

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Rok Gomišček

**Analiza slovenskega političnega  
prostora na Twitterju**

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM PRVE STOPNJE  
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Dejan Lavbič

Ljubljana 2013



Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

*Besedilo je oblikovano z urejevalnikom besedil  $\text{\LaTeX}$ .*





Št. naloge: 00109/2013

Datum: 11.04.2013

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **ROK GOMIŠČEK**

Naslov: **ANALIZA SLOVENSKEGA POLITIČNEGA PROSTORA NA TWITTERJU**  
**ANALYSIS OF SLOVENIAN POLITICS ACTIVITY ON TWITTER**

Vrsta naloge: Diplomsko delo univerzitetnega študija prve stopnje

Tematika naloge:

Stik z javnostjo in oglaševanje sta pomembna elementa predvolilnih kampanj, kjer je javnim osebam in politikom v zadnjem času na voljo vedno več komunikacijskih kanalov, podprtih s sodobno tehnologijo. Predvsem na družbenih omrežjih in forumih se zbirajo številni volivci, ki izražajo svoje mnenje in se opredeljujejo za njim najbližjo politično opcijo. Eno izmed takšnih omrežij je tudi Twitter, ki ga v zadnjem času uporablja vse več akterjev v tujem in tudi slovenskem političnem prostoru.

V okviru diplomske naloge naj študent na podlagi obstoječih raziskav napovedovanja volitev in odkrivanja mrež uporabnikov analizira slovenski politični prostor. Pri tem naj poskuša identificirati aktivnost strank v izbranem družabnem omrežju, vsebino zapisov ter morebitne korelacije vzorcev na spletu in v realnem svetu. Pri zajemu podatkov o političnem prostoru v Sloveniji se naj osredotoči na podatke o parlamentarnih strankah in poslancih na Twitterju.

Mentor:

doc. dr. Dejan Lavbič



Dekan:

prof. dr. Nikolaj Zimic



## IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Rok Gomišček, z vpisno številko **63100428**, sem avtor diplomskega dela z naslovom:

*Analiza slovenskega političnega prostora na Twitterju*

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Dejana Lavbiča,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, dne 16. avgusta 2013

Podpis avtorja:





# Kazalo

Povzetek

Abstract

<b>1</b>	<b>Uvod</b>	<b>1</b>
1.1	Metodologija . . . . .	3
1.2	Slovarček izrazov . . . . .	3
<b>2</b>	<b>Pregled obstoječih pristopov</b>	<b>5</b>
2.1	Napovedovanje izida volitev . . . . .	5
2.2	Odkrivanje mrež uporabnikov . . . . .	11
<b>3</b>	<b>Analiza parlamentarnih strank na Twitterju</b>	<b>13</b>
3.1	Sledilci in prijatelji . . . . .	14
3.2	O čem govorijo . . . . .	15
3.3	Podobnosti med strankami . . . . .	21
<b>4</b>	<b>Mreže uporabnikov</b>	<b>23</b>
4.1	Polarizacija sledilcev . . . . .	23
4.2	Povezave med uporabniki . . . . .	25
<b>5</b>	<b>Poslanci in Twitter</b>	<b>31</b>
5.1	O čem govorijo . . . . .	33
5.2	Povezave med sledilci poslancev . . . . .	37



# Povzetek

Twitter je hitro rastoče družbeno omrežje, ki zaradi velikega števila uporabnikov privablja tudi politike in raziskovalce. Prvim ponuja način za doseganje volivcev, drugim pa ogromno količino podatkov za analizo; predvsem si želijo najti način, kako uspešno napovedati izid volitev. V svoji diplomski nalogi bom najprej pregledal pretekle pristope raziskovanja političnega dogajanja na Twitterju, v nadaljevanju pa bom analiziral dejavnost slovenske politike na Twitterju, najprej na nivoju strank, nato še na nivoju njihovih poslancev. Pri analizi dejavnosti bom raziskoval koreliranost volilnega izida, aktivnosti, števila objav in števila sledilcev, leksikalno raznolikost besed, hashtagov in omemb v tweetih ter povezave med sledilci strank in sledilci poslancev.



# Abstract

Twitter is a rapidly growing social network, whose vast number of users attracts politicians and researchers alike. The former see it as a way of connecting with their constituents, while it offers a huge amount of data for analysis to the latter, who are especially interested in finding a way to predict elections. I will begin my thesis by covering past attempts at researching political activities on Twitter and later I will analyze the Slovenian political activity on Twitter, both on a party and on a deputy level. During my analysis I will research the correlation between election results, activity, number of tweets and number of followers, the lexical diversity of words, hashtags and mentions used in tweets and the connection between the followers of the parties and the followers of the deputies.



# Poglavje 1

## Uvod

Twitter je spletno mesto, namenjeno pisanju in branju kratkih sporočil (do 140 znakov), imenovanih tweeti. Ustvarjen je bil marca 2006 in je do februarja 2013 zrasel na 200 milijonov uporabnikov, ki prispevajo 400 milijonov tweetov vsak dan [1]. Izvirno naj bi bili tweeti namenjeni objavljanju posodobitev o sebi, a danes se jih uporablja za marsikaj, vključno za stik z javnostjo in oglaševanje. Na Twitterju tako lahko najdemo zelo širok nabor informacij, saj uporabniki lahko z vsemi delijo svoje mnenje o čemer koli želijo. Med uporabniki lahko poleg osebnih računov "navadnih" ljudi najdemo tudi ogromno podjetij, ki oglašujejo svoje izdelke, račune novinarjev in novinarskih strani, znane osebe in skupine iz sveta zabave ter tudi politike in politične stranke.

Stik z javnostjo in oglaševanje sta pomemben del predvolilnih kampanj, saj politikom ponujata nov način za doseganje volivcev. Politikom in strankam Twitter ponuja način, kako širiti svoje sporočilo, naj bo to vabilo na srečanje v živo, povezava na drugo spletno stran ali komentiranje aktualnih dogodkov. Družbeni mediji so zato postali del predvolilnih kampanj in mnogo politikov jih uspešno uporablja tudi, ko ni čas volitev. A ne "tweetajo" vedno politiki sami, včasih imajo temu namenjeno ekipo. Predsednik Republike Slovenije Borut Pahor, na primer, ima Twitter račun (@BorutPahor), a v opisu jasno piše, da sporočila piše ekipa, v primeru, da piše predsednik, pa

so tweeti označeni z BP [2].

Družbeni splet omogoča raziskovalcem, da opazujejo mnenja velikega števila politično aktivnih ljudi, kaj si na primer mislijo o kandidatih ali strankah, ali kakšno je njihovo stališče do aktualnih tem. Zaradi svoje rasti je Twitter vedno bolj zanimiv tudi za raziskovalce političnega dogajanja, kjer raziskovalce zanima, ali se da družbena omrežja, kot je Twitter, uporabiti za razpravljanje o pomembnih temah, za napovedovanje volilnega izida in ali se da ugotoviti politično pripadnost uporabnikov. Zaradi specifik posameznih volilnih sistemov (večstrankarski/dvostrankarski, večinski/proporcionalni) se ne da vseh pristopov, ki so pokazali dobre rezultate na določenih volitvah, avtomatično uporabiti na volitvah v drugi državi.

V svoji diplomski nalogi bom analiziral dejavnost slovenske politike na Twitterju. Najprej bom v 2. poglavju pregledal različne pristope, ki so jih raziskovalci družbenih omrežij uporabili pri raziskovanju političnega dogajanja na Twitterju. Kot bomo videli, so raziskovalci na različnih volitvah uporabljali različne pristope in dosegali različne rezultate, od neuspešnih napovedi do zelo natančnih. V naslednjem poglavju se bom osredotočil na prijatelje, sledilce in tweete posameznih strank. Stranke bom primerjal po teh lastnostih in primerjal z volilnim izidom. Tweete bom razbil na besede, hashtage in omembe, v poskusu ugotoviti, o čem stranke govorijo na Twitterju. V 4. poglavju bom pogledal, kako se sledilci strank razporedijo med strankami in kako se povezujejo med seboj. Za konec bom v 5. poglavju pozornost usmeril na poslance parlamentarnih strank in preveril, o čem tweetajo, koliko sledilcev imajo in kako si jih razdelijo.

Tekom naloge bom poskusil odgovoriti na naslednja vprašanja:

- Katere stranke so najbolj aktivne?
- Ali obstaja povezava med popularnostjo stranke na Twitterju in njenim izidom na volitvah?
- O čem stranke govorijo?
- Ali se politična podobnost strank kaže tudi na Twitterju?



- Ali se da iz skupnih sledilcev sklepati na podobnost strank?
- Kako so sledilci strank povezani med seboj?
- Ali imajo poslanci iste sledilce kot njihova stranka?
- O čem pišejo poslanci?
- Ali sledilci poslancev sledijo vsem poslancem iz iste stranke?

## 1.1 Metodologija

Twitter za dostop do podatkov o uporabnikih in tweetih ponuja svoj API [3], do katerega sem dostopal z modulom `twitter` [4] za programski jezik Python [5]. Iz dobljenih podatkov sem grafe ustvaril z modulom `networkx` [6] in jih vizualiziral z odprtokodnim programom `Gephi` [7]. Zaradi omejitev pri prenosu podatkov sem obširnejše podatke shranjeval v podatkovno bazo MongoDB [8], do katere sem dostopal z modulom `pymongo` [9]. Za izris histogramov v 4. in 5. poglavju sem uporabil modula `matplotlib` [10] in `numpy` [11]. Za izračune sem uporabil tudi module `collections`, `math`, `scipy` [12], `time` in `datetime`.

Izvorna koda je na voljo na <https://github.com/rokgomiscek/Diploma-FRI>.

## 1.2 Slovarček izrazov

- followers (sledilci) - uporabniki, ki prejemaajo posodobitve na svojo časovnico (domačo stran na Twitterju)
- hashtag (#) - uporabniki lahko z znakom # izberejo besedo ali več, ki označuje temo ali ciljno publiko
- mention (omemba) - uporabniki lahko v svojih tweetih omenijo druge uporabnike z uporabo znaka @

- retweet - uporabniki lahko sporočilo nekoga drugega ponovno objavijo tako, da ga vidijo tudi njihovi sledilci
- tweet - največ 140 znakov dolga sporočila na Twitterju

# Poglavje 2

## Pregled obstoječih pristopov

Družbena omrežja privabljajo veliko število ljudi, ki se med seboj povezujejo in objavljajo različne prispevke. To posledično privablja raziskovalce z različnih področij, ki želijo iz ogromne količine podatkov, ki je na voljo, odkriti zakonitosti. Mnogo raziskovalcev je poskušalo ugotoviti, ali se s pomočjo družbenih omrežij da napovedati izid volitev in politično pripadnost uporabnikov.

### 2.1 Napovedovanje izida volitev

Napovedovanje izida volitev na tradicionalni način (anketiranje) je drago in zamudno opravilo. Zato so se raziskovalci vprašali, ali se da s pomočjo družbenih omrežij enostavneje in predvsem ceneje napovedati izid volitev z enako točnostjo, kot z anketami.

Tumasjan et. al. [13] so izpeljali študijo na podlagi nemških zveznih volitev 27. septembra 2009. Uporabili so 104.003 tweetov, objavljenih v tednih pred volitvami (med 13. avgustom in 19. septembrom). Tweete so zbirali tako, da je vsak vseboval ime ene od strank, ki so sodelovale na volitvah, ali ime vidnejših politikov iz teh strank. Namen njihove študije je bil trodelen: pregledati, ali ljudje uporabljajo Twitter za politične pogovore; oceniti, ali sporočila na Twitterju odražajo dejansko politično voljo; analizirati, ali se

da aktivnost na Twitterju uporabiti za napovedovanje popularnosti strank in koalicij. V svoji raziskavi ugotavljajo, da zaradi posebnosti Twitterja ne morejo nanj enostavno aplicirati enakih pristopov, kot pri raziskovanju bolj tradicionalnih družbenih medijev, kot so na primer blogi in strani z novicami. To pripisujejo predvsem dejstvu, da so zapisi na Twitterju precej krajši ter da precejšen del sporočil sestoji samo iz povezave na drugo spletno stran. Da bi ugotovili razpoloženje sporočil, so uporabili program za analizo besedil, ki izračuna, v katero vnaprej definirano kategorijo spada sporočilo (npr. pozitivna/negativna čustva, žalost, jeza). Iz raziskave so ugotovili, da se da, kljub kratkosti, na Twitterju razpravljati tudi o pomembnih vprašanjih. Ugotovili so tudi, da približno ena tretjina tweetov vsebuje znak ”@”, kar nakazuje omenjanje drugih uporabnikov (bodisi za neposredno sporočanje, bodisi zgolj omenjanje). Našli so tudi veliko število retweetov (19,1%). Oboje nakazuje na to, da ljudje na Twitterju dobijo zanimive politične informacije, ki jih delijo s prijatelji in sledilci. Ko so analizirali enakost sodelovanja, so ugotovili, da je 4% uporabnikov prispevalo kar 44% vseh sporočil. S pomočjo analize besedila so ustvarili več-dimenzionalne profile za politike v vzorcu. Ugotovili so podoben vzorec pri vseh vodilnih kandidatih, kjer je več pozitivnih kot negativnih čustev. Profili drugih vidnih politikov kažejo nekaj razlik, predvsem se vidi odražanje njihovih vlog (ministra za finance in ekonomijo imata na primer bolj poudarjeni dimenziji delo in denar). Izračunali so tudi razdaljo med politiki in med strankami ter ugotovili, da odseva dejansko politično bližino med strankami pred volitvami. Povezavo med strankami nakazuje tudi število sporočil, v katerih sta skupaj omenjeni dve stranki. Nenazadnje, preverjali so tudi napovedno moč Twitterja. Primerjali so delež pozornosti, ki ga je bila posamezna stranka deležna na Twitterju, z dejanskim volilnim izidom. Zaključili so, da je samo število, kolikokrat je stranka omenjena, verjeten odraz volilnega rezultata, njegova točnost pa se približa točnosti tradicionalnih volilnih anket. Kljub uspehu raziskave, avtorji svarijo pred njeno omejenostjo. Svarijo pred nereprezentativnostjo vzorca, možnostjo, da v vzorec niso sprejeli vseh relevantnih tweetov, težavami analize tako kratkih

besedil in prevajanje iz nemščine v angleščino, obravnavo vseh sporočil enako, ne glede na datum, kdaj so bila napisana in da so pregledali celotno politično področje in ne posameznih vprašanj, ki bi volivce lahko razdeljevala.

Ali je Twitter dobro orodje za ocenjevanje trenutnega političnega razpoloženja in napovedovanje priljubljenosti strank in koalicij, so raziskovalci preverjali tudi na različnih volitvah v Španiji [14]. Razvili so spletno aplikacijo TaraTweet, ki ima dva cilja: spremljanje tweetov z izbranimi hashtagi in štetje pojavitev izbranih ključnih besed. V raziskavo so vključili regionalne, splošne in volitve v Andaluziji, tweete pa zbirali dalj časa pred volitvami in tudi še nekaj časa po. Pri vseh volitvah so upoštevali celotno število omemb in ugotovili visoko povezanost z dejanskimi izidi. Med težavami raziskave so izpostavili probleme pri iskanju relevantnih tweetov, saj uporabniki lahko uporabljajo hashtag, ki jih raziskovalci niso predvideli in vključitev tweetov, ki niso povezani z izbranimi volitvami, temveč se nanašajo na druge. Kljub temu so optimistični, da je možno Twitter uporabljati za napovedovanje izida volitev.

Število tweetov kot osnovo za napovedovanje izida volitev so uporabili tudi na posebnih volitvah za senat ZDA v Massachusettsu, 19. januarja 2010 [15]. Zbrali so 234.697 tweetov 56.165 različnih uporabnikov med 13. in 20. januarjem. Ugotovili so, da zgolj število tweetov ni dovolj za napovedovanje izida volitev in predlagajo, da je treba upoštevati tudi pomen tweeta. Tako so za oba kandidata tweete razdelili na podporne in nasprotujoče. A polarnost tweeta je težavno avtomatično razbrati zaradi napak pri klasifikaciji besed, težav pri besedah z večimi pomeni in zapisov, ki jim polarnost dajejo ločila. Zato svojo raziskavo zaključujejo z ugotovitvijo, da enostavne metode napovedovanja izida volitev z upoštevanjem pomena ne prinašajo dobrih rezultatov.

Nekaj raziskovalcev se je osredotočilo na sledenje mnenja, ki ga izražajo uporabniki Twitterja skozi čas [16]. Analizirali so vzdušje na Twitterju v povezavi s predsedniškimi volitvami v ZDA leta 2008 in merili podporo predsedniku ZDA v letu 2009 ter oboje primerjali z rezultati javnomnenjskih

raziskav. Raziskavo so razdelili na dva podproblema: prepoznati tweete, ki so povezani s temo in oceniti, ali tweeti vsebujejo pozitivno ali negativno mnenje. Tweete, povezane s temo, so dobili tako, da so iskali izbrane ključne besede, polarnost mnenja pa so izračunali s pomočjo seznama besed, označenih kot negativne ali pozitivne. Naključno so preverjali, kako so bili tweeti klasificirani in opazili več primerov napačno označenih. To pripisujejo predvsem temu, da so uporabljali seznam, namenjen dobro napisanim besedilom in ne neformalnemu zapisu na Twitterju, kjer je mnogo drugače črkovanih besed in se uporabljajo čustveni simboli. Ker se polarnost mnenj iz dneva v dan zelo močno spreminja, so to spreminjanje ublažili z uporabo 15-dnevnega drsečega povprečja. Raziskava je sicer pokazala, da se razpoloženja na Twitterju ne da uporabiti za napovedovanje volitev, podpora predsedniku na Twitterju pa je podobna mnenju v resničnem življenju, a se sprašujejo, ali je to zaradi političnih novic in dogodkov. Na koncu povedo, da bi bilo potrebno izboljšati sisteme za ocenjevanje besedil in upoštevati demografske značilnosti uporabnikov.

Uporabo števila tweetov in analizo razpoloženja so uporabili tudi za napovedovanje izida šestih volitev v ZDA leta 2010 [17, 18]. Pri enih volitvah so ugotovili, da število tweetov v celotnem opazovanem obdobju sicer pravilno napoveduje izid, a je v to obdobje všteti tudi čas po zaključku volitev. Če bi pa gledali samo obdobje pred volitvami, bi dobili napačno napoved. Gledano v celoti, je vsaka metoda pravilno napovedala izid volitev le v polovici primerov, v enem primeru sta obe metodi pravilno napovedali, v enem pa obe napačno. Izračunali so, da analiza razpoloženja sicer ima manjšo napako kot merjenje števila tweetov, a še vedno ne dosega ravni profesionalnih anket. Da bi ugotovili natančnost avtomatičnega klasificiranja tweetov, so naključno izbrane tweete še ročno klasificirali. Ko so primerjali avtomatično klasificirane tweete z ročno klasificiranimi, so ugotovili, da je metoda s slovarjem narobe klasificirala mnogo tweetov in je le nekoliko boljša od naključnosti. Boljšo natančnost tradicionalnih metod za merjenje javnega mnenja pripisujejo temu, da se njihovih kompleksnosti ne da duplicirati z vzorčenjem socialnega me-

dija. Profesionalne javnomnenjske raziskave upoštevajo tudi verjetne volivce, to je tiste, ki kažejo večjo verjetnost, da bodo dejansko šli volit, poleg tega so prirejene glede na starost, na družbenih omrežjih pa tega ne moremo zagotovo izvesti. Tudi če bi imeli dostop do teh podatkov, še vedno ne bi mogli dobiti popolnoma naključnega vzorca, saj dobimo podatke le od oseb, ki so se odločile objaviti svoje mnenje. Dodatno težavo predstavljajo spammerji in propagandisti, ki ustvarjajo lažne račune in reklamirajo izbrano stran, oziroma blatijo nasprotnike. Avtorji predlagajo štiri standarde, ki bi jih morali upoštevati za napovedovanje izida volitev. Kot prvi standard predlagajo uporabo algoritma, ki bi se ga dalo nekoliko spremeniti, glede na specifične posameznih volitev. Pred samimi volitvami bi morali jasno določiti način pridobivanja podatkov, način obdelave podatkov, algoritem z vhodnimi podatki in semantiko, s katero bi interpretirali rezultate. Z drugim standardom opominjajo na to, da so podatki iz družbenih medijev bistveno drugačni od podatkov naravnih pojavov, zlasti zaradi spammerjev in propagandistov. Njihova tretji standard predlaga bodočim raziskovalcem, naj ustvarijo preverljivo teorijo, zakaj in kdaj pravilno napovedujejo. Pomembno je namreč razumevanje, zakaj neka metoda vrača pravilno napoved volilnega izida. V povezavi s tem predlagajo vzpostavitev neke izhodiščne napovedne točnosti, pod katero ne smejo pasti. Z zadnjim standardom pa predlagajo učenje od profesionalnih javnomnenjskih raziskovalcev, kar sicer ni nujno, a je ena od možnosti, da bi napovedovanje volitev s pomočjo družbenih medijev lahko delovalo. Izpostavljajo predvsem prepoznavanje verjetnih volivcev in pridobivanje nepristranskih vzorcev.

Da rudarjenje mnenj na družbenem spletu ne more zamenjati javnomnenjskih raziskav, se zavedajo Bravo-Marquez et. al. v svoji raziskavi [19], saj, kot pravijo, uporabniki družbenega spleta niso nujno reprezentativni vzorec resničnega sveta. Kljub temu vidijo nekaj prednosti mnenj, pridobljenih iz družbenih medijev, v primerjavi s tradicionalnim anketiranjem. Prva prednost je cenejša obdelava večjega števila podatkov, druga pa, da so taka mnenja primernejša za preučevanje časovnih lastnosti javnega mnenja. V

raziskavi posvečajo posebno pozornost dinamiki mnenj, torej kako se javno mnenje razvija skozi čas. Raziskavo so opravili na predsedniških volitvah v ZDA leta 2008 in zbrali 250.000 tweetov, ki jih je ustvarilo 20.000 uporabnikov med 1. junijem in 11. novembrom 2008. Razpoloženje so ocenili s štetjem negativnih in pozitivnih besed, kot so definirane v danem slovarju. Ker so bili tweeti v zbirki označeni s časovnim žigom, so jih lahko združevali glede na časovno obdobje (dan, teden, mesec). Merili so aktivnost, pozitivnost, negativnost in polarnost. Aktivnost predstavlja število tweetov o izbranem dogodku v danem časovnem obdobju. Pozitivnost za nek dogodek so definirali kot povprečno število pozitivnih besed na tweet, podobno so negativnost definirali kot povprečno število negativnih besed, polarnost pa kot razliko med pozitivnostjo in negativnostjo. Ena prvih ugotovitev raziskave je bila, da pozitivnost in negativnost nimata linearnega razmerja z nivojem aktivnosti v izbranem časovnem obdobju. Ugotovili so tudi, da se nihajnost mnenj s časom zmanjšuje. Ena izmed možnih razlag, ki jo ponujajo avtorji za to, je ta, da so na začetku volivci bolj odprti za nove podatke in je večja negotovost, za koga bodo glasovali, sčasoma pa se njihova izbira zjasni, kar vodi do manjšega nihanja. Zaključujejo, da je zaradi nihanja skozi čas težko uporabiti podatke s Twitterja za dolgoročno napovedovanje.

Med slovenskimi predsedniškimi volitvami leta 2012 je podjetje Gama System računalo politični indeks kandidatov. Tega so računali tako, da so poiskali vse objave na temo volitev, jih sentimentalno ocenili in poiskali absolutno razliko med številom pozitivnih in negativnih objav [20]. Izvedena sicer ni bila nobena primerjava tako dobljenega političnega indeksa z anketami ali z rezultatom na volitvah, a razvidno je, da je politični indeks kandidata, ki je zmagal na volitvah, večino časa višji od političnega indeksa ostalih kandidatov.



## 2.2 Odkrivanje mrež uporabnikov

Nekatere raziskovalce je zanimalo, kako se uporabniki Twitterja povezujejo med seboj in ali se iz tega da razbrati pripadnost določeni politični opciji. Conover et. al. so raziskovali mrežo retweetov in mrežo omenjanja na podlagi političnega komuniciranja na Twitterju pred vmesnimi volitvami v ZDA, leta 2010 [21, 22]. Kot politično komunikacijo so definirali vse tweete, ki vsebujejo vsaj en politično relevanten hashtag. Nepristransko množico političnih hashtagov so dobili z uporabo iskanja sopojavnosti. Začeli so z najbolj popularnim hashtagom vsake politične opcije in poiskali, s katerimi hashtagi se pojavijo skupaj vsaj enkrat. Z izbrano mejo podobnosti so dobili 66 hashtagov, od katerih so jih umaknili 11, ker so preveč splošni. S tem so dobili 252.300 politično relevantnih tweetov. Ugotovili so, da je omrežje retweetov sestavljeno iz dveh gruč, kar nakazuje, da uporabniki posredujejo zapise, ki so jim politično blizu. Po drugi strani pa v omrežju omemb niso našli posebnih struktur, kar pripisujejo temu, da uporabniki z obeh političnih strani pozivajo uporabnike z nasprotne strani. Opazili so tudi, da nekateri uporabniki označijo svoje tweete z nasprotno politično opcijo, čeprav ta nima zveze z vsebino napisanega.

Song in Kim sta se v svoji raziskavi lotila iskanja mrež na podoben način, kot je opisano zgoraj, a bolj omejeno [23]. V tem primeru sta raziskovalca s sistemom RT<sup>2</sup>M (Real-Time Twitter Trend Mining) ustvarila seznam vseh uporabnikov, ki so bili omenjeni skupaj z iskano besedo. Poleg tega sistem lahko izpiše tudi seznam pojmov, ki se pojavijo skupaj z izbranim, izračuna, koliko sta si dva uporabnika blizu in prikaže trend za izbran pojem. Sistem sta preizkusila na predsedniških volitvah v Južni Koreji leta 2012. Kot primer uporabe sta prikazala vizualizacijo mreže, ki sta jo dobila z uporabo imena ene izmed kandidatk za predsednico. Ugotovila sta, da je mreža sestavljena iz večjega števila dobro povezanih gruč z redkimi povezavami med njimi.

Iz dosedanjih raziskav lahko razberemo, da se raziskovalci političnega dogajanja na Twitterju soočajo z različnimi težavami. Prva je vsekakor pridobivanje vseh relevantnih tweetov, saj se ne moremo zanesti, da bodo upo-

rabniki svoje zapise označili. Druga težava je avtomatično razbiranje vsebine tweetov, saj se je izkazalo, da so zapisi pogosto narobe klasificirani.

## Poglavje 3

# Analiza parlamentarnih strank na Twitterju

Vse parlamentarne stranke, razen DeSUS, imajo svoj Twitter račun (dva poslanca iz DeSUSa imata Twitter račun, a več o tem v 5. poglavju). To gre verjetno pripisati dejstvu, da DeSUS nagovarja predvsem ciljno publiko, ki po večini ne uporablja Twitterja. Zato bom iz nadaljnih raziskav to stranko izvzel. V raziskavo bom tako vključil naslednje stranke (v vrstnem redu glede na rezultat na predčasnih volitvah v državni zbor, 4. 12. 2011):

- Pozitivna Slovenija (@PozitivnaSlo; 20. 9. 2012)
- Slovenska demokratska stranka (@strankaSDS; 25. 2. 2009)
- Socialni demokrati (@strankaSD; 26. 10. 2011)
- Državljska lista (@strankaDL; 19. 11. 2012)
- Slovenska ljudska stranka (@strankaSLS; 21. 4. 2010)
- Nova Slovenija (@NovaSlovenija; 29. 4. 2010)

V oklepaju je zapisano strankino uporabniško ime na Twitterju in datum, ko je bil račun ustvarjen. Opazimo, da stranke uporabljajo enostavno prepoznavna uporabniška imena. Prva parlamentarna stranka, ki je začela uporabljati Twitter, je SDS, leto pred ostalimi.

V tabeli 3.1 je podan izid na zadnjih parlamentarnih volitvah (4. 12. 2011) v odstotkih ter število sledilcev in objav v času pisanja te naloge (avgust 2013). Iz teh podatkov in iz tega, koliko časa stranka že ima račun, lahko izračunamo strankino aktivnost na Twitterju (povprečno število tweetov na dan). Ugotovimo, da stranke uporabljajo Twitter z različno aktivnostjo. Najbolj aktivni so pri PS, s 6,4 tweeti na dan, sledi jim SDS, s 5,2 tweeti na dan. Najmanj aktivni pa so pri NSi, kjer v povprečju objavijo 1,1 tweet na dan.

Modul `scipy` vsebuje funkcijo `pearsonr`, ki izračuna Pearsonov koeficient korelacije, ki nam za navedena seznama vrednosti pove, če sta povezana ali ne. Če primerjamo število objav s številom sledilcev, vidimo visoko koreliranost (96,7%), kar nakazuje na to, da stranke z večjim številom objav privabijo več sledilcev. Število sledilcev je tudi nekoliko korelirano s tem, koliko časa obstaja strankin Twitter račun (82,5%), ni pa povezano z aktivnostjo stranke (43,7%). Zdaj lahko že odgovorimo na vprašanje, ali obstaja povezava med številom sledilcev (in, zaradi koreliranosti, objav) stranke in njenim rezultatom na volitvah. Koreliranost izida na volitvah in števila sledilcev je 42,2%, kar pomeni, da nista korelirani. To se lahko vidi tudi iz samega števila sledilcev, saj ima prvouvrščena stranka nekaj manj sledilcev kot zadnja, druga pa daleč največ. Obstaja pa koreliranost med aktivnostjo in izidom na zadnjih volitvah (89,9%). A koreliranost ni dokaz za vzročnost, zato je možna razlaga te koreliranosti tudi to, da stranke z boljšim uspehom na volitvah dobijo več funkcij (poslanci, ministri ipd.) in, ker na twitterju sporočajo o aktivnostih svojih članov, imajo več objav.

### 3.1 Sledilci in prijatelji

Odnosi med uporabniki na Twitterju so tipično asimetrični. Uporabniki lahko sledijo drugim, ne da bi tem bilo treba kar koli potrditi ali slediti nazaj. Za popularne račune je značilno, da imajo dosti sledilcev, ki tako prejemajo posodobitve, manj pa sami sledijo drugim.

Stranka	Izid volitev 2011	Število sledilcev	Število objav
PozitivnaSlo	28,51	1868	2058
strankaSDS	26,19	6748	8435
strankaSD	10,52	3067	1924
strankaDL	8,37	708	511
DeSUS	6,97	/	/
strankaSLS	6,83	3652	4618
NovaSlovenija	4,88	2061	1362

Tabela 3.1: Izid volitev leta 2011 [24], število sledilcev in število objav (podatki s Twitterja dne 9. 8. 2013).

V tabeli 3.2 je prikazano število sledilcev posamezne stranke, število prijateljev (računi, ki jim stranka sledi) in presek teh dveh množic, torej koliko je takih računov, ki sledijo stranki in stranka njim. Razmerje med sledilci in prijatelji je drugačno glede na stranko, a pri vseh velja, da imajo več sledilcev kot prijateljev. Največ prijateljev imajo pri SLS, procentualno gledano pa pri PS. O povezavah med uporabniki bomo izvedeli več v 4. poglavju.

## 3.2 O čem govorijo

Poleg tega, kdo je publika strank na Twitterju, nas zanima tudi, kaj stranke pišejo. Twitter nam za danega uporabnika vrne največ 3.200 tweetov, kar je za večino strank dovolj, da dobimo njihovo celotno zgodovino. Iz tweetov lahko pridobimo marsikateri podatek, a tu se bomo omejili na same besede, omembe drugih uporabnikov in hashtage. Russell [25] predlaga nekaj enostavnih prijemov za analizo besedil, ki sem jih priredil za namene te naloge.

Prva primerjava je leksikalna raznolikost. Izračunamo jo tako, da delimo število različnih besed s številom vseh besed v tweetih. Zasede vrednost med 0 in 1, višja vrednost pomeni več različnih besed. Na področju medosebne komunikacije predstavlja kvantitativno mero raznolikosti komu-

Stranka	Št. sledilcev	Št. prijateljev	Presek
PozitivnaSlo	1868	803	550
strankaSDS	6748	424	251
strankaSD	3067	430	236
strankaDL	708	227	73
strankaSLS	3652	899	440
NovaSlovenija	2061	528	233

Tabela 3.2: Število sledilcev in prijateljev strank (podatki s Twitterja dne 9. 8. 2013).

niciranja. Pričakujemo lahko, da uporabniki z višjo leksikalno raznolikostjo pokrivajo večje število tem in se manj ponavljajo. V tem primeru jo lahko apliciramo na besede, omembe in hashtage. Rezultati (prikazani v tabeli 3.3) kažejo na podobno leksikalno raznolikost besed in omemb pri vseh strankah, nekoliko se razlikujejo pri raznolikosti uporabe hashtagov. Večina strank ima tudi tam podobno vrednost, ena stranka precej višjo, druga pa precej nižjo. Stranke imajo zelo nizko raznolikost omemb, kar kaže na to, da omenjajo pretežno iste uporabnike. Raznolikost besed in hashtagov sta sicer višji kot raznolikost omemb, a še vedno precej nizki. To si lahko razložimo tako, da stranke uporabljajo svoje Twitter račune za omejeno število tem. V tabeli 3.3 vidimo tudi povprečno število besed na tweet, ki v povprečju za vse stranke znaša 15,7, kar nakazuje, da v tweetih strank prevladujejo celi stavki ter sporočila in ne krajši zapisi.

V tabeli 3.4 so prikazane najbolj pogoste besede s frekvencami po strankah. Pri vseh strankah opazimo besedo "RT", ki na Twitterju označuje retweet, oziroma objavo tweeta nekoga drugega. Večino ostalih pogostih besed pri vseh strankah predstavljajo imena vidnejših članov. Pri treh strankah je med najbolj pogostimi besedami tudi "danes", pri eni pa "Vabimo", kar si lahko razlagamo s tem, da stranke uporabljajo Twitter predvsem za pošiljanje obvestil svojim sledilcem o raznih dogodkih.

Stranka	Leksikalna raznolikost			
	besed	omemb	hashtagov	$\bar{n}$
PozitivnaSlo	0,38	0,16	0,32	14,3
strankaSDS	0,37	0,11	0,43	15,0
strankaSD	0,30	0,24	0,19	18,0
strankaDL	0,42	0,23	0,42	15,4
strankaSLS	0,34	0,19	0,30	16,1
NovaSlovenija	0,38	0,16	0,53	15,4
Povprečje	0,36	0,18	0,37	15,7

Tabela 3.3: Leksikalna raznolikost besed, omemb in hashtagov ter povprečno število besed na tweet po strankah (podatki s Twitterja dne 9. 8. 2013).

V prejšnjem odstavku smo ugotovili, da stranke pogosto retweetajo. V retweetu je pogosto zapisan tudi avtor originalnega tweeta, zato lahko s tabelo 3.5, kjer so zapisane najbolj pogoste omembe v tweetih strank, poskusimo ugotoviti, čigave tweete stranke ponovno objavijo. Vidimo, da večino omemb predstavljajo člani stranke in novinarski računi. Iz tega lahko sklepamo, da stranke retweetajo zapise svojih članov in novice, ki so jim blizu.

Ker se s hashtagi označi tweet glede na njegovo vsebino, predstavljajo nekakšno enobesedno obnovo tweeta. Z njimi lahko torej na hiter in enostaven način razberemo, o čem se stranke pogovarjajo. V tabeli 3.6 so prikazani najbolj pogosti hashtagi po strankah. Skoraj vse stranke (izjema je strankaDL) imajo med pogosto uporabljenimi hashtagi #soocenje, najverjetneje so s tem vabile k ogledu in komentirale razna predvolilna soočenja. Pogosti hashtagi so tudi o raznih volitvah (npr. #volitve2011, #volitve2012) in dogodkih znotraj stranke (npr. #kongresPS, #10kongresSDS).

PozitivnaSlo	PS (235), RT (219), Alenka (137), Bratušek (123), Bratušek: (96), DZ (88), danes (85), Jani (85), mag. (70), lahko (67), @DaniloTurk2012 (67)
strankaSDS	RT (945), via (230), Reporter.si (206), @RevijaReporter (184), SDS (152), .@JJansaSDS (142), Dr. (141), @Planetsiolnet (133), .@MilanZver (125), Vabimo (119)
strankaSD	SD (737), dr. (383), Lukšič: (259), Predsednik (256), Igor (223), predsednika (147), Lukšič (140), Borut (109), danes (107), DZ (89)
strankaDL	RT (86), Minister (67), Državljanske (59), dr. (52), Virant (47), @SenkoPlicanic: (45), liste (44), DL (43), odbora (36), Facebook (32)
strankaSLS	SLS (1057), RT (504), :) (333), Žerjav (209), Zerjav (191), Franc (182) mag. (161), #volitve11 (157), Žerjav: (148), pri (148)
NovaSlovenija	NSi (464), RT (361), Ljudmila (220), Novak (127), Novak: (106), @strankaSDS: (100), @LjudmilaNovak: (96), Matej (92), @LjudmilaNovak (67), danes (64)

Tabela 3.4: Najbolj pogoste besede po strankah (podatki s Twitterja dne 9. 8. 2013).



PozitivnaSlo	@DaniloTurk2012 (95), @MJaniPS (89), @RomanJakic (67), @TVOdmevi (62), @TamaraVonta (50), @PozitivnaSlo (31), @ggmPS (28), @majusdimi (17), @24ur_com (13), @stanka_d (11), @rtvslo (11)
strankaSDS	@JJansaSDS (356), @RevijaReporter (265), @MilanZver (255), @PlanetSiolnet (155), @Demokracija1 (105), @PortalPolitikis (56), @finance_si (38), @AndrejaValic (35), @vinkovasle1 (32), @Libertarec (30)
strankaSD	@BorutPahor (70), @strankaSD (63), mojcakleva (42), @matevzfrangez (40), @lajnar_EU (25), @TheProgressives (24), @tfajon (23), @RTV_Slovenija (21), @rtvslo (18), @Hannes_Swoboda (18)
strankaDL	@SenkoPlicanic (49), @mastenMMC (8), @STA_novice (7), @StudioCity_ (7), @rtvslo (4), @RTV_Slovenija (4), @PIArovkaMNZ (4), @gregor_virant (3), @TVOdmevi (3), @jeanmark00 (3),
strankaSLS	@STA_novice (133), @Janez_Tomsic (79), @NG_SLS (72), @JasminaOpec (54), @rtvslo (40), @RTV_Slovenija (39), @Dnevnik_si (37), @24ur_com (37), @petrasovdat (34), @Demokracija1 (33)
NovaSlovenija	@LjudmilaNovak (171), @strankaSDS (102), @MilanZver (84), @aleshojs (29), @MO_RS (26), @jozefhorvat (24), @MladaSlovenija (21), @RevijaReporter (12), @NovaSlovenija (12), @Demokracija1 (10)

Tabela 3.5: Najbolj pogoste omembe po strankah (podatki s Twitterja dne 9. 8. 2013).

PozitivnaSlo	#PV (37), #soocenje (34), #DZ (16), #PSgorenjska (14), #konvencija (11), #zaskupnodobro (8), #PSmladi (8), #PSstajerska (6), #PSprimorska (6), #kongresPS (5)
strankaSDS	#dz (35), #soocenje (26), #10kongresSDS (15), #volitve2012 (10), #dandrzavnosti (10), #predsednik12 (9), #EPP (7), #ljprotest (7), #rtvolitve2012 (6), #podlupo (6), #sooceneje (6)
strankaSD	#tribunaSD (85), #soocenje (59), #kongresSD (51), #mladiSD (25), #volitve2011 (18), #volitve11 (14), #spopad (14), #volitve (11), #Luksic (9), #RelaunchingEurope (9),
strankaDL	#tovsak (7), #ustavnespremembe (2), #StudioOb17ih (1), #pličanič (1), #plinskiterminal (1)
strankaSLS	#volitve11 (157), #volitve2011 (49), #volitve (44), #soocenje (41), #kongresSLS (29), #25letSLS (24), #Finance_si (16), #mandatar (14), #spomladanskiposvet (13), #GZS (12), #delamo (12), #pogledislovenije (12)
NovaSlovenija	#blizuljudem (10), #Finance_si (5), #tabornsi (3), #medijskanoc (3), #soocenje (2), #podlupo (2),

Tabela 3.6: Najbolj pogosti hashtagi po strankah (podatki s Twitterja dne 9. 8. 2013).

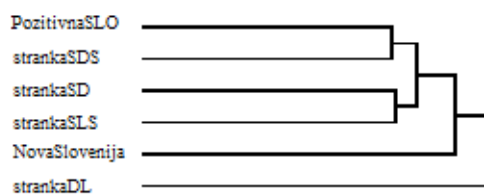
### 3.3 Podobnosti med strankami

Obravnavane stranke lahko s pomočjo sklenjenih koalicij po volitvah v zadnjih desetih letih razdelimo na dva pola. K enemu polu sem štel PS in SD, k drugemu polu pa SDS in NSi. Za to delitev sem se odločil, ker so v zadnjem desetletju stranke z istega pola bile skupaj v koaliciji, z nasprotnega pa ne. Ostalih dveh strank nisem uvrstil k nobenemu izmed polov, ker sta bili v koaliciji s strankami z obeh polov.

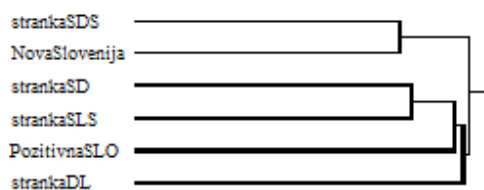
Na podlagi pridobljenih podatkov o uporabljenih hashtagih in omembah, lahko poskusimo ugotoviti, koliko so si stranke med seboj podobne. To sem dosegel tako, da sem vsako stranko dal v svojo skupino, nato poiskal najbližji si skupini in ju združil. To sem ponavljal, dokler ni ostala samo ena skupina. Za mero razdalje med skupinami sem vzel povprečno kosinusno razdaljo med vsemi pari primerov. Združevanje v skupine sem prikazal z dendrogramom.

Pričakoval sem, da bo združevanje strank glede na omembe pokazalo, da so si stranke, ki pripadajo isti politični polarnosti, bližje, kot stranke z nasprotnega pola. Pri združevanju strank glede na uporabljene hashtage pa tega ne pričakujem, saj predvidevam, da stranke razpravljajo o podobnih temah.

Združevanje strank v skupine glede na uporabljene hashtage prikazuje slika 3.1, glede na omembe pa slika 3.2. Na obeh slikah vidimo, da so si stranke precej različne glede teh dveh lastnosti, saj se združijo precej pozno. Združitve so tudi drugačne od takih, kot bi jih pričakovali glede na pretekle koalicije. To gre verjetno pripisati dejstvu, da stranke pretežno omenjajo svoje člane in uporabljajo hashtage, ki so specifični za posamezno stranko.



Slika 3.1: Dendrogram podobnosti strank glede na uporabljene hashtage.



Slika 3.2: Dendrogram podobnosti strank glede na omenjene uporabnike.

# Poglavje 4

## Mreže uporabnikov

Šestim slovenskim strankam na Twitterju sledi 9.745 različnih uporabnikov (podatki za 10. 8. 2013). Ker so sledilci tisti uporabniki, ki prejema jo posodobitve na svojo časovnico, predstavljajo strankam primarno publiko, torej koliko uporabnikov bo neposredno videlo tweete. V poglavju 3.1 smo govorili o številu sledilcev, ki jih ima posamezna stranka. V nadaljevanju si bomo pogledali, koliko strankam sledijo ti uporabniki in kako so povezani med seboj.

V tabeli 4.1 je prikazano celotno število sledilcev po posameznih strankah, koliko uporabnikov sledi samo tej stranki ter kolikšen delež vseh uporabnikov predstavljajo. Ugotovimo lahko, da večina sledilcev vseh strank sledi še kakšni drugi stranki, izjema je le SDS, saj polovica njenih sledilcev ne sledi nobeni drugi stranki. Vsem strankam pa sledi 239 uporabnikov.

### 4.1 Polarizacija sledilcev

Na družbenih omrežjih so podatki sicer pogosto prosto dostopni vsem, a uporabniki lahko izbirajo, kaj bodo videli. Zaradi tega obstaja nevarnost, da postanejo mnenja uporabnikov vse bolj ekstremna, saj so bolj izpostavljeni podobnim idejam in manj nasprotujočim [21].

Ali se da iz skupnih uporabnikov sklepati na podobnost strank? Predpo-

Stranka	Št. sledilcev	Samo tej stranki	Delež
PozitivnaSlo	1868	402	0,22
strankaSDS	6750	3434	0,51
strankaSD	3069	738	0,24
strankaDL	707	61	0,09
strankaSLS	3651	864	0,24
NovaSlovenija	2065	237	0,11

Tabela 4.1: Število sledilcev in ekskluzivnih sledilcev (podatki s Twitterja dne 10. 8. 2013).

stavimo, da večino sledilcev neke stranke predstavljajo potencialni volivci in simpatizerji strankine politične opcije. Iz tega sledi, da bodo ti uporabniki sledili strankam iz iste, ne bodo pa sledili strankam iz nasprotnne politične opcije. Predpostavko lahko preverimo, če preštejemo število uporabnikov Twitterja, ki sledi dvema strankama.

V tabeli 4.2 je prikazano, koliko uporabnikov sledi vsaj danima dvema strankama. Na prvi pogled bi lahko ovrgli prejšnjo hipotezo, saj ima NSi sicer največ skupnih sledilcev s SDS, a ta ima več skupnih sledilcev s SD. Na drugem polu je situacija podobna, kjer imata obe stranki več skupnih sledilcev s stranko nasprotnega pola kot med seboj. Lahko pa primerjamo povprečni delež skupnih sledilcev. Tega sem dobil tako, da sem izračunal delež, ki ga predstavljajo skupni sledilci za posamezen par strank, seštel deleža ter delil z dva, da sem dobil srednjo vrednost. Tak izračun je pokazal, da sta si najbolj podobni SLS in NSi (53,1%), sledijo SDS in SLS, SD in SLS, SDS in NSi ter PS in DL (45%). Na drugi strani lestvice pa sta si najmanj podobni PS in NSi (35,8%), sledijo PS in SDS, DL in NSi, PS in SLS ter SDS in DL (39,3%). Vidimo torej, da je možno iz skupnih sledilcev razbrati podobnost oziroma različnost strank.

Presek sledilcev dveh strank nam ne pove vsega. Vidimo, na primer, da DL in NSi sledi 398 uporabnikov, ne izvemo pa, da samo tema dvema stran-

Stranka	NSi	SLS	DL	SD	SDS	PS
PozitivnaSlo	707	947	460	1038	1093	-
strankaSDS	1513	2302	509	1838	-	
strankaSD	1090	1605	483	-		
strankaDL	398	496	-			
strankaSLS	1406	-				
NovaSlovenija	-					

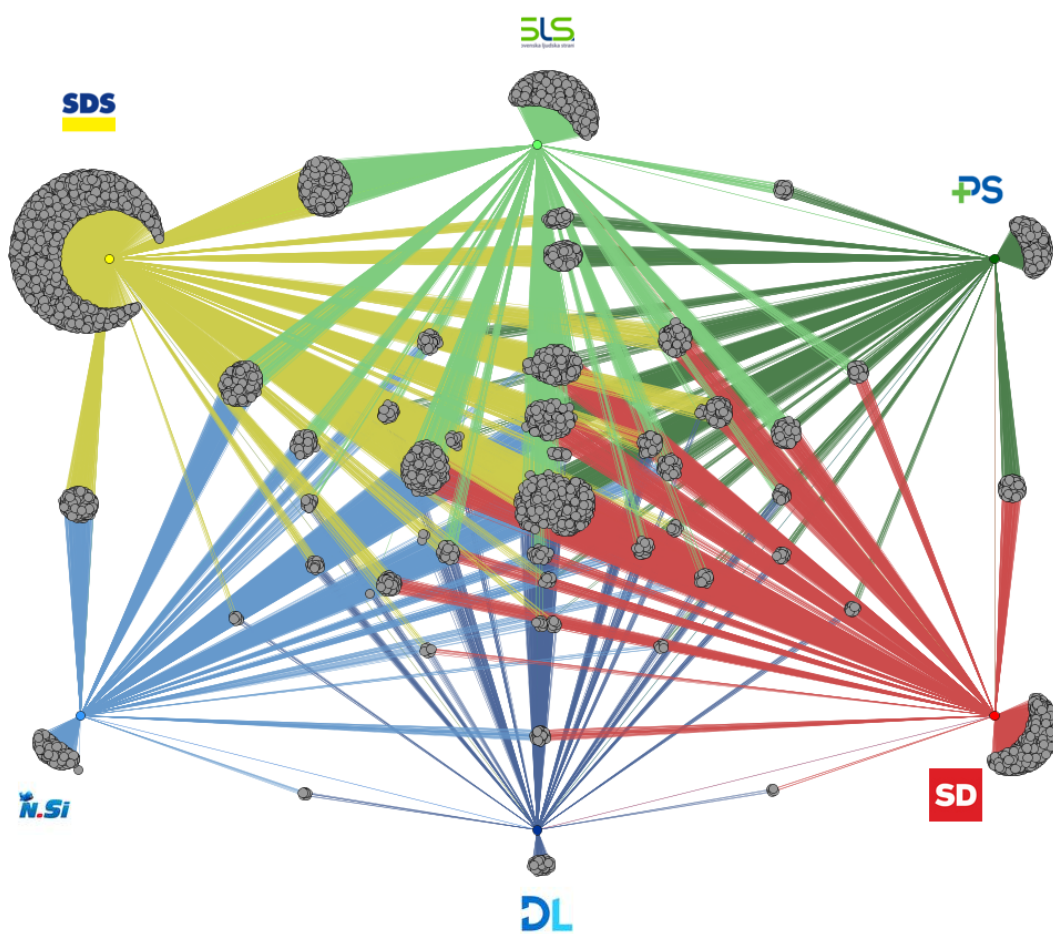
Tabela 4.2: Število skupnih sledilcev (podatki s Twitterja dne 10. 8. 2013).

kama sledi le 6 uporabnikov. Ker bi bila razpredelnica z vsemi možnimi kombinacijami, koliko uporabnikov sledi izbranim strankam, nepregledna, sem se odločil te povezave prikazati v obliki grafa, kot to prikazuje slika 4.1. Pobarvane točke predstavljajo stranke, sive točke predstavljajo sledilce, povezava pa nakazuje sledenje. Siva vozlišča v bližini strank predstavljajo uporabnike, ki sledijo samo eni stranki, vozlišča med strankami pa predstavljajo uporabnike, ki sledijo več strankam.

## 4.2 Povezave med uporabniki

Prikazati graf, kako so sledilci vseh strank povezani med seboj, bi bilo zaradi velikega števila povezav (skoraj 420.000) nepraktično. Tudi razbitje grafa na več manjših grafov po strankah se pokaže kot nepraktično za prikaz. Zato se bom osredotočil na stopnjo vozlišč v posameznih grafih. Ker nam stopnja vozlišča pove, koliko povezav ima določeno vozlišče, lahko iz tega sklepamo na to, koliko se uporabniki povezujejo med seboj.

Na sliki 4.2 so prikazane frekvence stopenj vozlišč za grafe posameznih strank. Prikaz sem na abscisni osi omejil, saj je pri višjih stopnjah le nekaj uporabnikov. Ti uporabniški računi praviloma pripadajo novinarskim stranem ali znanim osebam. Pri vseh strankah opazimo enak vzorec, in sicer, da je velika večina vozlišč (sledilcev) prve stopnje. Iz tega sledi, da se sledilci



Slika 4.1: Graf sledilcev strank na Twitterju.

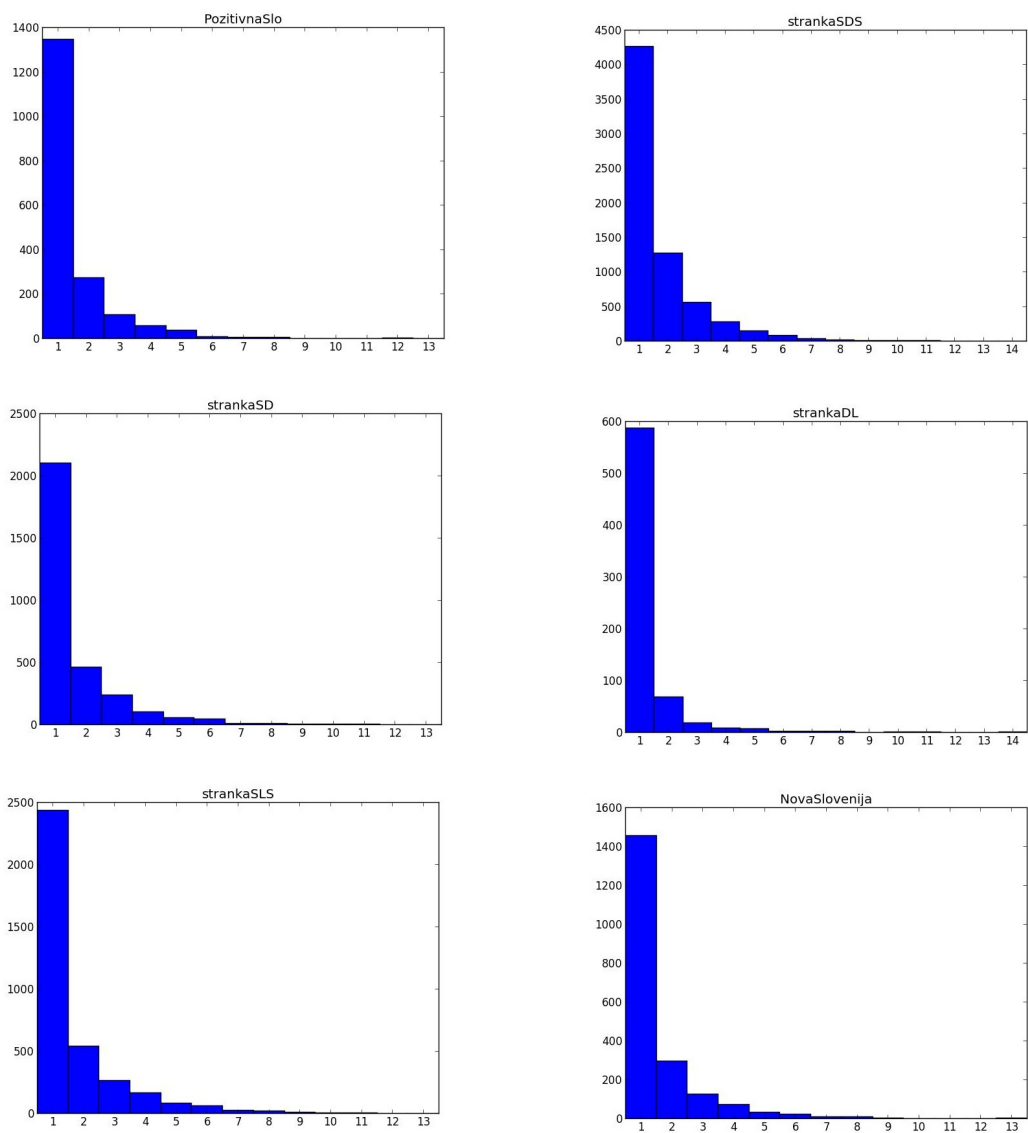


posameznih strank redko povežejo med seboj.

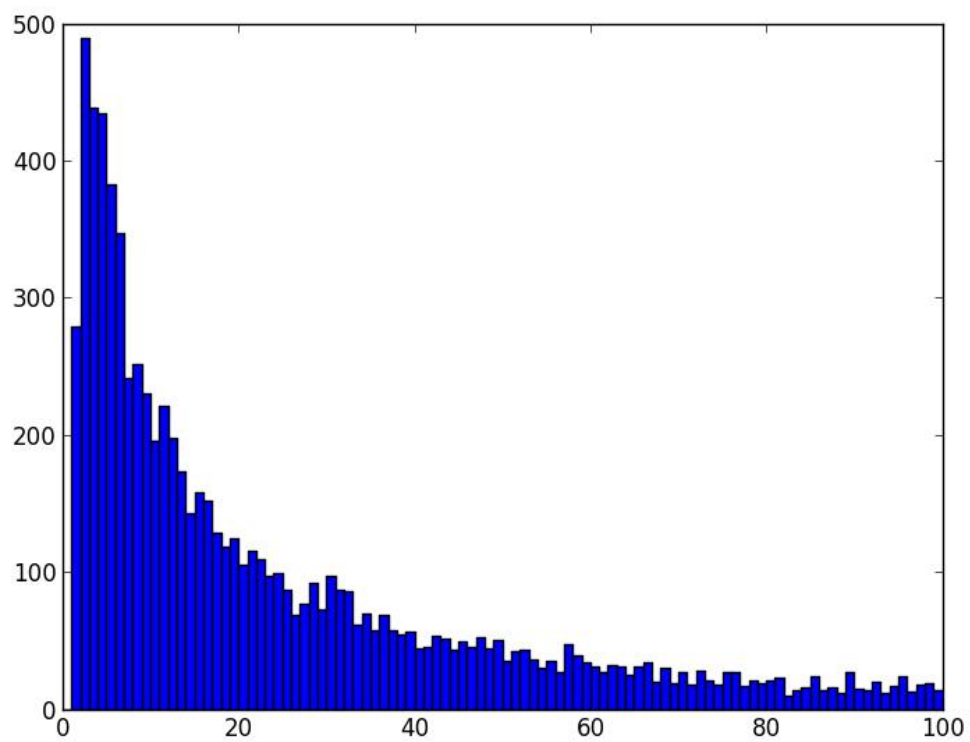
Če pa pogledamo razporeditev stopenj vozlišč za celoten graf (slika 4.3), opazimo drugačen vzorec. Kot lahko pričakujemo, imamo vozlišča tudi s preko 6.000 povezavami, a ker se vzorec padanja števila vozlišč z večjim številom povezav ne spremeni bistveno in zaradi preglednosti, sem histogram omejil do vozlišč stopnje 100. Glede na sliko 4.1 smo lahko pričakovali večje število vozlišč z do šest povezavami, torej uporabnike, ki sledijo vsem strankam. Opazimo lahko relativno nizko število vozlišč z eno povezavo, predvsem če primerjamo s številom uporabnikov, ki sledijo samo eni stranki (tabela 4.1). Če primerjamo še s podatki s slike 4.2, kjer smo ugotovili, da se znotraj podgrafa posameznih strank sledilci večinoma ne povezujejo med seboj, lahko sklepamo, da se sledilci ene stranke povezujejo s sledilci druge stranke. Iz histograma vidimo tudi, da se upadanje števila vozlišč upočasni pri okoli 40 povezavah.

Predpostavko, da vozlišča z večjim številom povezav predstavljajo uporabniške račune novinarskih strani in znanih oseb, lahko preverimo, saj imamo seznam uporabnikov in njihovo število povezav. Ker za stranke že poznamo število sledilcev, sem jih iz seznama odstranil. Uporabniki z največ povezavami so:

- JJansaSDS (6408)
- 24ur.com (4748)
- vladaRS (4260)
- nmusar (4200)
- zzTurk (3902)
- KatarinaKresal (3888)
- Delo (3687)
- SpletnaMladina (3684)



Slika 4.2: Frekvence stopenj vozlišč po strankah.



Slika 4.3: Frekvence stopenj vozlišč za sledilce vseh strank.

- rtvslo (3643)
- BorutPahor (3635)

Ugotovimo, da predpostavka drži, saj so na seznamu politiki, informacijska pooblaščenka in novinarski računi.

## Poglavje 5

# Poslanci in Twitter

V tem delu naloge bom s parlamentarnih strank pozornost usmeril na poslance s Twitter računom. Dobil sem jih tako, da sem v Twitter iskalnik vpisal vsakega poslanca posebej [26] in si izpisal tiste, ki imajo Twitter račun. Ker nas še vedno zanima, kako parlamentarne stranke uporabljajo Twitter, sem iz analiz izvzel nepovezane poslance in poslanca narodnih skupnosti. V tabeli 5.1 je zapisano število poslancev s Twitter računom po strankah. Skupno ima Twitter račun 30 poslancev iz strank, med njimi jih ima največ SDS, kar 13, oziroma polovico vseh svojih poslancev. Procentualno gledano ima enak delež poslancev s Twitter računom NSi, približata se tudi DL in DeSUS. A, kot je razvidno iz tabele, več poslancev s Twitter računom še ne pomeni, da bodo ti poslanci tudi ustvarili več sporočil.

V tabeli 5.2 so prikazani Twitter računi poslancev in njihovo število tweetov po strankah. V prejšnjem odstavku smo ugotovili, da več računov ne pomeni več sporočil, zdaj pa lahko preverimo, koliko so posamezni poslanci aktivni na Twitterju. Pri SDS na primer vidimo, da trije niso nikoli nič objavili, štirje so objavili manj kot 10 tweetov, in še to zadnjič leta 2011. Pri DL pa opazimo, da je veliko večino tweetov prispeval en poslanec.

Naslednje, kar si bomo pogledali je, koliko sledilcev imajo poslanci. Oziroma, ali imajo poslanci iste sledilce kot stranke, katerim pripadajo. Pričakujem, da bo presek precej velik, saj predvidevam, da kdor sledi nekemu politiku,

Stranka	Št. Twitter računov	Št. poslancev	Št. tweetov
PS	6	27	4309
SDS	13	26	3217
SD	3	10	1241
DL	3	7	13707
DeSUS	2	5	26
SLS	1	6	50
NSi	2	4	317

Tabela 5.1: Poslanci na Twitterju (podatki s Twitterja dne 19. 8. 2013).

Stranka	Poslanci in število tweetov
PS	@MJaniPS (3094), @majusdimi (1015), @aljosajeric (108), @GObrane (86), @barbarazgajner (19), @RBrunskole (0)
SDS	@Andrejcus (1702), @drVinkoGorenak (791), @pucnik1 (305), @SONJARAMSAK (202), @MPogacnik (163), @BrankoGrims1 (41), @StefanTisel (8) @dragutinmate (3), @RoberHrovat (2), @IvanGrill (1), @KDanijel (0), @RomanaTomc (0), @BreznikFranc (0)
SD	@matevzfrangez (885), @ljubicajelusic (355), @BevkSamo (1)
DL	@jeanmark00 (13562), @PoloncaKomar (102), @KatarinaHocevar (44),
DeSUS	@IvanHrsak (23), @JanaJanca (3)
SLS	@JasminaOpec (50)
NSi	@LjudmilaNovak (188), @jozefhorvat (129)

Tabela 5.2: Poslanci s Twitter računom, urejeni po padajočem številu objav (podatki s Twitterja dne 19. 8. 2013).

Stranka	Št. različnih sledilcev	Št. sledilcev stranke	Presek
PS	1964	1857	901
SDS	4772	6735	2670
SD	2343	3062	1233
DL	1386	703	229
SLS	97	3642	63
NSi	1459	2048	820

Tabela 5.3: Sledilci poslancev in strank (podatki s Twitterja dne 17. 8. 2013).

bo sledil tudi njegovi stranki. Zaradi tega pričakujem tudi, da bodo stranke imele več sledilcev kot njihovi poslanci skupaj. Da bi preveril prvo hipotezo, sem primerjal množico sledilcev poslancev posameznih strank s sledilci stranke, ki ji pripadajo. Rezultat je prikazan v tabeli 5.3. Če izračunamo povprečen delež, ki ga predstavlja presek, ugotovimo, da se pri večini strank približa polovici, nižji je le pri SLS, kjer tveeta samo ena poslanka, in DL, kjer imajo poslanci dvakrat toliko sledilcev kot stranka. Pri teh dveh strankah predstavlja presek 33,3% oziroma 24,5%. S pomočjo prvih dveh stolpcev lahko ovržemo drugo hipotezo, saj imata dve stranki (PS in DL) manj sledilcev kot njuni poslanci skupaj.

## 5.1 O čem govorijo

Nadalje lahko primerjamo leksikalno raznolikost poslanskih tweetov. Pri strankah smo odkrili nizko leksikalno raznolikost, kar smo pripisali predvsem temu, da stranke Twitter uporabljajo predvsem za vabila na razne dogodke in k ogledu različnih oddaj. Leksikalno raznolikost sem računal na enak način kot v 3. poglavju, le da sem tokrat v analizo vzel tweete poslancev posameznih strank. Rezultati izračuna leksikalne raznolikosti so prikazani v tabeli 5.4 in jih lahko primerjamo z rezultati v tabeli 3.3. Opazimo, da je

Stranka	Leksikalna raznolikost			
	besed	omemb	hashtagov	$\bar{n}$
PS	0,36	0,14	0,54	14,0
SDS	0,39	0,20	0,49	13,2
SD	0,50	0,28	0,51	14,7
DL	0,37	0,14	0,84	12,2
DeSUS	0,72	1,0	0	11,7
SLS	0,80	0,64	1,0	12,0
NSi	0,59	0,38	0,87	13,1
Povprečje	0,53	0,40	0,61	13,0

Tabela 5.4: Leksikalna raznolikost besed, omemb in hashtagov ter povprečno število besed na tweet po strankah glede na poslance (podatki s Twitterja dne 17. 8. 2013).

tokrat povprečna leksikalna raznolikost za vse tri kategorije višja, manjše je pa povprečno število besed na tweet. To nakazuje, da poslanci uporabljajo Twitter za komuniciranje o večjem številu zadev kot same stranke.

S pogledom na najbolj pogoste besede v poslanskih tweetih (prikazane v tabeli 5.5), si lahko ustvarimo predstavo o tem, kaj poslanci pišejo. Opazimo, da so tudi pri njih retweeti zelo pogosti. Pogosto omenjajo tudi stranko, kateri pripadajo.

Med pogostimi omembami uporabnikov (prikazane v tabeli 5.6) ne preseneča, da so na seznamih računi strank. Zanimive so omembe športnikov, kar namiguje na to, da poslanci spremljajo aktualno športno dogajanje in ga komentirajo prek svojih Twitter računov.

Podobnost s strankami lahko opazimo tudi pri hashtagih, ki jih poslanci uporabljajo, saj je #soocenje pogosto uporabljen hashtag, tako kot pri strankah. Opazimo tudi Olimpijo in olimpijske igre, kar namiguje na to, da so med poslanci tudi športni navdušenci.



PozitivnaSlo	RT (2440), :) (313), @union_olimpija: (274), @PozitivnaSlo: (255), @union_olimpija (164), @TamaraVonta (144),
strankaSDS	RT (813), :) (249), @strankaSDS: (145), SDS (111), danes (93), lahko (85), #soocenje
strankaSD	RT (149), ;) (58), @multikultivator (50), @Maribor2020: (35), lahko (27)
strankaDL	RT (249), @petrasovdat (218), :-) (170), @Rok_Novak (128), lahko (127)
DeSUS	ZA (10), uspešno (7), Slovenijo! (6), DeSUS (5), glas (5)
strankaSLS	:) (6), RT (6), vedno (4)
NovaSlovenija	RT (59), NSi (32), @NovaSlovenija: (30), Slovencev (13), Ljudmila (11), Slovenija (10),

Tabela 5.5: Najbolj pogoste besede po strankah glede na poslance (podatki s Twitterja dne 17. 8. 2013).

PozitivnaSlo	@union_olimpija (467), @PozitivnaSlo (297), @TamaraVonta (289), @RomanJakic (232), @stanka_d (156), @MJaniPS (118)
strankaSDS	@strankaSDS (165), @MilanZver (77), @JJansaSDS (67), @organizacijaSDM (66), @JernejStromajer (45), @KlemenMesarec (41)
strankaSD	@multikultivator (54), @Maribor2020 (52), @PetraSlanic (30), @MatevzNovak (21), @krajncb (20), @MariborSD (18)
strankaDL	@petrasovdat (240), @Rok_Novak (148), @AlesGolli (98), @MladiMisleci (77), @Svarun_K (77), @rava_ (65)
DeSUS	@markodulmin (1)
strankaSLS	@strankaSLS (3), @JasminaOpec (3), @Sentisch (2)
NovaSlovenija	@NovaSlovenija (39), @TinaMaze (6), @MilanZver (5), @jozefhorvat (5), @RevijaReporter (4), @matjazta (4)

Tabela 5.6: Najbolj pogoste omembe po strankah glede na poslance (podatki s Twitterja dne 17. 8. 2013).

PozitivnaSlo	#soocenje (77), #OlimpijaOLE (19), #OSCE (17), #olimpija (12), #devotion (11), #kongresPS (11)
strankaSDS	#soocenje (86), #mislisvobodno (22), #predsednik12 (15), #oi2012si (14), #soocenje (6)
strankaSD	#soocenje (19), #mladiSD (13), #Maribor2020 (8), #kongresSD (6)
strankaDL	#fail (6), #Finance_si (6), #likealot (5), #delo (4), #mrezaidej (4)
strankaSLS	#zivaknjiznica (1), #podezelje (1), #vladni (1), #zivaknjiznica (1), #centralizem (1)
NovaSlovenija	#oi2012si (3)

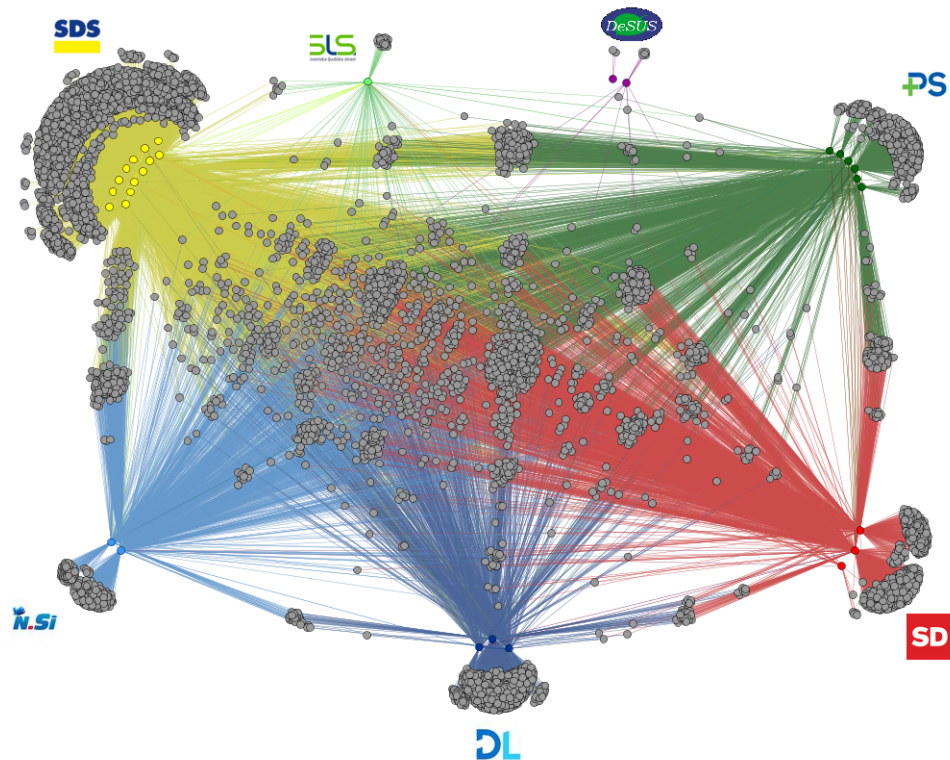
Tabela 5.7: Najbolj pogosti hashtagi po strankah glede na poslance (podatki s Twitterja dne 17. 8. 2013).

## 5.2 Povezave med sledilci poslancev

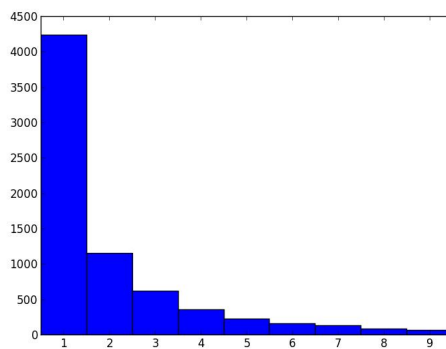
Na sliki 5.1 so prikazane povezave med poslanci in njihovimi sledilci. Pobarvane točke predstavljajo poslance, sive točke predstavljajo sledilce, povezava pa nakazuje sledenje. Podobno kot pri grafu sledilcev strank, opazimo precej uporabnikov, ki sledijo samo enemu poslancu ali poslancem ene stranke. Prisotnih je tudi kar nekaj uporabnikov v sredini grafa, torej takih, ki sledijo poslancem iz različnih strank. Opazimo lahko zanimivost, da ima ena poslanka iz DeSUS samo dva sledilca, ki ne sledita drugim poslancem.

Tudi pri tem grafu lahko pogledamo frekvence stopenj vozlišč (kot je prikazano na sliki 5.2). Tudi tu sem skrajšal abscisno os, saj so vozlišča z več povezavami redka, pa še ta praviloma pripadajo poslancem. Histogram nam potrди, da veliko uporabnikov sledi samo enemu poslancu.

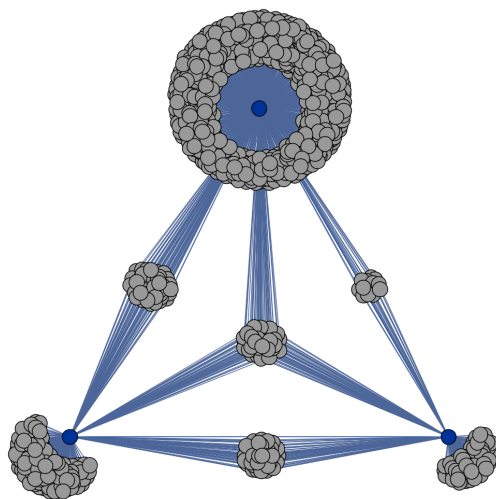
V nadaljevanju (slike 5.3 do 5.7) si bomo podrobneje pogledali povezave med sledilci poslancev iste stranke. Iz tega pregleda sta izvzeta SLS, ker Twitter uporablja samo ena poslanka, ter DeSUS, ker ni povezav med poslancema s Twitter računom. Pri vseh strankah opazimo, da število sledilcev



Slika 5.1: Graf sledilcev poslancev na Twitterju.

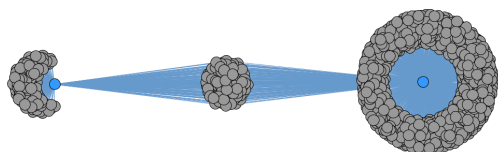


Slika 5.2: Frekvence stopenj vozlišč za sledilce poslancev.

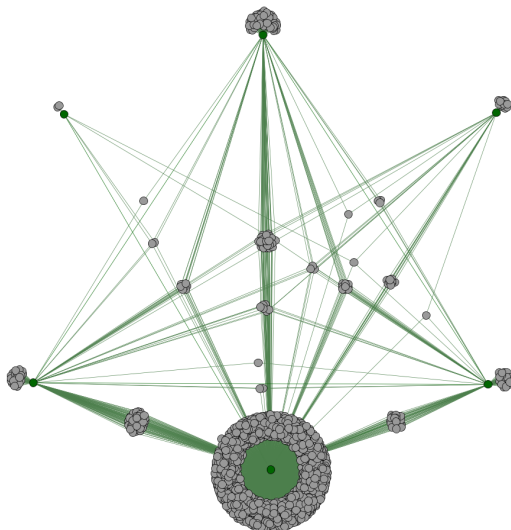


Slika 5.3: Graf sledilcev poslancev DL na Twitterju.

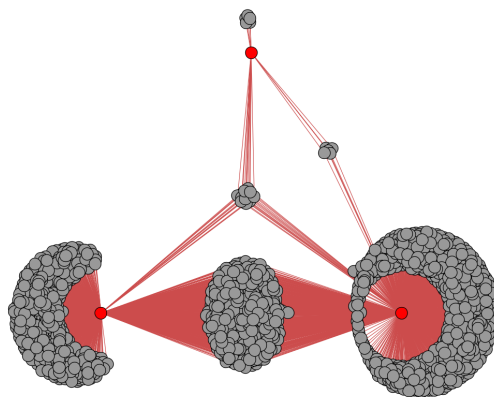
ni enakomerno porazdeljeno med vsemi poslanci, temveč obstajajo manj in bolj popularni. Morda bi lahko pričakovali, da bodo uporabniki, ki ne sledijo samo enemu poslancu iz stranke, sledili vsem poslancem iz te stranke, a iz grafov je razvidno, da temu ni tako. Kar nekaj uporabnikov namreč sledi dvema poslancema, ne pa tudi tretjemu, to je še posebej opazno pri strankah DL (slika 5.3) in SD (slika 5.6).



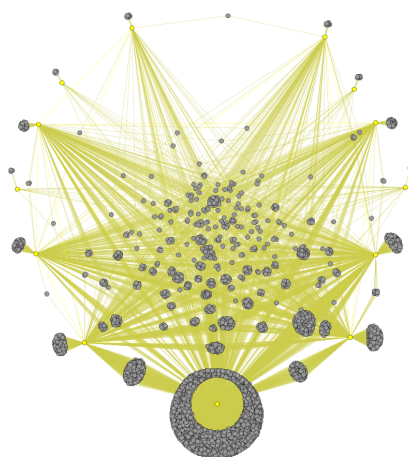
Slika 5.4: Graf sledilcev poslancev NSi na Twitterju.



Slika 5.5: Graf sledilcev poslancev PS na Twitterju.



Slika 5.6: Graf sledilcev poslancev SD na Twitterju.



Slika 5.7: Graf sledilcev poslancev SDS na Twitterju.





# Poglavje 6

## Zaključek

Zaradi velikega števila uporabnikov je Twitter zanimiva platforma za komuniciranje med politiki in volivci. Politično komuniciranje je zanimivo tudi za raziskovalce in analitike, ki želijo s pomočjo družbenega spleta na cenejši in hitrejši način razbrati voljo ljudstva.

V uvodnem poglavju sem zastavil nekaj vprašanj, na katera sem v nalogi iskal odgovore. Za konec si še enkrat pogledajmo ta vprašanja z odgovori.

### **Katere stranke so najbolj aktivne?**

Aktivnost sem definiral kot povprečno število objavljenih tweetov na dan, od dneva, ko je stranka ustvarila Twitter račun. Po taki definiciji je najbolj aktivna stranka PS, najmanj pa NSi.

### **Ali obstaja povezava med popularnostjo stranke na Twitterju in njenim izidom na volitvah?**

Z izračunom koreliranosti nisem odkril povezave med številom sledilcev in izidom na volitvah. Opaziti pa je povezavo med številom objav in številom sledilcev, številom sledilcev in časom obstoja računa ter aktivnostjo in izidom na volitvah.

### **O čem pišejo stranke?**

Z uporabo leksikalne raznolikosti smo ugotovili, da stranke razpravljajo o manjšem obsegu tem. Poleg tega smo odkrili, da pogosto retweetajo, predvsem svoje člane in novinarje, ter da uporabljajo Twitter predvsem za

sporočanje sledilcem o raznih dogodkih.

### **Ali se politična podobnost strank kaže tudi na Twitterju?**

Podobnost med strankami sem računal na osnovi omemb in hashtagov, za mero razdalje pa sem vzel povprečno kosinusno razdaljo. Izračun je pokazal, da so si stranke precej različne, pri uporabi hashtagov sta si najbližje celo stranki iz nasprotnih polov.

### **Ali se da iz skupnih sledilcev sklepati na podobnost strank?**

Zaradi različnega števila sledilcev strank se iz absolutnega števila skupnih sledilcev dveh strank ne vidi podobnosti, ki bi odražala resnično. Če pa primerjamo deleže, ki jih predstavljajo skupni sledilci, najdemo tako podobnost in različnost, kot se kaže tudi v realnem svetu.

### **Kako so sledilci strank povezani med seboj?**

Sledilci strank se povezujejo s sledilci drugih strank, med sledilci pa zasledimo tudi uporabnike z večjim številom povezav, ti pa predstavljajo znane osebe in novinarske račune.

### **Ali imajo poslanci iste sledilce kot njihova stranka?**

Pri vseh parlamentarnih strankah velja, da imajo nekaj skupnih sledilcev s svojimi poslanci, a pri nobeni stranki ni povprečen delež večji od polovice, najnižji povprečni delež pa je samo 24,5%.

### **O čem pišejo poslanci?**

Pri tweetih poslancev smo odkrili višjo leksikalno raznolikost kot pri tweetih strank, kar nakazuje na to, da poslanci prek Twitterja komunicirajo o večjem številu tem, med drugim pišejo tudi o športu. Tudi pri poslancih opazimo veliko retweetov in omemb njihovih strank ter drugih članov strank.

### **Ali sledilci poslancev sledijo vsem poslancem iz iste stranke?**

Ugotovili smo, da največ (preko 4000) sledilcev poslancev sledi samo enemu poslancu. Iz grafov strank smo videli tudi, da se znotraj iste stranke sledilci ne razporedijo enakomerno med poslance, temveč obstajajo poslanci s precej več sledilci kot drugi.

Kot smo videli, stranke in njihovi poslanci uporabljajo Twitter z različno aktivnostjo, a ugotovili smo nekaj podobnih načinov uporabe (zlasti vabila

k ogledu oddaj in na dogodke). Zaradi kratkosti sporočil Twitter morda ni najprimernejši medij za razpravljanje o aktualnih tematikah, ponuja pa hiter in javen način za izražanje lastnega mnenja. Ponuja tudi dodaten stik med politiki in volivci, kjer je komunikacija možna v obe smeri. Vprašanje je le, koliko sta obe strani pripravljene ta stik izkoristiti.



# Literatura

- [1] K. Wickre. (2013) Celebrating #Twitter7. [Online]. Available: <https://blog.twitter.com/2013/celebrating-twitter7>
- [2] Twitter. (2013) Borut Pahor (BorutPahor) on Twitter. [Online]. Available: <https://twitter.com/BorutPahor>
- [3] Twitter api v1.1 resources. [Online]. Available: <https://dev.twitter.com/docs/api/1.1>
- [4] M. Verdone. Python twitter tools. [Online]. Available: <http://mike.verdone.ca/twitter/>
- [5] Python. [Online]. Available: <http://www.python.org/>
- [6] Networkx. [Online]. Available: <http://networkx.github.io/>
- [7] Gephi. [Online]. Available: <https://gephi.org/>
- [8] MongoDB. [Online]. Available: <http://www.mongodb.org/>
- [9] Pymongo 2.6. [Online]. Available: <http://api.mongodb.org/python/current/>
- [10] matplotlib python plotting. [Online]. Available: <http://matplotlib.org/>
- [11] Numpy. [Online]. Available: <http://www.numpy.org/>
- [12] Scipy. [Online]. Available: <http://www.scipy.org/>

- 
- [13] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welp, "Predicting elections with twitter: What 140 characters reveal about political sentiment." *ICWSM*, vol. 10, pp. 178–185, 2010.
- [14] J. Soler, F. Cuartero, and M. Roblizo, "Twitter as a tool for predicting elections results," in *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, 2012, pp. 1194–1200.
- [15] J. E. Chung and E. Mustafaraj, "Can collective sentiment expressed on twitter predict political elections?" in *AAAI*, 2011.
- [16] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series." *ICWSM*, vol. 11, pp. 122–129, 2010.
- [17] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj, "Limits of electoral predictions using twitter." in *ICWSM*, 2011.
- [18] P. Metaxas, E. Mustafaraj, and D. Gayo-Avello, "How (not) to predict elections," in *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, 2011, pp. 165–171.
- [19] F. Bravo-Marquez, D. Gayo-Avello, M. Mendoza, and B. Poblete, "Opinion dynamics of elections in twitter," in *Web Congress (LA-WEB), 2012 Eighth Latin American*, 2012, pp. 32–39.
- [20] G. System. Predsedniške volitve 2012. [Online]. Available: <http://predsedniskevolitve.si/>
- [21] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, "Political polarization on twitter." in *ICWSM*, 2011.
- [22] M. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer, "Predicting the political alignment of twitter users," in *Privacy*,

*security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, 2011, pp. 192–199.

- [23] M. Song and M. C. Kim, “RT<sup>2</sup>M: Real-time twitter trend mining system,” in *Social Intelligence and Technology (SOCIETY), 2013 International Conference on*, 2013, pp. 64–71.
- [24] D. volilna komisija. (2011) Predčasne volitve v državni zbor 2011. [Online]. Available: <http://www.dvk-rs.si/arhivi/dz2011/>
- [25] M. A. Russell, *Mining the Social Web, 2nd Edition*, 2013.
- [26] Državni zbor - po abecedi. [Online]. Available: <http://www.dz-rs.si/wps/portal/Home/ODrzavnemZboru/KdoJeKdo/PoslankeInPoslanci>