bilo nemoguće jednoznačno dekodirati pojedino anonimizirano područje. Slika Slika 10 prikazuje jednu od mogućih podjela županija, dok slika Slika 11, radi bolje preglednosti, prikazuje podjelu Grada Zagreba na zamišljena područja.



Slika 10. podjela županija na "zamišljena" područja



Slika 11. podjela Grada Zagreba na zamišljena područja

Tablica Tabela 16 prikazuje broj stanovnika po županijama i broj susjednih područja pojedine županije nakon podjele.

Tabela 16. broj stanovnika i susjednih područja nakon podjele

Ime administrativnog područja - Županije	Br. stanovnika	Br. stanovnika nakon podjele	Br. susjednih područja
Grad Zagreb	790 017	158 004	5
		158 004	5
		158 003	4
		158 003	4
		158 003	4
Splitsko-dalmatinska	454 798	151 600	4
		151 599	3
		151 599	2
Zagrebačka	317 606	158 803	5
		158 803	4
Osječko-baranjska	305 032	152 516	5
		152 516	2
Primorsko-goranska	296 195	148 098	4
		148 097	4
Istarska	208 055	104 028	2
		104 027	3
Vukovarsko-srijemska	179 521	179 521	3
Varaždinska	175 951	175 951	5
Sisačko-moslavačka	172 439	172 439	5
Zadarska	170 017	170 017	2
Brodsko-posavska	158 575	158 575	5
Krapinsko-zagorska	132 892	44 298	5
		44 297	3
		44 297	2
Karlovачka	128 899	42 967	4
		42 966	3
		42 966	3
Dubrovačko-neretvanska	122 568	122 568	2
Bjelovarsko-bilogorska	119 764	119 764	5
Koprivničko-križevačka	115 584	115 584	5
Međimurska	113 804	113 804	2
Šibensko-kninska	109 375	109 375	2
Virovitičko-podravska	84 836	42 418	4
		42 418	3
Požeško-slavonska	78 034	39 017	5
		39 017	5
Ličko-senjska	50 927	50.927	3

Vidimo da je sada broj stanovnika po županijama ujednačen, te broj susjednih županija nekog područja varira između 2 i 5, time niti jedna županija nije jednoznačno određena opisanim parametrima. Županije razdvojene na zamišljena područja prikazane se u tablici Tabela 17.

Tabela 17 podjela županija na "zamišljena" područja

Ime administrativnog područja - Županije	Br. stanovnika nakon podjele	Br. susjednih područja
Grad Zagreb 1	158 004	5
Grad Zagreb 2	158 004	5
Grad Zagreb 3	158 003	4
Grad Zagreb 4	158 003	4
Grad Zagreb 5	158 003	4
Splitsko-dalmatinska 1	151 600	4
Splitsko-dalmatinska 2	151 599	3
Splitsko-dalmatinska 3	151 599	2
Zagrebačka 1	158 803	5
Zagrebačka 2	158 803	4
Osječko-baranjska 1	152 516	5
Osječko-baranjska 2	152 516	2
Primorsko-goranska 1	148 098	4
Primorsko-goranska 2	148 097	4
Istarska 1	104 028	2
Istarska 2	104 027	3
Vukovarsko-srijemska	179 521	3
Varaždinska	175 951	5
Sisačko-moslavačka	172 439	5
Zadarska	170 017	2
Brodsko-posavska	158 575	5
Krapinsko-zagorska 1	44 298	5
Krapinsko-zagorska 2	44 297	3
Krapinsko-zagorska 3	44 297	2
Karlovačka 1	42 967	4
Karlovačka 2	42 966	3
Karlovačka 3	42 966	4
Dubrovačko-neretvanska	122 568	2
Bjelovarsko-bilogorska	119 764	5
Koprivničko-križevačka	115 584	5
Međimurska	113 804	2
Šibensko-kninska	109 375	2
Virovitičko-podravska 1	42 418	4
Virovitičko-podravska 2	42 418	3
Požeško-slavonska 1	39 017	5
Požeško-slavonska 2	39 017	5
Ličko-senjska	50 927	3

Za kraj je područjima potrebno pridružiti njihov anonimizirani parametar i svakom području pridružiti anonimizirane nazine onih područja koja su mu susjedna, kao što je prikazano u tablici Tabela 18 (zbog jednostavnijeg prikaza anonimizirano područje označeno je s PX gdje X označava jedinstveni broj tog područja).

Tabela 18. anonimizacija podijeljenih područja

Ime administrativnog područja - Županije	Anonimizirano ime administrativnog područja - Županije	Susjedna područja
Grad Zagreb 1	P1	P2, P3, P4, P5, P22
Grad Zagreb 2	P2	P1, P3, P4, P5, P9
Grad Zagreb 3	P3	P1, P2, P4, P5
Grad Zagreb 4	P4	P1, P2, P3, P5
Grad Zagreb 5	P5	P1, P2, P3, P4
Splitsko-dalmatinska 1	P6	P7, P8, P32, P28
Splitsko-dalmatinska 2	P7	P6, P8, P28
Splitsko-dalmatinska 3	P8	P6, P7
Zagrebačka 1	P9	P10, P2, P18, P25, P22
Zagrebačka 2	P10	P9, P30, P29, P19
Osječko-baranjska 1	P11	P12, P17, P21, P36, P34
Osječko-baranjska 2	P12	P11, P17
Primorsko-goranska 1	P13	P14, P26, P15, P16
Primorsko-goranska 2	P14	P13, P16, P27, P37
Istarska 1	P15	P16, P13
Istarska 2	P16	P15, P13, P14
Vukovarsko-srijemska	P17	P11, P12, P21
Varaždinska	P18	P31, P30, P9, P22, P23
Sisačko-moslavačka	P19	P25, P10, P29, P35, P21
Zadarska	P20	P37, P32
Brodsko-posavska	P21	P19, P35, P36, P11, P17
Krapinsko-zagorska 1	P22	P23, P24, P18, P9, P1
Krapinsko-zagorska 2	P23	P22, P24, P18
Krapinsko-zagorska 3	P24	P22, P23
Karlovačka 1	P25	P26, P27, P9, P19
Karlovačka 2	P26	P25, P27, P13
Karlovačka 3	P27	P25, P26, P14, P37
Dubrovačko-neretvanska	P28	P6, P7
Bjelovarsko-bilogorska	P29	P10, P30, P33, P35, P19
Koprivničko-križevačka	P30	P10, P29, P33, P18, P31
Međimurska	P31	P18, P30
Šibensko-kninska	P32	P6, P20
Virovitičko-podravska 1	P33	P34, P30, P29, P35, P36,
Virovitičko-podravska 2	P34	P33, P36, P11
Požeško-slavonska 1	P35	P36, P33, P29, P19, P21
Požeško-slavonska 2	P36	P35, P33, P34, P21, P11
Ličko-senjska	P37	P27, P14, P20

6. Zaključak

Velika popularnosti društvenih mreža i porast broja podataka koje opisuju korisnički profil, rezultiralo je time da su društvene mreže postale idealna podloga za izradu sve većeg broja aplikacija koje prikupljaju određene podatke o korisnicima kako bi na razne načine izračunale njegovu utjecajnost, najbolje prijatelje i slično. Razvoj spomenutih aplikacija omogućava lakše prikupljanje podataka o korisnicima društvenih mreža, a samim time povećavaju se mogućnosti analize istih. Prikupljeni korisnički podaci mogu se:

- zadržati za vlastitu provedbu analize,
- javno objaviti za detaljniju analizu.

Ukoliko se prikupljeni podaci zadrže za provedbu vlastitih istraživanja, biti će puno teže ili vrlo često i nemoguće provesti detaljniju analizu istih. Javno objavljivanje prikupljenih podataka omogućuje izvršavanje analize nad njima od strane većeg broja stručnih timova, a samim time analiza će biti detaljnija i kvalitetnija. Međutim, javno objavljivanje takvih podataka uzrokuje razne povrede privatnosti korisnika čiji su podaci objavljeni.

Rezultat ovog rada je analiza i prikaz raznih metoda i algoritama za anonimizaciju skupova podataka, kao i prikaz postojećeg rješenja, korištenog od strane Sveučilišta Stanford, za anonimizaciju velikog skupa podataka korisnika društvenih mreža. Također rezultat rada su 3 predložene metode za anonimizaciju geolokacijskog položaja, te usporedba predloženih metoda, a zatim i provedba same anonimizacije odabranom metodom.

7. Literatura

- [1] About social networking, <http://www.whatissocialnetworking.com/>, pristup: 02.05.2015.
- [2] Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, Siddharth Srivastava, "Anonymizing Social Networks", Technical Report No. 07-19, University of Massachusetts Amherst Computer Science Department, March 2008.
- [3] Amirreza Masoumzadeh, James Joshi, "Top Location Anonymization for Geosocial Network Datasets", School of Information Sciences, University of Pittsburgh, 2013.
- [4] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, Muthuramakrishnan Venkitasubramaniam, " ℓ -Diversity: Privacy Beyond k -Anonymity", Department of Computer Science, Cornell University.
- [5] Stanford Large Network Dataset Collection, <http://snap.stanford.edu/data/egonets-Facebook.html>, pristup 4.5.2015.
- [6] Slika županija preuzeta s: <http://www.croatia.eu/images/03-06/zupanje.gif>, pristup 16.5.2015.
- [7] Podaci o broju stanovnika po županijama preuzeti s: <http://www.croatia.eu/article.php?lang=1&id=30>, pristup 1.6.2015.

8. Sažetak

Izazovi privatnosti na društvenim mrežama: anonimizacija korisničkih podataka

U radu su opisani potreba za anonimizacijom skupa podataka o korisnicima društvenih mreže, te problemi i izazovi koje je potrebno riješiti prilikom izvedbe same anonimizacije. Poseban naglasak stavljen je na anonimizaciju geolokacijskog područja, te su u radu nalaze opisi triju predloženih metoda koje bi se mogle koristiti za provedbu anonimizacije nekog skupa podataka koji sadrži podatak o korisnikovoj lokaciji. Naposljetku je provedena usporedba predloženih metoda po definiranim kriterijima, te je, uz pomoć metode koja se pokazala najprikladnijom, dan prijedlog kako anonimizirati geolokacijske podatke za područje Republike Hrvatske.