

UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE

Borut Penko

**Analiza slovenskih uporabnikov družbenega omrežja
Twitter**

Diplomsko delo

Ljubljana, 2014

UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE

Borut Penko

Mentor: doc. dr. Aleš Žiberna

**Analiza slovenskih uporabnikov družbenega omrežja
Twitter**

Diplomsko delo

Ljubljana, 2014

*Zahvaljujem se mentorju doc. dr. Alešu Žiberni
za čas, pomoč in nasvete pri pisanju diplome.*

Analiza slovenskih uporabnikov družbenega omrežja Twitter

S pojavom in širjenjem različnih socialnih omrežij na spletu, prihaja do vprašanj o merjenju in analizi takih omrežij. Tu kot metoda raziskovanja v poštev pride analiza socialnih omrežij, ki je že do sedaj veljala za pomembno metodološko orodje. Pri tej je pomembno predvsem opazovanje enot in relacij med njimi. Namen diplomske naloge je tako v prvem koraku oblikovanje orodja za pridobivanje podatkov izbranega spletnega socialnega omrežja, v drugem koraku pa analiza takega omrežja s pomočjo metode analize socialnih omrežij. Omrežje, ki ga bom analiziral je podomrežje spletnega socialnega omrežja Twitter, predstavljajo pa ga aktivni slovenski uporabniki tega omrežja. Podatki so zbrani z orodjem napisanim v programskem jeziku Python, kjer je bila uporabljena knjižnica Tweepy, s pomočjo katere lahko preko vmesnika, API-ja komuniciramo s Twitterjevimi. Analiza omrežja je bila opravljena s programom Pajek, v katerem sem iskal pomembne enote v omrežju in analizirali povezave med enotami.

Ključne besede: analiza socialnih omrežij, spletna socialna omrežja, vizualizacija socialnih omrežij, podatkovno rudarjenje, Twitter, Pajek, Python, Tweepy.

The analysis of slovenian users of social networking site Twitter

With the emergence and spreading of different social networks on the web, the question that occurs is how to measure and analyse these kind of networks. The method of social network analysis, considered as an important methodological tool, proves as an applicable option. Especially important aspect of this method is to observe actors and the relations among them. The purpose of this paper is firstly to create an application for gathering data from chosen online social network, and secondly the analysis of this social network. The network I chose to analyse is a subnetwork of an online social network Twitter and is composed of active slovenian Twitter users. Data is gathered with application written in programming language Python and opensourced Python library Tweepy, which allows us to communicate with Twitter through its API. Analysis of social network was conducted with program Pajek, which helped me to search for important actors in the network and to analyse relation between these actors.

Keywords: social network analysis, online social networks, visualization of social networks, data mining, Twitter, Pajek, Python, Tweepy.

KAZALO

1 UVOD	7
2 ANALIZA SOCIALNIH OMREŽIJ.....	9
2.1 ZGODOVINA ANALIZE SOCIALNIH OMREŽIJ	10
2.2 PRIKAZ (VIZUALIZACIJA) SOCIALNIH OMREŽIJ.....	11
2.2.1 PAJEK – PROGRAM ZA ANALIZO IN VIZUALIZACIJO VELIKIH OMREŽIJ	12
2.3 ZNAČILNOSTI SOCIALNIH OMREŽIJ	13
2.4 STOPNJE TOČK.....	14
2.5 KOHEZIVNE PODSKUPINE	14
2.5.1 KOMPONENTE OMREŽJA.....	15
2.5.2 JEDRA	15
2.5.3 OTOKI	15
2.6 MERE SREDIŠČNOSTI IN POMEMBNOSTI.....	15
2.6.1 MERE SREDIŠČNOSTI	16
2.6.3 MERE POMEMBNOSTI	16
3 SPLETNA SOCIALNA OMREŽJA.....	17
3.1 TWITTER.....	18
3.1.1 ZNAČILNOSTI TWITTERJA	19
3.1.2 POVEZOVANJE UPORABNIKOV V SPLETNEM SOCIALNEM OMREŽJU TWITTER	19
3.1.3 PODATKOVNO RUDARJENJE PO TWITTERJU.....	20
4 VZOREC IN ZBIRANJE PODATKOV.....	22
4.1 PYTHON.....	22
4.2 TWEETPY IN TWITTER API.....	23
4.3 POTEK IN KORAKI V PROCESU ZBIRANJA PODATKOV	23
5 ANALIZA OMREŽJA.....	26
5.1 SESTAVA IN ZNAČILNOSTI OMREŽJA	26
5.2 STOPNJE TOČK OMREŽJA	27
5.3 KOHEZIVNE PODSKUPINE OMREŽJA.....	28
5.3.1 KOMPONENTE	28
5.3.2 JEDRA	28
5.3.3 OTOKI	28
5.4 MERE POMEMBNOSTI IN USREDINJENOST OMREŽJA.....	30

5.5 MERE OBMOČJA VPLIVA IN BLIŽINE (PRESTIŽA)	31
5.6 PRIMERJAVA PODOBNOSTI MER POMEMBNOСТИ	32
6 ZAKLJUČEK	33
7 LITERATURA	36
PRILOGE	40
PRILOGA A: KODA – ZBIRANJE TWEETOV	40
PRILOGA B: KODA – ZBIRANJE FOLLOWEES	41
PRILOGA C: KODA – ZBIRANJE FOLLOWERS	43
PRILOGA Č: PAJEK – OSNOVNE LASTNOSTI OMREŽJA	45
PRILOGA D: PAJEK – STOPNJE TOČK	45
PRILOGA E: PAJEK – MERE SREDIŠČ. IN POMEMBNOСТИ	46
PRILOGA F: PAJEK – OBMOČJE VPLIVA IN BLIŽINA	47
PRILOGA G: PAJEK – KAZALA IN OPISI	48

KAZALO TABEL

Tabela 5.1: Stopnja točk desetih najbolj aktivnih enot (razvrščeno po skupni stopnji točk)..	27
Tabela 5.2: Prikaz mer pomembnosti za najpomembnejše enote	30
Tabela 5.3: Mere usredinjenosti omrežja	30
Tabela 5.4: Vrednosti najboljših kazal in opisev	32
Tabela 5.5: Primerjava mer pomembnosti	32

KAZALO TABEL

Slika 5.1: Najdaljša najkrajša pot v omrežju	26
Slika 5.2: Prikaz otokov velikosti 30 do 100 enot (Prikaz s sivinami).....	29
Slika 5.3: Primer zanimivega otoka v omrežju (Prikaz s sivinami)	29

1 UVOD

Analiza socialnih omrežij se v družbenih in vedenjskih znanostih uporablja že nekje od sredine tridesetih let prejšnjega stoletja. Čeprav so bili začetki počasni pa je v zadnjih desetletjih zanimanje za metodo močno naraslo. Z zanimanjem za analizo socialnih omrežij je prišlo tudi do ogromnih napredkov in tehničnih izboljšav, ki so danes na voljo uporabnikom (Wasserman in drugi 2005, 1). Metoda je v družboslovju poleg bolj klasičnih metod statistike postala močno orodje, s katerim je mogoče preučevati povezave v različnih tipih omrežij, ki jih lahko sestavljajo ljudje, različne organizacije, države ali druge entitete. Mnogi raziskovalci v družboslovju so mnenja, da so medsebojne vezi v takih omrežjih pomembne, preučevanje teh kot struktur pa je lahko bogat vir informacij (de Nooy in drugi 2011, 1–3).

Primer nekega tipa socialnih omrežij tako lahko predstavljajo tudi različna spletna socialna omrežja. S pojavom in razvojem spleta smo torej pridobili nov vir socialnih omrežij, ki jih danes poznamo predvsem kot spletne strani ali aplikacije, ki uporabnikom omogočajo povezovanje na podlagi različnih interesov, zanimanj ali celo kulturnih in drugih ozadij. Sestavni del takih spletnih aplikacij so različna orodja, ki omogočajo povezovanje, komunikacijo, delitev različnih vsebin, komentiranje ter izmenjavo mnenj (Boyd 2011, 39–44).

Eno takih spletnih socialnih omrežij, ki je v zadnjem času izredno priljubljeno, je tudi Twitter. Gre za novejše socialno omrežje, ki temelji na mikrobloggingu oziroma izmenjavi omejenih kratkih sporočil. To omogoča hitrejši način komunikacije, kjer so objave lahko produkt nekega trenutka, ali pa dobro načrtovana sporočila. Omrežje deluje na deljenju tweetov med osebami, ki si med seboj sledijo, vendar ne nujno obojestransko (Russell 2013; Huberman in drugi 2009).

Za analizo slovenskih uporabnikov spletnega omrežja Twitter sem odločil iz več razlogov. Prvi izmed teh so sama spletna omrežja, ki predstavljajo zanimiv fenomen in vir informacij, z naborom katerih lahko preučujemo povezovanje ključnih oseb in njihovo vlogo v takih omrežjih. Twitter je prosto dostopen in tako omogoča spletno rudarjenje in pridobitev podatkov na relativno lahek način. Po drugi strani pa me je k nalogi pritegnilo ravno oblikovanje nekega uporabnega orodja s pomočjo programerskega jezika Python. Tako sem s pomočjo Pythona in temu prirejene knjižnice Tweepy sprogramiral aplikacijo, ki komunicira s Twitterjevim vmesnikom (API-jem), preko katerega lahko iz Twitterja izvlečemo zelene podatke. Uporabnike sem pridobil z zbiranjem tweetov v časovnem obdobju med 12.

novembrom in 12. decembrom 2013, pri čemer sem lokacijske meje zbranih sporočil začrtal s pomočjo geokodiranja. Tu sem izbral območje, ki ima center določen s koordinatami zemljepisne dolžine 46.05 in zemljepisne širine 14.5 (Ljubljana) ter v radiju 60 kilometrov. S tem postopkom sem pridobil nekakšno Twitterjevo podomrežje, sestavljeno iz uporabnikov, ki so v izbranem obdobju in lokaciji predstavljali aktivne uporabnike Twitterja. Potrebno pa je poudariti, da pridobivanje tweetov na ta način ne zajame vseh aktivnih uporabnikov, pač pa le tiste, ki omogočijo funkcijo beleženja lokacije, iz katere je tweet poslan. Po nekaterih ocenah pridobljenih skozi testiranja, naj bi beleženje geolokacije bilo prisotno v 35 do 50 odstotkih slovenskih tweetov (Martinc 2013, 12). Vseeno sem ocenil, da je ta opcija bolj primerna kot pridobivanje uporabnikov glede na njihovo navedbo lokacije. Veliko uporabnikov namreč ne beleži tega podatka ali pa je ta izmišljen oziroma ni napisan pravilno. Končno omrežje tako sestavljajo aktivni slovenski uporabniki Twitterja v radiju 60 kilometrov okoli Ljubljane.

Program, v katerem sem analiziral svoje generirano omrežje, je Pajek. Ta predstavlja orodje za analizo in vizualizacijo velikih omrežij. S pomočjo programa in nekaterih metod analize bom poskušal ugotoviti in prikazati nekatere zanimive lastnosti omrežja. Kdo sledi največjemu številu oseb v našem omrežju, katera oseba ima največ sledilcev, kdo ima največji vpliv, ter kdo največjo podporo. Preučeval bom torej le povezovanje oseb v omrežju, ter poskusil prikazati nekaj tistih, ki v omrežju igrajo ključno vlogo.

Raziskovalni vprašanji sta torej:

Katere od enot v omrežju imajo največji vpliv in katere največjo podporo?

Ali se v omrežju pojavljajo kakšne podskupine?

2 ANALIZA SOCIALNIH OMREŽIJ

Pojem socialna omrežja in posledično metode analize socialnih omrežij so v zadnjih desetletjih pridobile precejšno pomembnost in se v družbeni in vedenjski znanosti uveljavile kot pomemben raziskovalni pristop. Velik del tega lahko pripišemo predvsem cilju metode analize socialnih omrežij, ki se fokusira na preučevanje odnosov med enotami, ter analizo vzorcev in posledic, ki jih taki odnosi prinašajo. Tak pristop omogoča preučevanje različnih okolij, v katerih nastopajo enote, ki so med seboj v relaciji. Z vidika analize socialnih omrežij so določena družbena okolja namreč predstavljena kot pravilni vzorci ali strukture v odnosih med enotami, ki so v medsebojni interakciji (Wasserman in Faust 1994, 3; de Nooy in drugi 2011, 3).

Analiza socialnih omrežij se tako osredotoča na preučevanje povezav med enotami, ki so lahko ljudje, skupine ljudi, organizacije ali druge izbrane entitete. Vezi s povezovanjem enot tvorijo omrežja, ki jih z metodami analize socialnih omrežij tudi preučujemo. Družbene znanosti se namreč osredotočajo na strukture, ki jih lahko predstavljajo skupine ljudi, skupnosti, organizacije, trgi, družbe itn. Ključni vidik analize socialnih omrežij je torej ta, da so povezave med enotami pomembne, ne glede na tip enot, ki sestavljajo neko omrežje. Iz takih povezav je namreč mogoče razbrati vedenja, odnose ali informacije. Gre torej za metodologijo, ki omogoča analizo družbenih relacij (de Nooy in drugi 2011, 1–3).

V današnjem času z napredkom računalniških tehnologij analiza socialnih omrežij predstavlja vse bolj priljubljeno metodo raziskovanja družbenih pojavov. Še posebej primerna je kot metoda za analizo večjih omrežij, kar jim danes omogoča napredek računalniških zmogljivosti (Hlebec in Kogovšek 2006, 7). Spletna socialna omrežja, ki so v zadnjem desetletju močno pridobila na priljubljenosti in jih tako lahko sestavlja ogromno število enot, pa se izkažejo kot primeren kandidat za uporabo take metode.

Prednost uporabe analize socialnih omrežij se pokaže tudi v tem, da jo lahko uporabimo kot samostojno metodo ali v povezavi z bolj klasično sociološko analizo. V prvem primeru lahko preučujemo socialna omrežja različnih tipov enot in strukture, ki se pojavljajo znotraj takih omrežij. Torej na kakšne načine se člani takšnih omrežij med seboj povezujejo, kakšna je moč teh povezav ter v čem so si podobni oziroma se med seboj razlikujejo. V drugem primeru pa lahko analiziramo, kako se mnenja in stališča o določenih stvareh razširjajo, spreminjajo, kateri člani omrežja so za takšno širjenje ključni, ter katere so skupne značilnosti teh ključnih članov omrežja (Hlebec in Kogovšek 2006, 7–8).

2.1 ZGODOVINA ANALIZE SOCIALNIH OMREŽIJ

Analiza socialnih omrežij je v osnovi interdisciplinaren produkt. Koncept metode se je razvil s prepletom različnih področij, od družbene teorije do matematičnih, statističnih in računalniških ved. Posamezni koncepti, ki sestavljajo analizo socialnih omrežij, kot so relacija, omrežje in struktura, so se na posameznih področjih že prej pojavljali neodvisno. Tako tudi začetniki metode prihajajo iz različnih disciplin (Wasserman in Faust 1994, 10).

Freeman (2011, 26) pravi, da pristop analize socialnih omrežij vsebuje štiri lastnosti, ki ga definirajo:

- predpostavko, da so vezi med družbenimi akterji pomembne;
- osnovanost na zbirki in analizi podatkov, kjer so zbrani družbeni odnosi, ki povezujejo akterje;
- osnovanost na grafičnem posnetkih, ki morajo razkriti in prikazati vzorce teh povezav;
- razvijanje matematičnih modelov, ki opisujejo in razlagajo te vzorce.

Pravi, da naj bi se že pred 13. stoletjem pojavljala dela, ki so se naslanjala na vsaj eno izmed teh lastnosti. Šele okoli leta 1930 pa naj bi bile uporabljene vse štiri lastnosti hkrati, s čimer se pojavi sodobno področje analize socialnih omrežij.

V prvi modernejši obliki je bila analiza socialnih omrežij predstavljena s strani Jacoba L. Morena in Helene Jennings, ki sta delovala na področju psihologije. Svoj pristop sta poimenovala sociometrija. Ta je sprva privabil določen interes, ki pa ni trajal dolgo, saj so se raziskovalci na področju psihologije in sociologije kmalu vrnili k bolj tradicionalnim metodam, ki so se osredotočale bolj na lastnosti posameznika. V enakem časovnem obdobju sta se pristopa analize socialnih omrežij lotili še dve ekipi. Prvo je vodil antropolog W. Lloyd Warner, druga pa je delovala pod okriljem psihologa Kurta Lewina. Nobena od treh začetnih ekip pa ni uspela razviti metode oziroma pristopa, ki bi se razširil in obdržal na različnih področjih raziskovanja (Freeman 2001, 26–27).

Vpeljavo izraza socialna omrežja v znanstvenem kontekstu se večinoma pripisuje antropologu Johnu Barnesu, ki je pojem uporabil v članku »Class and Committees in a Norwegian Island Parish« leta 1954. Barnes in ostali člani manchesterske šole veljajo za ene prvih, ki so s svojim terenskim delom preučevanja različnih skupnosti začrtali pot analizi socialnih omrežij. Delu Barnesu je kasneje sledil tudi Siegfried Nadel, psiholog, katerega delovanje na področju teorije socialne strukture je vplivalo na kasnejši razvoj analize socialnih omrežij. Prav tako je

eno ključnih vlog v razvoju analize socialnih teorij odigralo delo raziskovalcev Harvarda, ki jih je vodil profesor Harrison White (Wasserman in Faust 1994, 10; Scott 2000, 28–37).

V Sloveniji se z raziskovanjem socialnih omrežij prvič srečamo kar nekaj desetletij kasneje in sicer šele v drugi polovici osemdesetih let. Tudi takrat pa je pravzaprav to raziskovanje imelo drugačno konceptualno ozadje, kot ga poznamo danes (Iglič v Novak 2004, 7). Šele v devetdesetih letih prejšnjega stoletja lahko govorimo o obdobju, ko je raziskovanje socialnih omrežij pri nas dobilo teoretsko in metodološko konceptualizacijo, predstavljenih pa je bilo tudi več del o metodah, merjenju in aplikacijah na tem področju. Proti koncu tisočletja pa se je pričelo tudi z praktičnim testiranjem prej zgolj teoretsko in metodološko utemeljenih orodij (Novak 2004, 7).

2.2 PRIKAZ (VIZUALIZACIJA) SOCIALNIH OMREŽIJ

Vizualizacija in orodja za vizualni prikaz so pomembna komponenta za znanstveni razvoj na različnih področjih raziskovanja. Iz preprostega grafičnega prikaza je mogoče razbrati ogromno informacij, ne glede na to ali gre za linijski graf, matriko ali piramidni prikaz podatkov (Chu in drugi, 2013, 2). Zgodovinar Alfred Crosby je celo mnenja, da je vizualizacija oziroma prikaz poleg meritve, najpomembnejši element v razvoju celotne moderne znanosti (Crosby v Freeman 2000, 1).

Oba elementa igrata središčno vlogo tudi na področju analize socialnih omrežij. Merjenje oziroma razvoj postopkov merjenja je v raziskovanju socialnih omrežij pogost in reden pojav (Wasserman in Faust 1994, 10–27). Prav tako pa že od začetka razvoja področja vizualni prikaz predstavlja ključno vlogo (Wolfe; Hage in Harary v Freeman 2000, 1).

Za vizualni prikaz socialnih omrežij sta uporabljena predvsem dva različna načina. Prvi temelji na točkah in premicah, kjer točke ponavadi predstavljajo družbene akterje, premice pa povezave med akterji. Ta način je za prikaz socialnih omrežij bolj pogosta izbira, razvil pa se je v petih fazah, kjer so bili prikazi sprva ročno delo, z razvojem računalniških tehnologij pa se je povečalo tudi število programov in orodij, ki omogočajo bolj zmogljivo računalniško vizualizacijo omrežij. Zadnja faza razvoja vizualizacije omrežij omogoča prikaz teh z uporabo barv, animacijami v tridimenzionalnem prostoru ali celo s prikazom v spletnem brskalniku. Druga manj pogosta metoda prikaza pa so matrike, kjer vrstice in stolpci predstavljajo družbene akterje, števila ali simboli v celicah matrike pa prikazujejo družbene relacije med akterji (Freeman 2000, 1–13).

Danes se pojavlja več različnih programskih paketov za analizo socialnih omrežij, ki se razlikujejo po svoji osnovni namembnosti in karakteristikah. Inflow je namenjen mapiranju omrežij, NetDraw in NetMiner vizualizaciji, Pajek vizualizaciji velikih omrežij, StOCNET statistični analizi omrežij, UCINET pa deluje kot celovito orodje. Namen orodij za vizualizacijo socialnih omrežij je pretvorba zapletenih analiz podatkov v transparenten in razumljiv prikaz, ki ga uporabnik lažje razume. S takim procesom vizualizacije lahko torej prikazujemo celotna omrežja, podomrežja, izvajamo različne vizualno analitske tehnike za odkrivanje anomalij ali zgolj raziskujemo sestavo omrežja. Orodja, ki so namenjena predvsem vizualizaciji omrežij lahko razdelimo v tri različne skupine, glede na način prikaza. Node-Link prikaz: Graphviz, Guess, Igraph, InFlow, Pajek, Ucinet, Vizster. Hibridni prikaz: MatLink in NodeTrix. Cluster(blockmodeling): Treemap (Dráždilová in ostali 2010, 291–310).

2.2.1 PAJEK – PROGRAM ZA ANALIZO IN VIZUALIZACIJO VELIKIH OMREŽIJ

Velika omrežja, ki imajo na tisoče ali celo več točk oziroma »akterjev« in še več povezav med njimi, lahko najdemo na različnih področjih raziskovanja: molekule v kemiji, spletna socialna omrežja, omrežja organizacij itd. Zaradi velikosti teh omrežij so nekateri standardni postopki in algoritmi za analizo takih omrežij neprimerni, saj bi analiza z njimi vzela preveč časa in pomnilniških kapacitet. Tako je pomemben razvoj nekaterih algoritmov in programskih orodij, ki analizo takih omrežij sploh omogočajo (Batagelj in Mrvar 1998, 48).

Pajek je programsko orodje, ki deluje v operacijskem sistemu Windows in je za namene nekomercialne uporabe dostopno zastonj. Namenjeno je analizi in vizualizaciji velikih omrežij, ki lahko vsebujejo tisoče ali celo milijone točk. Razvoj Pajka se je začel novembra leta 1996 in je bil ustvarjen s strani Andreja Mrvarja in Vladimirja Batagelja ter z nekaj pomoči Matjaža Zaveršnika. Glavni povod za nastanek programa je bil obstoj takih virov za generiranje velikih omrežij, ki so že bili računalniško berljivi. Pajek tako vsebuje orodja za analizo in vizualizacijo različnih tipov omrežij: spletna omrežja, rodovniki in druga sorodstvena omrežja, organske molekule v kemiji, omrežja citiranj, transportna omrežja itd (Mrvar in Batagelj 2014, 3–5).

Glavni cilji in nameni, ki jih program omogoča, so:

- podpreti abstrakcijo z (rekurzivno) razčlenitvijo velikega omrežja na več manjših omrežij, ki jih je mogoče nadalje analizirati z uporabo nekaterih bolj primernih metod;

- uporabniku omogočiti prikaz omrežij z različnimi zmogljivimi vizualizacijskimi orodji;
- nuditi nabor različnih učinkovitih algoritmov za analizo velikih, bolj obsežnih omrežij.

Tako je s Pajkom v izbranem omrežju mogoče najti skupine oziroma podomrežja (komponente, sosede bolj pomembnih točk v omrežju, jedra, itn.), lahko najdemo točke, ki pripadajo istemu podomrežju, ter jih predstavimo posebej ali kot dele nekega konteksta, lahko prikažemo povezave med skupinami ali odkrivamo ključne akterje omrežja (Mrvar in Batagelj 2014, 3–5).

2.3 ZNAČILNOSTI SOCIALNIH OMREŽIJ

Socialna omrežja so definirana z različnimi lastnostmi, ki nam lahko pomagajo pri izbiri pravih postopkov in analiz pri obravnavi določenega omrežja. Ključne osnovne lastnosti po katerih omrežja razlikujemo so velikost, gostota, ter različni tipi omrežij (Wasserman in Faust 1994, 28–42).

Ena izmed delitev je na egocentrična in popolna omrežja. Pri slednjih opazujemo vse enote in njihove relacije. Tu nas zanimajo različne mere, kot so gostota, mere središčnosti in pomembnosti, lahko pa si pri analizi takega omrežja pomagamo tudi z bločnim modeliranjem. Za najprimernejše mere se odločamo glede na raziskovalen problem. Egocentrično omrežje pa sestavlja nek začetni seznam enot, ki jih imenujemo egi. V analizah se osredotočamo na omrežje nekega ega in na povezave ega z ostalimi enotami, ki jih imenujemo alterji. Ponavadi gre pri proučevanju v egocentričnih omrežjih za opazovanje slučajno izbranih enot (Hlebec in Kogovšek 2006, 11–12).

Neusmerjena omrežja so tista, pri katerih so relacije med točkami simetrične. Pri usmerjenih omrežjih pa so relacije med enotami usmerjene, pomembna je torej smer povezave. V nekaterih primerih je za željene rezultate potrebna tudi manipulacija relacij. Dve usmerjeni povezavi med parom točk lahko zamenjamo za neusmerjeno, včasih pa storimo obratno (de Nooy in drugi 2011, 7–14).

Omrežja lahko ločimo na mala, srednja in velika. Mala omrežja so tista z nekaj 10 enotami, srednja z nekaj 100 enotami, velika omrežja pa so tista z več kot 1000 enotami (de Nooy in drugi 2011, 12). Velika omrežja je mogoče najti na različnih področjih raziskovanja. Podatke lahko dobimo iz različnih virov, ponavadi pa so že dostopni v računalniški obliki. Kljub

računalniškimi zmogljivostim, pa velika omrežja lahko povzročajo težave pri računanju nekaterih algoritmov (Batagelj in Mrvar 1998, 47–51).

Večje število povezav v omrežju naj bi pomenilo boljše povezanost, saj gre za bolj povezano strukturo. Gostota je mera, ki je podana s številom vseh povezav v omrežju v primerjavi z vsemi možnimi povezavami. Maksimalna gostota vrednosti 1 nastopi takrat, ko je vsaka enota v omrežju povezana z vsemi ostalimi enotami omrežja. Mera gostote z naraščanjem omrežja ponavadi pada, saj z večanjem števila točk, strmo narašča število vseh možnih povezav v omrežju. V primeru, da želimo primerjati vrednost gostote omrežij različnih velikosti, torej nastane problem (de Nooy in drugi 2011, 73–74).

2.4 STOPNJE TOČK

Stopnja točke izhaja iz števila relacij s katerimi je točka v povezavi. V usmerjenih omrežjih ločimo tudi med vhodno in izhodno stopnjo točke, kjer je prva podana s številom povezav, ki v točko vstopajo, druga pa s številom povezav, ki iz točke izstopajo. Ker je gostota omrežja lahko varljiva, saj je odvisna od velikosti omrežja, je dobro preveriti tudi vrednosti stopenj točk. Točke z višjo stopnjo se ponavadi nahajajo v bolj gostih delih omrežja, saj nosijo večje število povezav. Tako je možno izračunati tudi povprečno stopnjo vseh točk, s katero lahko merimo strukturno povezanost omrežja. Tak način merjenja povezanosti, je boljši kot gostota, saj lahko vrednosti primerjamo tudi med omrežji različnih velikosti (de Nooy in drugi 2011, 73–74).

2.5 KOHEZIVNE PODSKUPINE

V omrežjih se pogosto nahajajo neke podskupine enot, ki so med sabo bolj povezane, kot pa z ostalimi enotami v omrežju. Pri tem lahko domnevamo, da povezave med njimi ne pomenijo zgolj preproste interakcije, ampak se enote povezujejo tudi na osnovi nekih skupnih lastnosti in strukturnih karakteristik, ki jih imajo. Predvidevamo namreč lahko, da take enote pogosteje vstopajo v interakcijo, kot pa enote, ki so si različne. Za opazovanje kohezivnih skupin obstaja več različnih tehnik. Tako lahko v nekem omrežju iščemo komponente, jedra in klike. Zadnje je v velikih omrežjih težje iskati, saj način določanja teh vzame veliko časa (de Nooy in drugi 2011, 71–85). V svoji nalogi sem tako iskal zgolj komponente in jedra.

2.5.1 KOMPONENTE OMREŽJA

Čeprav omrežje ni vedno povezano v celoti, lahko odkrivamo dele, kjer se točke med seboj povezujejo. Pri komponentah tako razlikujemo krepko povezane, šibko povezane in dvopovezane komponente. Skupina točk predstavlja krepko povezano komponento takrat, ko ob upoštevanju smeri povezav pridemo iz vsake točke v tej skupini v vsako drugo točko skupine. Če smer povezave med točkami v skupini ni pomembna, potem govorimo o šibko povezani komponenti. Dvopovezana komponenta se pojavi, ko lahko iz take skupine točk odstranimo katerokoli točko, omrežje pa ostane povezano. Točke, ki povzročajo razpad omrežja na več komponent imenujemo presečiščne točke (de Nooy in drugi 2011, 77–162).

2.5.2 JEDRA

S pomočjo jeder odkrivamo podomrežja, ki nastanejo na podlagi minimalne vrednosti stopnje točk v taki skupini. Taka podomrežja se imenujejo k-jedra, kjer k predstavlja minimalno stopnjo točk znotraj jedra. V 2-jedru imajo vse točke znotraj jedra, ki se med seboj povezujejo, stopnjo točke dva ali več. Glede na vhodne in izhodne povezave ločimo vhodna in izhodna jedra ter jedra, pri katerih upoštevamo vhodne in izhodne povezave skupaj (de Nooy in drugi 2011, 81–82).

2.5.3 OTOKI

Otoki predstavljajo kohezivne skupine glede na vrednosti točk ali povezav. Če se vrednosti pojavljajo na točkah govorimo o točkovnih otokih, v primeru vrednosti na povezavah pa o povezavnih otokih. Otok tako predstavlja podomrežje točk povezanih direktno ali posredno, s povezavami, ki imajo večje vrednosti, kot so vrednosti povezav, ki vodijo izven otoka (de Nooy in drugi 2011, 124–125).

2.6 MERE SREDIŠČNOSTI IN POMEMBNOСТИ

Z merami središčnosti in pomembnosti v omrežjih ugotavljamo in pojasnujemo položaje enot v omrežju ter tako poskušamo najti pomembne osebe. Središčnost in usredinjenost predstavljata dva izmed starejših konceptov v analizi socialnih omrežij. Mere središčnosti ali centralnosti se računajo za neusmerjena omrežja, medtem ko pri usmerjenih omrežjih, kjer razlikujemo med izvorom in ciljem vezi, računamo mere pomembnosti (Hlebec in Kogovšek 2006, 11–12).

2.6.1 MERE SREDIŠČNOSTI

Mere središčnosti in centralnosti nam prikažejo enote, ki so v omrežju blizu relacijskega dogajanja. Enote z večjo mero središčnosti, ki zavzemajo bolj centralno pozicijo v omrežju, imajo višjo stopnjo od ostalih točk, prav tako pa se nahajajo na večjem številu najkrajših poti med ostalimi enotami. Tako imajo lahko boljši dostop do informacij. Mere središčnosti se nanašajo na posamezne enote v omrežju. Kadar pa vrednosti računamo za celotno omrežje, govorimo o merah usredinjenosti. Omrežje je zelo usredinjeno, če obstaja jasna meja med središčem in periferijo (de Nooy in drugi 2011, 141).

Mera središčnosti glede na stopnjo neke točke je v neusmerjenem omrežju lahko kar skupna stopnja te točke. V usmerjenih omrežjih lahko to mero uporabimo tudi kot mero pomembnosti, kjer ločimo med vhodnimi in izhodnimi središčnostjo glede na stopnjo. V primeru primerjave vrednosti z ostalimi omrežji jo lahko normaliziramo in dobimo relativno mero središčnosti glede na stopnjo (de Nooy in drugi 2011, 144).

Naslednja je mera središčnosti glede na dostopnost. Ta poleg neposrednih upošteva tudi posredne sosede neke točke in predstavlja boljšo mero kot središčnost glede na stopnjo. Mera torej upošteva tudi razdaljo med izbrano točko in vsemi ostalimi točkami, kjer večja razdalja prinese manjšo vrednost središčnosti glede na dostopnost. Tudi pri usmerjenih omrežjih govorimo o vhodni in izhodni središčnosti glede na dostopnost. Prva mera nam pove kako blizu je neka enota ostalim v omrežju, druga pa kako blizu v omrežju so vse ostale enote izbrani enoti (Wasserman in Faust 1994, 183–188).

Mera središčnosti glede na vmesnost izhaja iz tega, da je neka enota bolj središčna, če v omrežju nastopa kot nekakšen posrednik informacij. Gre za enote, ki se nahajajo na najkrajših poteh med različnimi pari enot in tako zavzemajo strateško pomemben položaj (de Nooy in drugi 2011, 150).

2.6.3 MERE POMEMBNOSTI

V usmerjenih omrežjih, kjer je smer povezave pomembna, lahko računamo mere pomembnosti. Tu do izraza pride razlikovanje med danimi izbirami in prejetimi izbirami. Ljudje, ki velikokrat izbrani s strani ostalih in imajo večjo vhodno mero, so tako bolj pomembni in imajo prestiž. Pri meritvah pomembnosti so tako važne vhodne mere enot (Hlebec in Kogovšek 2006, 11–12).

Obstaja več mer za določanje pomembnosti točke. Prva izmed teh je lahko kar vhodna stopnja točke. Pri merjenju pomembnosti to sicer ni najboljša opcija, saj ima mera glede na stopnjo določene slabosti. Stopnja točke namreč upošteva le neposredne izbire, prav tako pa ne upošteva pomembnosti ostalih enot, ki so jo izbrale (de Nooy in drugi 2011, 217–228).

V našem primeru, ko gre za bolj redko omrežje, je bolj primerna mera območje vpliva, ki predstavlja število vseh drugih točk v usmerjenem omrežju, iz katerih lahko ob upoštevanju smeri povezav dosežemo izbrano točko. Pomaga nam torej določiti točke, ki imajo v omrežju strukturno bolj pomemben položaj. Slabost metode pa je v tem, da ne razlikuje med posrednimi in neposrednimi izbirami. Mera, ki med takimi izbirami razlikuje pa je bližina. Ta domneva, da neposredne nosijo večjo težo, kot posredne. Tako je bližina mera, ki predstavlja delež točk v območju vpliva neke točke glede na povprečno oddaljenost vseh točk v območju vpliva (de Nooy in drugi 2011, 221–228).

Še dve zanimivi meri pomembnosti najdemo v kazalih in opisih. Dobra kazala v omrežju predstavljajo pomembne točke, ki kažejo na pomembne opise. Ravno obratno pa je točka dober opis, če nanjo kaže več dobrih kazal (de Nooy in drugi 2011, 153).

3 SPLETNA SOCIALNA OMREŽJA

Spletna socialna omrežja v zadnjem času poznamo kot popularne spletne aplikacije, ki so namenjene različnim tipom oziroma profilom posameznikov in ponujajo različne tipe storitev. Večinoma so namenjene druženju ljudi s prijatelji, deljenju informacij, ki so relevantne različnim skupinam, in pojavljanju z namenom biti viden in ostati v stiku. Spletna socialna omrežja so tako lahko namenjena zgolj prijateljskemu druženju, navezovanju in iskanju poslovnih stikov ter priložnosti, v nekaterih primerih pa tudi združevanju z namenom političnega udejstvovanja in izražanja. Najpomembnejše lastnosti, ki sestavljajo neko spletno socialno omrežje, so uporabniški profili, sezname prijateljev, orodja namenjena javnemu komentiranju in tekoč, sproten način objavljanja (Boyd 2011, 39–44).

Spletne storitve, ki definirajo neko stran oziroma aplikacijo kot spletno socialno omrežje, so po Boydovi in Ellisonovi (2007, 211) naslednje:

- kreacija javnega ali delno javnega profila v omejenem sistemu;
- oblikovanje seznama uporabnikov s katerimi si posameznik deli povezavo;
- vpogled v lasten seznam omrežja in v sezname narejene s strani drugih uporabnikov.

Večina spletnih socialnih omrežij uporabnikom omogoča, da z vpogledom v prijateljev profil pridobijo tudi vpogled v njegovo listo prijateljev. Tu se pokaže zanimiv aspekt socialnih omrežij, ki ne predstavlja zgolj spoznavanja novih ljudi, temveč uporabnikom omogoča prikaz in deljenje njihovega celotnega omrežja. Kljub temu, da ta storitev omogoča povezovanje med uporabniki, kjer to drugače ne bi bilo mogoče, pa v večini primerov socialnih omrežij primarni cilj ostaja komunikacija z ljudmi, ki predstavljajo del našega razširjenega omrežja. Prav tako večina socialnih omrežij omogoča objavljanje in komentiranje na profil prijateljev, kakor tudi pisanje zasebnih komentarjev. Komentarji ne predstavljajo zgolj pisnih oblik, ampak tudi objavo in deljenje fotografij, video posnetkov in objav iz drugih internetnih strani (Boyd in Ellison 2007, 211–214).

Glede na zgornjo definicijo spletnega socialnega omrežja naj bi se prva taka bolj prepoznavna spletna aplikacija pojavila leta 1997. Gre za SixDegrees.com, ki je omogočal kreacijo profila, oblikovanje seznama prijateljev, leto kasneje pa je bila omogočena tudi storitev pregledovanja seznamov prijateljev. Kljub temu, da so se podobne oblike storitev v posameznih vlogah pojavljale že v preteklosti, pa so v primeru SixDegrees.com prvič nastopile v združeni vlogi. Ne glede na to, da se ta spletna aplikacija ni obdržala, so sledile številne druge, ki so nekako izoblikovale, definirale in popularizirale današnji pogled na spletna socialna omrežja. Med najpomembnejša spadajo Friendster, MySpace in Facebook. Čeprav večina današnjih spletnih aplikacij uporablja podobna tehnološka orodja in funkcije, ki omogočajo njihovo delovanje, pa prihaja predvsem do delitve glede na namen takih spletnih socialnih omrežij. Nekatera združujejo raznolike profile uporabnikov, medtem ko se na drugih povezujejo predvsem ljudje z nekimi skupnimi lastnostmi (Boyd in Ellison 2007, 214–219).

3.1 TWITTER

Twitter je spletno socialno omrežje, ki so ga leta 2006 ustvarili oz. ustanovili Jack Dorsey, Evan Williams, Biz Stone in Noah Glass. Do danes je močno pridobilo na priljubljenosti in je z 271 milijoni mesečno aktivnimi uporabniki eno izmed najpriljubnejših spletnih socialnih omrežij. Dnevno je poslanih približno 500 milijonov »Tweetov« oziroma kratkih sporočil, ki so glavna značilnost Twitterja (Twitter Inc. 2014). Ta namreč temelji na mikrobloggingu, komunikacijski storitvi, ki uporabniku omogoča objavljanje sporočil dolžine do največ 140 znakov (Gilpin 2011, 232).

3.1.1 ZNAČILNOSTI TWITTERJA

Java in drugi (2007, 56–57) ugotavljajo, da uporabniki mikroblogging uporabljajo za različne namene. Predvsem značilno je objavljanje dnevnih aktivnosti in uporaba Twitterja kot pogovornega orodja. Mikroblogging pa se izkaže tudi v primeru objavljanja mnenj, deljenja in iskanja informacij ter poročanja novic. V primerjavi z blogingom, mikroblogging predstavlja hitrejši način komunikacije. Ker je dolžina posameznega tweeta omejena, je s tem zmanjšana zamudnost in vložek časa, ki ga uporabnik pri samem pisanju porabi. S tem pa se poveča tudi frekvenca posameznih tweetov nekega aktivnega uporabnika, ki jih tako lahko objavi tudi več dnevno.

Najbolj prepoznaven znak Twitterja je sigurno »hashtag« oziroma lojtra (#). Znak se v tweetih pojavlja kot označba za ključne besede, teme ali skupine in ima vlogo kategoriziranja sporočil, npr. #volitve2014. Uporablja se lahko več takih označb na posamezen tweet, postavimo pa jih lahko kjerkoli v tweetu. Določene označbe, ki postanejo najbolj popularne, se pojavijo na profilu »Trending Topics«. Ta objavlja teme, o katerih se na Twitterju najbolj govori. Označba »hashtag« je Twitterju v uporabo prišla šele avgusta leta 2007 po predlogu uporabnika Chrisa Messine (Twitter Inc. 2014).

3.1.2 POVEZOVANJE UPORABNIKOV V SPLETNEM SOCIALNEM OMREŽJU TWITTER

Twitter v osnovi deluje na način »sledenja« uporabnikom. Uporabnik sledi drugim osebam, ki lahko predstavljajo njegove prijatelje (friends) ali pa tudi ne. Če uporabnik sledi osebi, ki ni njegov prijatelj, je potemtakem zgolj sledilec (follower). Tako pridemo do zaključka, da so povezave med uporabniki lahko recipročne ali pa le enosmerne (Java in drugi 2007, 56–57; Huberman in drugi 2009, 2–3).

Usmerjene povezave v družbenih omrežjih lahko predstavljajo vse od prijateljstva, skupnih interesov, poti informacij itd. Take povezave torej določajo smer pretoka informacij in so tako ključne v ugotavljanju pomembnih oseb v omrežju, tistih z največjim vplivom ali podporo. V Twitterju je vpliv mogoče meriti v treh stopnjah, preko vhodnih stopenj točke oziroma osebe, z »retweeti« oziroma tweeti deljenimi naprej preko drugih oseb ter z omembami uporabniškega imena v tweetih drugih uporabnikov. (Cha in drugi 2010, 10). Tudi sam v svojem omrežju iščem take ključne osebe in analiziram usmerjeno omrežje. Glede na to da

analiziram zgolj uporabnike v omrežju ne pa tudi njihovih sporočil, sem vpliv in podporo meril z metodami stopenj točk, dostopnostjo, vmesnostjo, območjem vpliva in bližino.

Huberman in drugi (2009, 5–6) v svoji raziskavi, ki so jo opravili na bazi 309.740 uporabnikov, ugotavljajo, da ima kar 98.8% uporabnikov več oseb katerim sledijo (followees), kot pa dejanskih prijateljev. Število prijateljev, ki jih imajo uporabniki, naj bi bilo v primerjavi z številom oseb katerim sledijo, izredno nizko. Tako Twitterjevo omrežje, definirano zgolj s številom sledilcev in tistim, ki jim uporabnik sledi, na prvi pogled deluje zgoščeno. Če pa enako omrežje obravnavamo skozi povezave prijateljev, torej oseb, ki uporabnikom sledijo nazaj, je tako omrežje bolj redko.

Uporabniki Twitterja lahko svoja sporočila pošiljajo na direkten ali indirekten način. Direktna sporočila so namenjena točno določenim uporabnikom, medtem ko so indirektna sporočila namenjena vsem, ki jih želijo prebrati. In čeprav so prva v osnovi namenjena določenim uporabnikom, so prav tako vidna vsem ostalim uporabnikom (Huberman in drugi 2009, 2–3).

V nalogi sem se osredotočil na aktivne uporabnike Twitterja, kar je pomenilo, da so v obdobju enega meseca objavili vsaj en tweet. Ta predpostavka je nujna, saj sem aktivne uporabnike pridobil iz zajema objavljenih tweetov v izbranem obdobju. Za mesečno aktivnost kot pogoj, sem se odločil na podlagi statistike, ki jo navaja Twitter. Tega naj bi mesečno uporabljalo 271 milijonov aktivnih uporabnikov (Twitter Inc. 2014). Prav tako sem za zgled vzel raziskavo v članku *Social networks that matter: Twitter under the microscope*, ki za preverjanje aktivnosti uporablja podoben pogoj (Huberman in drugi 2009).

3.1.3 PODATKOVNO RUDARJENJE PO TWITTERJU

V današnjem času smo preplavljeni s količino informacij in podatkov, ki vseskozi narašča. Svetovni splet je prostor, kjer je ta fenomen najbolj izrazit. Hkrati pa se tudi naše delovanje in dejanja na spletu beležijo in za seboj puščajo bolj ali manj vidne sledi. Z naraščanjem količine podatkov prihaja do tega, da se zmanjšuje del podatkov, ki smo jih ljudje dejansko sposobni razumeti. V vsej tej gmoti podatkov se torej nahajajo informacije, ki so nekako zakrite in jih je potrebno izluščiti in na bolj razumljiv, ekspliciten način tudi prikazati. Govorimo o iskanju vzorcev. Pri tem gre lahko za iskanje vzorcev pri volivcih in njihovem mnenju o politiki, za prepoznavanje tržnih priložnosti, ali iskanje vzorcev v navadah potrošnikov (Witten in drugi 2011, 3–5; Han in drugi 2011, 1–5).

Danes se podatkov ne zbira le ročno, ampak se v veliki meri avtomatsko zbirajo na internetu. V zadnjih letih se tako povečuje število podatkovnih baz. Baze so lahko sestavljene iz različnih vsakdanjih aktivnosti uporabnikov interneta, njihovih potrošniških odločitev ali pa participacije v različnih spletnih socialnih omrežjih. Witten in drugi (2001, 4–5) navajajo, da naj bi se število podatkovnih baz na internetu celo podvojilo vsakih dvajset mesecev. S takim porastom zbirk podatkov se definitivno povečuje tudi število možnosti za podatkovno rudarjenje. To je tako eden izmed glavnih načinov, ki nam pri taki količini podatkov sploh še omogoča iskanje vzorcev, iz katerih lahko izluščimo smiselne informacije in tako dobimo vpogled v navade ljudi, načrtamo tržne strategije ali pridobimo konkurenčno prednost. Pri podatkovnem rudarjenju gre torej za analizo informacij, ki se že nahajajo v nekaterih obstoječih podatkovnih bazah. Lahko ga torej definiramo kot proces odkrivanja vzorcev v podatkih, pri tem pa mora biti proces avtomatski ali polavtomatski. Odkriti vzorci morajo imeti smisel in morajo voditi v neke vrste prednost ali znanje (Witten in drugi 2011, 4–9).

Han in drugi (2011, 5–8) menijo, da bi bilo bolj pravilno poimenovanje »knowledge mining«, saj iz podatkov želimo dobiti uporabne informacije ali znanje. Prav tako razlagata o dveh različnih pomenih podatkovnega rudarjenja. Medtem ko prvi pomeni celoten proces ekstrakcije uporabnih informacij iz večje količine podatkov, je v drugem pomenu podatkovno rudarjenje zgolj bistven korak v procesu odkrivanja znanja, kjer so drugi koraki še čiščenje, integracija, selekcija podatkov itn. Kakorkoli že, podatkovno rudarjenje predstavlja enega izmed pionirskih pristopov v panogi podatkovnih baz in informacijskih sistemov ter pomembno interdisciplinarno razvojno orodje v informacijski tehnologiji.

Twitter podobno kot nekatera druga večja spletna socialna omrežja predstavlja bogat vir podatkov o navadah ljudi. Prednost Twitterja je odprtost in dostopnost do podatkov, ki pa je v zadnjem času vseeno nekoliko omejena (zbiranje podatkov oziroma količina zbranih podatkov je časovno omejena). Prav tako ima Twitter zelo urejen API (Application Programming Interface), veliko dokumentacije za pomoč pri upravljanju in zbiranju podatkov preko le-tega ter bogat nabor orodij in programov s katerimi preko API-ja lahko dostopamo do podatkov (Russell 2013, 5).

Eden izmed načinov, kako priti do nekaterih zbirk podatkov je torej API. Ta omogoča povezavo med različnimi platformami in aplikacijami, s katerimi želimo dostopati do podatkov. API si lahko predstavljamo tudi kot dostopanje do spletne strani, ki namesto urejene spletne strani vrača strukturirane zbirke podatkov. Zaradi takšnega strukturiranega vračanja podatkov, te

lažje razdelimo in iz njih izluščimo uporabne informacije. V primeru Twitterjevega API-ja želene podatke zbiramo z ukazi, ki jih implementiramo v našo aplikacijo. Tako lahko na primer zberemo vse tweete, ki so bili objavljeni v nekem časovnem obdobju in vsebujejo določeno besedo, tako da v naši aplikaciji uporabimo določen ukaz in željene parametre. Twitterjev API pa lahko poleg namena zbiranja podatkov, uporabimo tudi kot orodje v programiranju funkcionalnih aplikacij (Makice 2009, 133–134).

Podatki, pridobljeni s spletnega socialnega omrežja Twitter, so lahko zanimivi z vidika, da uporabniki svoje komentarje in objave pišejo instantno, momentalno in brez nekega posebnega načrtovanja. To pomeni, da so podatki, ki so sami po sebi večplastni, takoj uporabni in predstavljajo najširši vpogled v družbeno dogajanje na lokalni ali globalni ravni (Russell 2013, 5). Po drugi strani pa so tweeti lahko dobro načrtovana sporočila, ki služijo v marketinške, promocijske ali propagandne namene (Huberman in drugi 2009; Weimann 2010).

Pisanje in objava tweetov ter Twitterjeva funkcionalnost, ki uporabniku omogoča sledenje ostalim uporabnikom in obratno, tvorita način, ki ljudi povezuje na več različnih načinov. To povezovanje lahko prikažemo, kot zanimive kratke, smiselne dialoge, kot grafe, ki prikazujejo mreženje med ljudmi samimi, ali v odnosu uporabnikov do določene tematike. Prav lahka dostopnost do Twitterjevih podatkov se torej izkaže kot velika prednost pri podatkovnem rudarjenju, kjer se pokažejo številne možnosti manipulacije s podatki (Russell 2013, 5).

4 VZOREC IN ZBIRANJE PODATKOV

Analiziral sem omrežje, ki sem ga pridobil z izdelavo lastnega orodja v programu Python in s pomočjo Pythonu prirejene knjižnice Tweepy, ki omogoča komuniciranje programa s spletnim omrežjem Twitter. Spletno omrežje Twitter sem izbral, ker njegov vmesnik omogoča dostop do podatkov s pomočjo različnih programov. Postopek je sicer v primerjavi s preteklostjo sedaj nekoliko otežen, saj je zbiranje podatkov omejeno na klice v API, ki časovno dovoljujejo le določeno količino prenosa podatkov.

4.1 PYTHON

Python je priljubljen odprtokodni programski jezik, ki ga lahko uporabljamo za pisanje samostojnih programov ali kot skriptni jezik na širokem spektru področij. Njegove odlike so

zmogljivost, prenosnost in dejstvo, da je zastoj, hkrati pa se ponaša z berljivostjo kode. Je interpretativen, objektno orientiran programski jezik (Python 2014).

V današnjem času je na voljo veliko število programskih jezikov, ki imajo različne funkcionalnosti ter prednosti in slabosti. V našem primeru ima Python prednost v tem, da je programska skripta izvedena takoj in je tako lažje odpravljati napake, predvsem pa je prednost v naboru različnih dodatnih knjižnic, ki poleg standardne funkcionalnosti in namembnosti programskega jezika omogočajo programiranje aplikacij za najrazličnejša področja (Lutz 2013, 3–5).

4.2 TWEETPY IN TWITTER API

Ena izmed mnogih dodatnih razširitev, ki so na voljo programskemu jeziku Python, je tudi Tweepy. Ta poleg Twython in nekaterih drugih podobnih knjižnic omogoča poenostavljeno komunikacijo s Twitterjevim preko njegovega API-ja in tako predstavlja osnovo za našo aplikacijo zbiranja podatkov. Izbral sem jo zaradi dejstva, da je po spremembi API-ja ostala ena izmed redkih knjižnic, ki je bila še naprej nadgrajevana in delujoča.

Tweepy je torej odprtokodna knjižnica za komunikacijo s Twitterjem, ki je sprva uporabljala dostop preko »Basic Authentication«, sedaj ko ta metoda s strani Twitterja ni več podprta, pa dostop preko »Oauth« metode. Glavna razlika je v tem, da je preteklosti bil mogoč prost dostop do podatkov zgolj s Twitterjevim uporabniškim imenom in geslom, po spremembi pa je potrebno kreiranje aplikacij. Novejša metoda ima nekatere prednosti, ki pomenijo predvsem večjo varnost. Čeprav sprva na voljo ni bilo veliko primerov kode in dokumentacija knjižnice ni bila najbolj nazorna, je s pomočjo aktivne skupnosti Tweepy danes eden izmed bolj priljubljenih in uporabnih Pythonovih dodatkov, ki omogočajo komunikacijo s Twitterjem in gradnjo uporabnih aplikacij (Novalić 2013).

4.3 POTEK IN KORAKI V PROCESU ZBIRANJA PODATKOV

V prvem delu procesa pridobivanja podatkov in oblikovanja končnega omrežja sem sestavil osnovno bazo aktivnih slovenskih uporabnikov Twitterja. Ker se dejanskih oseb s socialnega omrežja Twitter ne da pridobivati direktno oziroma so klici zelo omejeni, sem moral sprogramirati prirejeno orodje. To je v osnovi delovalo tako, da sem zbiral tweete, ki so bili objavljeni na izbranem področju. Področje sem zasnoval s pomočjo geokodiranja, ki v spletnem omrežju Twitter omogoča geolokacijski prikaz izvora objavljenega tweeta. Tako

sem zbiral vse objavljene tweete na območju, ki ima center določen s koordinatami zemljepisne dolžine 46.05 in zemljepisne širine 14.5 (Ljubljana), ter v radiju 60 kilometrov. Tweete sem nato zbiral en mesec med 12. novembrom in 12. decembrom. Treba pa je poudariti, da se objavljene tweete lahko zbira le za 7 dni nazaj, tako da je bilo postopek potrebno nekajkrat ponoviti. S tem postopkom sem torej pridobil aktivne slovenske uporabnike Twitterja na izbranem območju.

Spletno omrežje Twitter načeloma omogoča tudi zapis lokacije uporabnika, kar pa pri sami registraciji ni zahtevano. Uporabniki tako pogosto tega podatka ne prikažejo ali pa v polje zapišejo izmišljeno lokacijo. Prav zaradi tega je bilo zbiranje uporabnikov oteženo in je bil potreben alternativen način zbiranja uporabnikov s pomočjo parametra geolokacije. Ta način ima slabost v tem, da vsi tweeti ne vsebujejo podatka o geolokaciji. Podatek je dodan le tweetom uporabnikov, ki to funkcijo omogočijo. V tem primeru se v podatek poslanega tweeta zabeleži, če je ta poslan iz mobilne naprave ali v primeru, da ima uporabnik v svojih podatkih navedeno lokacijo. V slednjem primeru se zabeleži približna lokacija nekega kraja (Twitter Inc. 2014). Po nekaterih testih in preverjanjih naj bi bilo geolociranih okoli 35–50 odstotkov slovenskih tweetov (Martinc 2013, 12).

V drugem delu procesa oblikovanja omrežja je bilo potrebno oblikovati orodje, ki omogoča pridobivanje prijateljev oziroma tistih, ki jim uporabnik sledi »following« ter »followerjev« ali tistih, ki sledijo uporabniku. Ta proces je pomemben, saj na ta način lahko oblikujemo končno želeno omrežje, ki predstavlja aktivne uporabnike Twitterja na območju Slovenije. Twitter namreč omogoča, da uporabnik sledi drugim uporabnikom, ki niso v Sloveniji, prav tako pa uporabniku lahko sledijo ostali uporabniki, ki so v tujini. Izdelati je bilo torej potrebno nekakšno podomrežje, ki predstavlja samo uporabnike, ki spadajo v naš vzorčni okvir.

Tako pridemo do zadnjega dela procesa oblikovanja zaprtega omrežja aktivnih slovenskih uporabnikov Twitterja. Od tu naprej sta bili pomembni predvsem osnovna baza uporabnikov, ki sem jih dobil prek tweetov, in pa baza »following«, torej oseb, ki jim osnovni uporabniki sledijo.

V prvem koraku je bilo bazi potrebno preveriti in izločiti razne anomalije. Tako sem odstranil zanke in podvojena sledenja, ki jih v bazi načeloma sploh ne bi smelo biti. O zankah govorimo v primeru, ko oseba sledi sama sebi. V primeru omrežja Twitter se to lahko zgodi le, če uporabnik sam z lastno spisano aplikacijo to omogoči. Prav tako pa ne bi smelo prihajati do dvojnih sledenj, saj eni osebi na Twitterju ne moremo slediti dvakrat. Ta napaka se pojavi

v redkih primerih, ko gre sicer za unikatna uporabniška imena, ki pa imajo isto identifikacijsko kodo. Do tega lahko pride, ko uporabnik ustvari in poveže več Twitter računov.

V drugem koraku sem spisal algoritem, ki glede na osnovno bazo iz baze »following« izloči osebe, ki se v osnovni bazi ne nahajajo, saj je nemogoče ugotoviti ali bi lahko bili tudi del našega izbranega omrežja uporabnikov Twitterja v Sloveniji. Kot že prej omenjeno, lahko uporabnik Twitterja sledi komurkoli in seveda obstaja možnost, da ima več uporabnikov osnovne baze za prijatelja isto osebo, ki pa ni nujno del našega izbranega zaprtega omrežja. V teoriji bi bilo mogoče pregledati tudi take osebe, vendar bi bilo potrebno ponovno ponoviti postopek pridobivanja njihovih sledilcev in pa oseb, ki jim sledijo sami. Ta postopek bi prav tako lahko pomenil premik iz našega izbranega omrežja, predvsem pa bi bil zelo dolgotrajen, saj vsaka oseba lahko sledi tudi milijon ostalim. Obstaja torej način s katerim bi lahko zbrali bolj popolno omrežje slovenskih uporabnikov Twitterja, s tem da bi poskusil dodati osebe, ki jih s pomočjo geokodiranja nisem zajel, vendar bi pri tem lahko zajeli tudi take, ki niso nujno aktivni uporabniki.

S tem procesom sem ustvaril končno izbrano omrežje. Gre za aktivne uporabnike spletnega omrežja Twitter v Sloveniji, ki so tweetali vsaj enkrat v obdobju med 12. novembrom in 12. decembrom, iz prej določenega območja zemljepisne dolžine 46.05 in zemljepisne širine 14.5 (Ljubljana), ter v radiju 60 kilometrov.

Postopek zbiranja in čiščenja podatkov je vzel kar nekaj časa, saj je Twitter onemogočil prej manj omejen dostop do svojih strežnikov s podatki. Sedaj je Twitterjev vmesnik, ki omogoča zbiranje podatkov omejen s številom klicev na strežnike vezanih tudi na določeno časovno obdobje. Tako je bilo v orodje potrebno integrirati tudi časovne premore, ki so omogočili nemoteno zbiranje, vendar pa se je postopek kar občutno podaljšal.

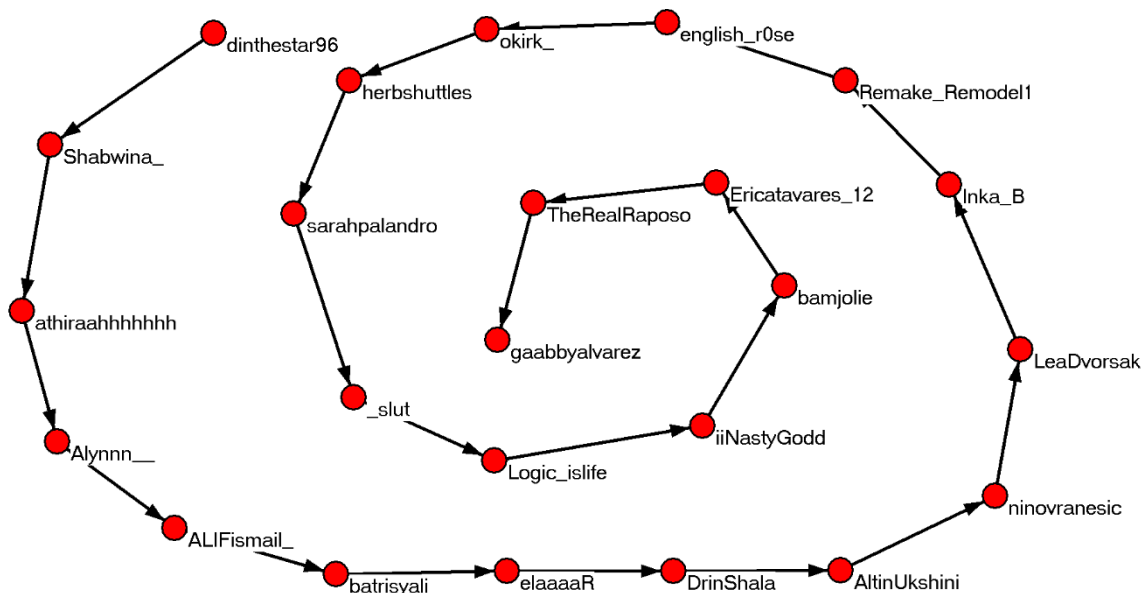
5 ANALIZA OMREŽJA

5.1 SESTAVA IN ZNAČILNOSTI OMREŽJA

Končno omrežje slovenskih uporabnikov Twitterja, ki sem ga uspel generirati, je enovrstno usmerjeno omrežje. Sestavlja ga 13807 enot oziroma slovenskih uporabnikov Twitterja. Gre torej za veliko socialno omrežje z kar 552389 usmerjenimi povezavami. V omrežju ne najdemo zank, saj osebe na Twitterju ne morejo slediti samim sebi, razen v redkih izjemah, kjer je za to potrebno spisati lastno aplikacijo. Prav tako neka oseba drugi ne more slediti večkrat. Gre za redko omrežje, saj gostota omrežja brez zank dosega vrednost približno 0.0029. Povprečna stopnja oziroma povprečno število oseb, ki jim posamezen uporabnik sledi, je 80.016.

Točki, ki sta v našem omrežju najbolj oddaljeni, sta točki »*dinthestar96*« in »*gaabbyalvarez*«. Razdalja med njima torej predstavlja najdaljšo najkrajšo pot in meri 23. Ta razdalja predstavlja tudi premer našega celotnega omrežja. Od ene do druge točke v triindvajsetih korakih pridemo na naslednji način, prikazan v spodnji sliki (glej Sliko 5.1).

Slika 5.1: Najdaljša najkrajša pot v omrežju



5.2 STOPNJE TOČK OMREŽJA

Vhodne in izhodne vrednosti točk so lahko prvi pokazatelj pomembnosti določenih enot v omrežju, saj nam povejo kako priljubljena je oseba s podatkom, koliko ljudi ji sledi, ter kako aktivna je oseba v smislu sledenja ostalim enotam v omrežju. Treba je ponovno poudariti, da se vrednosti nanašajo samo na naše generirano omrežje, kjer so izključene enote, ki v to omrežje ne spadajo.

Najvišja vrednost vhodne stopnje točk pripada enoti »Val202« in znaša 2165. Tolikšno je torej število drugih enot, ki sledijo uporabniku »Val202«. V omrežju je kar 2410 takih, ki imajo vhodno stopnjo točke z vrednostjo 0. To pomeni, da jim nihče v našem omrežju ne sledi.

Najvišjo izhodno stopnjo točke ima uporabnik »vladaRS«, z vrednostjo 1612. To pomeni, da uporabnik sledi kar 1612 osebam v našem omrežju. Po drugi strani kar 1934 takih enot, ki imajo izhodno stopnjo točke z vrednostjo 0 oziroma ne sledijo nobeni enoti v našem omrežju. V omrežju pa najdemo tudi 425 enot, ki sicer nimajo izhodnih povezav, imajo pa vhodne povezave in jim rečemo sprejemniki.

Najvišjo vrednost skupne stopnje točk ima prav tako enota »Val202« z vrednostjo 3385, ki ji sledi prej omenjena enota »vladaRS« z vrednostjo 3308. Kar 1509 enot ima skupno stopnjo točk z vrednostjo 0. Te enote torej v našem omrežju ne sledijo nikomur, prav tako pa jim ne sledi nihče iz našega omrežja. Seveda ni rečeno, da je tako v celotnem omrežju Twitterja, saj lahko sledijo drugim osebam, ki se lahko nahajajo izven našega začrtanega območja, prav tako pa tem enotam lahko sledijo druge osebe, ki niso v našem omrežju.

Tabela 5.1: Stopnja točk desetih najbolj aktivnih enot (razvrščeno po skupni stopnji točk)

Enota	Vhodne stopnje točk	Izhodne stopnje točk	Skupne stopnje točk
Val202	2165	1220	3385
vladaRS	1696	1612	3308
rtvslo	1873	839	2712
praprotnix	1495	802	2297
SpletnaMladina	1877	321	2198
RTV_Slovenija	1574	550	2124
toplovodar	1483	569	2052
TinaMaze	1937	0	1937
Dnevnik_si	1694	210	1904
Delo	1778	124	1902

5.3 KOHEZIVNE PODSKUPINE OMREŽJA

5.3.1 KOMPONENTE

V našem omrežju se nahaja 9 krepkih komponent, ki imajo vsaj 20 enot ali več. Največja komponenta ima kar 9625 točk in predstavlja skoraj 70 odstotkov vseh enot v omrežju, naslednja ima le še 239 točk oziroma slaba dva odstotka vseh enot v omrežju.

Obstaja torej 9 takih skupin točk, velikosti 20 ali več enot, pri katerih se, upoštevajoč smer povezave, lahko vsaj posredno pride iz vsake točke v drugo točko skupine. Z drugimi besedami lahko rečemo, da vsak uporabnik v eni izmed teh devetih skupin, sledi ostalim uporabnikom v svoji komponenti na vsaj posreden način.

Omrežje sestavlja 8 šibkih komponent, ki imajo 20 enot ali več. Največja šibka komponenta v omrežju je velikosti 11211 enot in zajema kar 81.2 odstotka vseh enot v omrežju. Pri šibkih komponentah pa smer povezave ne igra vloge, kar pomeni, da omrežje obravnavamo kot neusmerjeno. Tu je torej važno, da iz vsake točke neke komponente, vodi vsaj ena pot v vsako drugo točko te komponente, ne glede na smer povezave.

5.3.2 JEDRA

Pri generiranju jeder glede na vstopne povezave v točke ima najvišjo stopnjo 79-jedro, ki ga sestavlja 737 uporabnikov. To pomeni, da je v tem podomrežju 737 uporabnikov, med katerimi jim sledi vsaj 79 izmed teh 737 oseb.

Jedro najvišje stopnje glede na izhodne povezave iz točk je 86-jedro s 734 točkami. V tem podomrežju se torej nahaja 734 uporabnikov, ki sledijo vsaj 86 uporabnikom iz tega podomrežja.

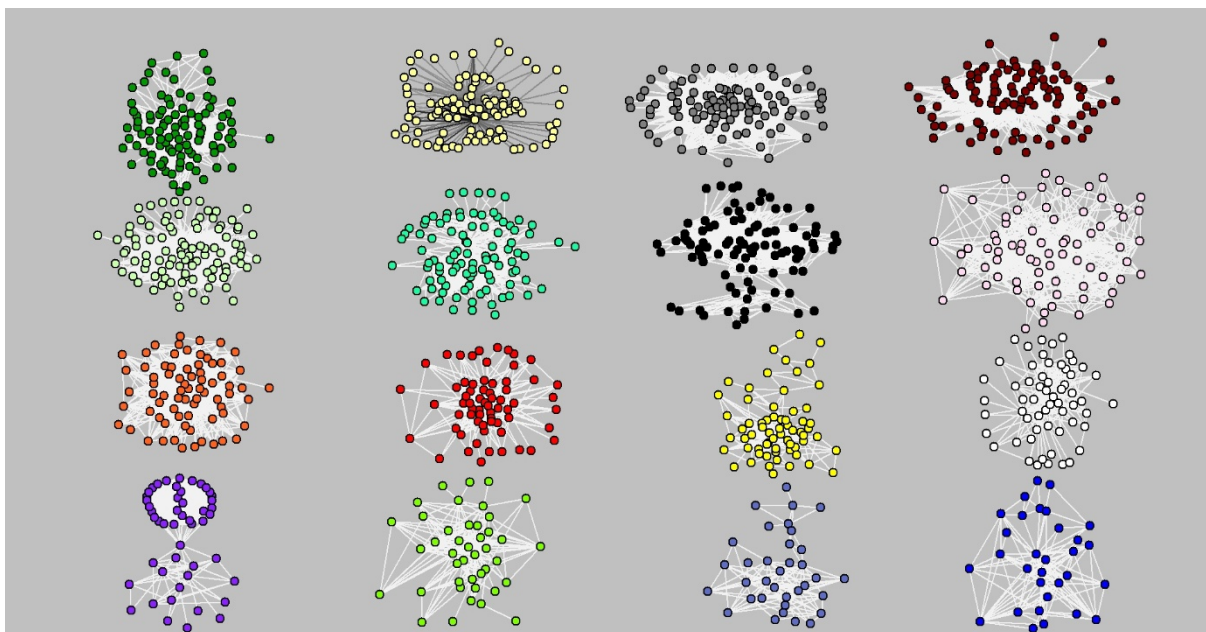
V primeru generiranja najvišje stopnje ne glede na smer povezave pa dobimo 220-jedro, kjer se nahaja 597 enot. Imamo torej podomrežje 597 uporabnikov, ki imajo vsaj 220 povezav.

5.3.3 OTOKI

Generiral sem nekaj primerov otokov z vrednostmi na povezavah. Ker je moje osnovno omrežje brez vrednosti na relacijah, je bilo te potrebno izračunati. To sem izvedel s postopkom štetja pojavitev določene relacije kot tranzitivne bližnice. Na ta način sem pridobil omrežje z vrednostmi na povezavah, kar je omogočilo izračun in prikaz otokov. V omrežju se torej pojavlja 16 otokov velikosti od 30 do 100 enot (Slika 5.2). Te enote znotraj

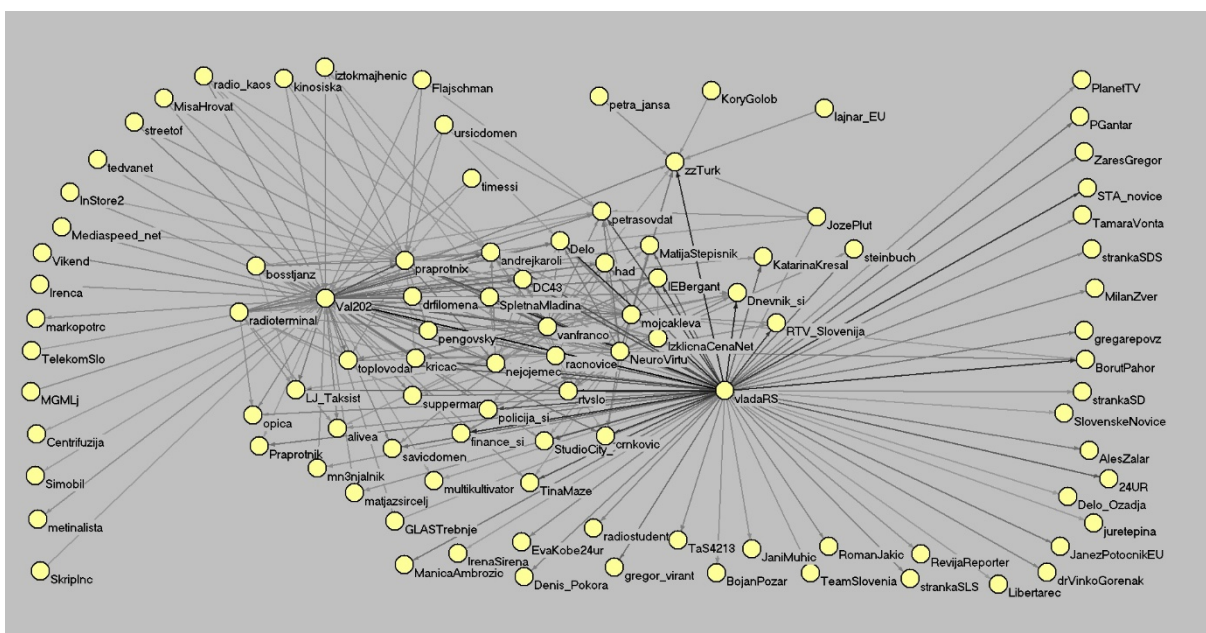
skupin so med seboj povezane s povezavami, katere imajo večje vrednosti, kot povezave, ki vodijo ven iz točk otoka.

Slika 5.2: Prikaz otokov velikosti 30 do 100 enot (Prikaz s sivinami)



Eden izmed teh otokov je še posebej zanimiv, saj prikazuje povezovanje nekaterih bolj znanih slovenskih imen. V otoku se nahajajo politiki, športniki, nekateri novinarji, medijske hiše, radijske postaje, telekomunikacijska podjetja in še nekatera poznana imena (Slika 5.3).

Slika 5.3: Primer zanimivega otoka v omrežju (Prikaz s sivinami)



5.4 MERE POMEMBNOСТИ IN USREDINJENOST OMREŽJA

Tabela 5.2: Prikaz mer pomembnosti za najpomembnejše enote

Enote	Vhodna stopnja	Izhodna stopnja	Vhodna dostopnost	Izhodna dostopnost	Vmesnost
Val202	0.15681588	0.08836738	0.233562	0.21626	0.018549
vladaRS	0.12284514	0.11676083	0.227085	0.225985	0.032241
TinaMaze	0.14030132	0	0.246182	0	0
Remake_Remodel1	0.00065189	0.00043459	0.177347	0.179909	0.035297

Vhodne stopnje pri meri pomembnosti govorijo o podpori oziroma pomenijo biti izbran. Največkrat izbrana enota je »Val202«, katere vrednost vhodne stopnje je približno 0.16 in zanjo lahko rečemo, da ima najvišjo podporo. Sledi ji torej največ ostalih oseb v našem omrežju. V primeru, da je enota izhodišče povezav, računamo kolikokrat posameznik izbere druge člane omrežja. Enota, ki sledi največ ostalim v omrežju je »vladaRS« z vrednostjo izhodne stopnje približno 0.12.

Mera središčnosti glede na dostopnost govori o enotah, ki lahko komunicirajo z ostalimi. Prednost teh vrednosti v primerjavi s središčnostjo glede na stopnjo je, da upošteva tudi posredne sosede. (de Nooy in drugi 2011, 141–152).

Največjo vrednost vhodne dostopnosti (≈ 0.25) ima oseba »TinaMaze«, ki je tako najbližje ostalim osebam v našem omrežju. Na to osebo torej kaže največ povezav, vključno s posrednimi. Od vseh ostalih točk v omrežju jo dosežemo v najmanj korakih.

Vrednost mere vmesnosti je največja pri enoti »Remake_Remodel1« in sicer približno 0.035. Ta enota se torej največkrat nahaja na najkrajši poti med ostalimi pari enot, kar je prav tako pomembna mera, kakor oddaljenost enot od ostalih. Enote z največjo mero vmesnosti so tiste, ki imajo največji nadzor nad pretokom informacij med pari enot znotraj omrežja.

Tabela 5.3: Mere usredinjenosti omrežja

	Vhodna stopnja	Izhodna stopnja	Vhodna dostopnost	Izhodna dostopnost	Vmesnost
Usredinjenost omrežja	0.15392916	0.11387121	/	/	0.018549

Najvišja vrednost usredinjenosti se v našem omrežju pojavlja pri usredinjenosti glede na vhodne stopnje, 0.154. Pravzaprav meri usredinjenosti glede na vhodno in izhodno stopnjo dosežeta precej višjo vrednost kot vmesnost, 0.019. Glede na nizke vrednosti tako

usredinjenosti glede na vhodno in izhodno stopnjo, ter usredinjenosti glede na vmesnost, lahko rečemo da gre za precej decentralizirano omrežje. Vrednosti usredinjenosti glede na vhodno in izhodno dostopnost ni mogoče izračunati, ker omrežje ni močno povezano.

5.5 MERE OBMOČJA VPLIVA IN BLIŽINE (PRESTIŽA)

Kot že prej omenjeno, je stopnja točke omejena mera pomembnosti (prestiza), saj upošteva samo neposredne izbire, prav tako pa je potrebno upoštevati, od katerih ljudi je neka oseba izbrana ali so te osebe same pogosto ali redko izbrane. Upošteva se torej lokalna struktura omrežja, ne pa tudi globalna. Mera, ki pride v poštev je območje vpliva, ki predstavlja število vseh drugih točk v usmerjenem omrežju, iz katerih lahko ob upoštevanju smeri povezav dosežemo izbrano točko. Ta mera pride v poštev v omrežjih, ki niso krepko povezana (de Nooy in drugi 2011, 221–228).

V našem primeru ima največje območje vpliva točka »*TinaMaze*«, njeno območje vpliva zajema 10523 uporabnikov. To pomeni, da lahko iz seznamov sledenja teh 10523 točk posredno ali neposredno pridemo do enote »*TinaMaze*«.

Slabost mere območja vpliva je, da ne razlikuje med posrednimi in neposrednimi povezavami. Neposredne povezave imajo večjo težo in tako je smiselno izračunati tudi vrednosti bližine. Največjo vrednost bližine ima prav tako točka »*TinaMaze*« (≈ 0.25), ki ima poleg največje vrednosti območja vpliva tudi največjo vrednost vhodne dostopnosti. Največja vrednost bližine pomeni, da je ta oseba najbolj pogosto neposredna izbira drugih uporabnikov v našem omrežju.

Kazala so točke, ki kažejo na eno ali več drugih enot v omrežju, opisi pa so točke, ki nekaj opisujejo. Dobra kazala so točke, ki kažejo na veliko dobrih opisov. Prav tako pa je točka dober opis, če kaže na veliko dobrih kazal. Najvišjo vrednost mere kazala dosega »*Val202*« z vrednostjo približno 0.075, hkrati pa dosega tudi največjo vrednost mere opisa, približno 0.115. Gre torej za točko, ki je dobro kazalo in opis, kar pomeni da nima zgolj velike vrednosti vhodnih in izhodnih stopenj, ampak sledi pomembnim uporabnikom, prav tako pa pomembni uporabniki sledijo tej enoti. Med dobrimi opisi je zanimivo videti tolikšno število medijskih entitet.

Tabela 5.4: Vrednosti najboljših kazal in opisev

Kazala		Opisi	
Enota	Vrednosti	Enota	Vrednosti
Val202	0.074709	Val202	0.114686
vladaRS	0.069848	SpletnaMladina	0.102821
vanfranco	0.069367	Delo	0.100647
bosstjanz	0.066588	rtvslo	0.098575
InStore2	0.065924	praprotnix	0.097809
mojcakleva	0.065496	Dnevnik_si	0.096476
ursicdomen	0.065237	vladaRS	0.095214
Flajschman	0.063562	crnkovic	0.094769
praprotnix	0.06333	IEBergant	0.093632
timessi	0.062884	petrasovdat	0.092553

5.6 PRIMERJAVA PODOBNOSTI MER POMEMBNOSTI

Tabela 5.5: Primerjava mer pomembnosti

	Vhodna stopnja	Vhodna dostopnost	Območje vpliva	Bližina	Vmesnost
Vhodna stopnja	1	0.344 ^{**}	0.182 ^{**}	0.344 ^{**}	0.248 ^{**}
Vhodna dostopnost	0.344 ^{**}	1	0.907 ^{**}	1.000 ^{**}	0.112 ^{**}
Območje vpliva	0.182 ^{**}	0.907 ^{**}	1	0.907 ^{**}	0.096 ^{**}
Bližina	0.344 ^{**}	1,000 ^{**}	0.907 ^{**}	1	0.112 ^{**}
Vmesnost	0.248 ^{**}	0.112 ^{**}	0.096 ^{**}	0.112 ^{**}	1

^{**}. Korelacija je signifikantna

Za ugotavljanje podobnosti med posameznimi merami pomembnosti, sem izračunal Pearsonov koeficient korelacije za vhodno stopnjo, vhodno dostopnost, območje vpliva, bližino in vmesnost. Pričakovana je korelacija med merami območja vpliva in bližino (0.907), saj drugo dobimo z delitvijo deleža točk v območju vpliva in povprečno oddaljenostjo izbrane točke od vseh drugih točk v temu območju. Še močnejša je korelacija med vhodno dostopnostjo in bližino (1), kar je ponavadi značilno za krepko povezana omrežja. Med ostalimi merami pomembnosti so koeficienti korelacij relativno nizki.

6 ZAKLJUČEK

V svoji diplomski nalogi sem se osredotočil na dva cilja. Prvi namen naloge je bila analiza aktivnih slovenskih uporabnikov Twitterja, z namenom prikaza ključnih točk in nekaterih podskupin v omrežju. Analiza je podala osebo, ki ji v našem omrežju sledi največ uporabnikov, ter osebo, ki sledi največjemu številu uporabnikov v omrežju. Z dodatnimi analizami središčnosti in pomembnosti sem odkril tudi osebe z največjo vplivnostjo, podporo, območjem vpliva in bližino. Drugi cilj pa je bila kreacija aplikacije v programskem jeziku Python, s katero sem tako omrežje sploh lahko zajel in oblikoval. Pri tem sem predstavil nekatere korake in brezplačna orodja s katerimi tako aplikacijo lahko izdelamo, ter nekatere ovire, ki se v procesu pojavljajo. Čeprav gre v diplomskem delu predvsem za osnovne analize, prihajam do zaključka, da ima analiza spletnega socialnega omrežja Twitter potencial tudi za bolj zahtevne in marketinško privlačnejše analize. Sedaj pa odgovor na prvo raziskovalno vprašanje:

Katere od enot v omrežju imajo največji vpliv in katere največjo podporo?

Analiza omrežja je pokazala naslednje rezultate. Pri pregledu vhodnih in izhodnih stopenj, sem ugotovil, da je oseba z največjim številom sledilcev enota »Val202«, ki ima tudi največje skupno število povezav (izhodnih in vhodnih). Največjemu številu oseb v omrežju pa sledi uporabnik »vladaRS«.

Absolutne stopnje točk vodijo do prvih, najbolj preprostih mer središčnosti in pomembnosti, in sicer relativnih mer središčnosti glede na stopnjo. Tako relativne mere središčnosti povejo, da sta najbolj aktivni enoti »Val202« in »vladaRS«. Za prvo lahko rečemo, da ima z največjo mero pomembnosti glede na vhodno stopnjo največjo podporo v omrežju. Točka »vladaRS« pa ima največjo mero pomembnost glede na izhodno stopnjo in tako največjo vplivnost.

Boljša mera središčnosti, ki ne upošteva zgolj neposrednih povezav, ampak tudi posredne, je mera središčnosti glede na dostopnost. Tu ima največjo mero pomembnosti glede na vhodno dostopnost enota »TinaMaze«, kar pomeni, da iz vseh preostalih točk v omrežju, to točko dosežemo najhitreje. Največjo mero pomembnosti glede na izhodno dostopnost pa ima točka »vladaRS«, kar pomeni, da iz te točke ostale točke v omrežju dosežemo v najmanj korakih.

Pomembnost enote glede na vmesnost je najvišja pri točki »Remake_Remodell«. Ta enota torej najpogosteje leži na najkrajših poteh med drugimi pari enot in ima tako nadzor nad pretokom informacij.

V našem primeru je omrežje dokaj redko, zato je bolj relevantna mera pomembnosti tudi območje vpliva, ki kakor kaže naša tabela primerjave mer pomembnosti, deluje dokaj podobno kot mera dostopnosti. Kot točka z največjim območjem vpliva se tu zopet pojavi točka »TinaMaze«. Ker pa ta mera ne loči med posrednimi in neposrednimi izbirami točke, sem izračunal še bližino, kjer v izračun dodamo še oddaljenost od točk. Tudi mera bližine pa je največja pri enoti »TinaMaze«, ki je torej največkrat neposredna izbira ostalih uporabnikov v omrežju.

Poiskal sem tudi točke, ki predstavljajo dobra kazala in opise. Tu prevladuje enota »Val202«, ki je dobro kazalo in dober opis. Med desetimi enotami z največjo vrednostjo opisov, se znajdejo kar tri imena tiskanih medijev, »rtvslo« in »IEBergant«. Med dobrimi opisi, na katere kažejo pomembne točke, torej prevladujejo medijska imena.

Kljub ogromnemu potencialu, ki ga ima Twitter kot vir podatkov, pa pri analizi tega prihaja do ovir, ki se nanašajo predvsem na razumevanje populacije njegovih uporabnikov. Na podlagi analiz omrežij uporabnikov Twitterja in analize tweetov, ki jih pošiljajo, je bilo narejenih že več različnih raziskav, od napovedi rezultatov volitev, borznih predikcij in napovedovanja uspešnosti filmov. Problem teh podatkov se nahaja v tem, da jih je težko aplicirati na različne demografske skupine. Twitter in tudi druga spletna socialna omrežja pri registraciji pogosto ne zahtevajo in ne preverjajo nekaterih vnesenih osebnih podatkov ali pa jih naknadno celo zavarujejo. Tako so na primer spol, starost, rasa in geografska porazdelitev informacije, ki jih lahko pridobimo le skozi nekatere bolj zapletene algoritme, ali pa sploh ne. Pri izvajanju takih algoritmov seveda lahko prihaja do manjkajočih podatkov ali pa do podatkov, ki niso nujno pravi. To seveda predstavlja mersko napako. Raziskave, ki delajo in bodo delale v to smer, sčasoma lahko prinesejo nekatere izboljšave, ki bodo ta problem omilile (Mislove in drugi 2011, 554–557). Drugo raziskovalno vprašanje pa se nanaša na potencialna podomrežja v našem omrežju.

Ali se v omrežju pojavljajo kakšne podskupine?

Zaradi zgoraj navedenih argumentov v svojem omrežju enot ne ločim glede na spol ali katere druge demografske kategorije. Posledično so analizirana podomrežja le nekatere kohezivne podskupine, kot so komponente, jedra in otoki.

V omrežju najdemo 9 krepkih komponent, ki vsebujejo vsaj 20 enot. Obstaja torej devet podskupin s frekvenco 20 enot ali več, kjer upoštevajoč smer povezave, pridemo iz vsake točke v drugo točko te skupine. Največjo krepko komponento sestavlja 9625 enot oziroma 70

odstotkov vseh enot našega omrežja. Omrežje vsebuje 8 šibkih komponent z vsaj dvajsetimi enotami. Toliko je podskupin velikosti 20 enot ali več, kjer ne glede na smer povezave iz vsake točke lahko dosežemo drugo točko neke podskupine. Največja zajema 11211 enot oziroma 82.2 odstotka vseh enot omrežja. Pojavlja se tud 99 dvopovezanih enot velikosti vsaj 3 enote. Največja izmed njih vsebuje celo 9504 enot.

Glede na vstopne povezave v točke je največje 79-jedro s 737 uporabniki. V tem podomrežju je torej 737 uporabnikov, ki jim sledi vsaj 79 izmed teh. Največje jedro glede na izhodne povezave iz točk je 86-jedro s 734 uporabniki. V tem podomrežju je 734 uporabnikov, ki sledijo vsaj 86 v tej podskupini.

Generirali sem tudi otoke velikosti 30 do 100 enot. V omrežju se jih pojavi 16. Podrobneje prikazan otok iz te skupine vsebuje nekatere bolj poznane osebnosti in podjetja na našem prostoru.

Za konec lahko navedem še nekatere potencialne izboljšave, ki jih je prikazala ta naloga. Sam način zajema slovenski uporabnikov Twitterja bi lahko poleg uporabe geokodiranja, dopolnil še z izpisom lokacij, ki bi jih preverjal s pomočjo aplikacije Google Maps. Prav tako bi način zbiranja tweetov lahko spremenil tako, da bi zbral vse objavljene tweete v izbranem obdobju in prikazal frekvence glede na obdobje ali uporabnika ter še kake druge zanimive podatke. Kot se nakazuje se že pojavljajo tudi nekateri algoritmi in načini, s katerimi bi lahko pridobil nekatere demografske podatke uporabnikov. Glede na to, da je zbiranje podatkov dolgotrajen proces, je torej pomembna dobra predhodna zasnova in načrt zbiranja. Analiza socialnega omrežja Twitter in podatkovno rudarjenje namreč imata potencial za prikaz nekaterih zanimivih podatkov o navadah, mreženju in razmišljanju ljudi.

7 LITERATURA

1. Batagelj, Vladimir in Andrej Mrvar. 1998. Pajek: A Program for Large Network Analysis. *Connections* 21 (2). Dostopno prek: http://www.insna.org/PDF/Connections/v21/1998_I-2.pdf (20. avgust 2014).
2. Boyd, Danah. 2011. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. V *A Networked Self: identity, community and culture on social network sites*, ur. Zizi Papacharissi, 39–58. New York: Routledge.
3. --- in Nicole B. Ellison. 2007. Social Network Sites: Definition, History and Scholarship. *Journal of Computer-Mediated Communication* 13 (1). Dostopno prek: <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/pdf> (27. julij2014).
4. Cha, Meeyoung, Hamed Haddadi, Fabrício Benevenuto in Krishna P. Gummadi. 2010. *Measuring User Influence in Twitter: The Million Follower Fallacy*. Dostopno prek: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1538/1826> (25. avgust 2014).
5. Chu, Kar-Hai, Heather Wipfli in Thomas W. Valente. 2013. Using Visualizations to Explore Network Dynamics. *Journal of Social Structure* 14 (4). Dostopno prek: <http://www.cmu.edu/joss/content/articles/volume14/ChuWipfliValente.pdf> (16. avgust 2014).
6. de Nooy, Wouter, Andrej Mrvar in Vladimir Batagelj. 2011. *Exploratory Social Network Analysis with Pajek, Revised and Expanded Second Edition*. New York: Cambridge University Press.
7. Dráždilová, Pavla, Gamila Obadi, Kateřina Slaninová, Jan Martinovič in Václav Snášel. 2010. Analysis and Visualization of Relations in eLearning. V *Computational Social Network Analysis: Trends, Tools and Research Advances (Computer Communications and Networks)*, ur. Ajith Abraham, Aboul-Ella Hassanien in Václav Snášel, 291–318. London: Springer.

8. Freeman, Linton C. 2000. Visualizing Social Networks. *Journal of Social Structure* 1 (1). Dostopno prek: <http://www.cmu.edu/joss/content/articles/volume1/Freeman.html> (15. avgust 2014).
9. --- 2011. The Development of Social Network Analysis – with an Emphasis on Recent Events. V *The SAGE Handbook of Social Network Analysis*, ur. John Scott in Peter J. Carrington, 26–39. London: SAGE Publications Ltd.
10. Gilpin, Dawn R. 2011. Working the Twittersphere: Microblogging as Professional Identity Construction. V *A Networked Self: identity, community and culture on social network sites*, ur. Zizi Papacharissi, 232–250. New York: Routledge.
11. Han, Jiawei, Micheline Kamber in Jian Pei. 2011. *Data Mining: Concepts and Techniques, Third Edition*. Massachusetts: Morgan Kaufmann Publishers.
12. Hlebec, Valentina in Tina Kogovšek. 2006. *Merjenje socialnih omrežij*. Ljubljana: Študentska založba.
13. Huberman, Bernardo A., Daniel M. Romero in Fang Wu. 2009. Social networks that matter: Twitter under the microscope. *First Monday* 14 (1–5). Dostopno prek: <http://firstmonday.org/ojs/index.php/fm/article/view/2317/2063> (15. julij 2014).
14. Java, Akshay, Tim Finin, Xiaodan Song in Belle Tseng. 2007. *Why We Twitter: Understanding Microblogging Usage and Communities*. Dostopno prek: http://ebiquity.umbc.edu/_file_directory_/papers/369.pdf (29. junij 2014).
15. Lutz, Mark. 2013. *Learnin Python, Fifth Edition*. Sebastopol, California: O'Reilly Media.
16. Makice, Kevin. 2009. *Twitter API: Up and Running*. Sebastopol, California: O'Reilly Media.
17. Martinc, Rok. 2013. *Merjenje sentimenta na družbenem omrežju Twitter: izdelava orodja in evaluacija*. Magistrsko delo. Ljubljana: Univerza v Ljubljani, Fakulteta za družbene vede.

18. Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela in J. Niels Rosenquist. 2011. *Understanding the Demographics of Twitter Users*. Dostopno prek: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816/3234> (26. avgust 2014).
19. Mrvar, Andrej in Vladimir Batagelj. 2014. *Pajek and Pajek – XXL: Programs for Analysis and Visualization of Very Large Networks, Reference Manual: List of commands with short explanation version 3.15*. Dostopno prek: <http://mrvar.fdv.uni-lj.si/pajek/pajekman.pdf> (22. avgust 2014).
20. Novak, Mojca. 2004. Uvod. V *Omrežja socialne opore prebovalstva Slovenije*, ur. Mojca Novak, 7–11. Ljubljana: Inštitut Republike Slovenije za socialno varstvo.
21. Novalić, Ahmet. 2013. *Introduction to Tweepy, Twitter for Python*. Dostopno prek: <http://www.pythoncentral.io/introduction-to-tweepy-twitter-for-python> (20. avgust 2014).
22. *Python*. Dostopno prek: <https://www.python.org> (20. avgust 2014).
23. Russell, Matthew A. 2013. *Mining the Social Web, 2nd Edition*. Sebastopol, California: O'Reilly Media.
24. Scott, John. 2000. *Social Network Analysis: A Handbook, Second Edition*. London: Sage Publications.
25. *Tweepy*. Dostopno prek: <http://www.tweepy.org> (21. junij 2014).
26. *Twitter Inc*. Dostopno prek: <https://twitter.com> (20. junij 2014).
27. Wasserman, Stanley in Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

28. Wasserman, Stanley, John Scott in Peter J. Carrington. 2005. Introduction. V *Models and Methods in Social Network Analysis*, ur. Peter J. Carrington, John Scott in Stanley Wasserman, 1–7. New York: Cambridge University Press.
29. Weimann, Gabriel. 2010. Terror on Facebook, Twitter, and Youtube. *Brown Journal of World Affairs* 16 (2). Dostopno prek: <http://brown.edu/initiatives/journal-world-affairs/16.2/terror-facebook-twitter-and-youtube> (22. avgust 2014).
30. Witten, Ian H., Eibe Frank in Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*. Massachusetts: Morgan Kaufmann Publishers.

PRILOGE

PRILOGA A: KODA – ZBIRANJE TWEETOV

```
import time
import datetime
import tweepy
#from tweepy import OAuthHandler
import collections
#from collections import Counter
import pickle

#-----

# To so kljuci in tokeni za avtorizacijo, ki se jih dobi s kreacijo nove aplikacije na dev.twitter
ckey = "*****"
csecret = "*****"
atoken = "*****"
asecret = "*****"

auth = tweepy.OAuthHandler(ckey, csecret)
auth.set_access_token(atoken, asecret)

api = tweepy.API(auth,
    retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503]),
    #monitor_rate_limit=True, wait_on_rate_limit=True
)

#-----

slovarMain = {}
slovarImena = {}
imena = []

zac = datetime.datetime.now()

for tweet in tweepy.Cursor(api.search,
    q = "",
    geocode = "46.05,14.5,60km",
    count = 100,
    result_type = "recent",
    include_entities=True,
    since = "2013-12-09",
    until = "2013-12-13"
).items():

    print tweet.user.screen_name, "\t", tweet.created_at, "\t", tweet.id
    imena.append(tweet.user.screen_name)
    slovarMain[tweet.user.screen_name.encode("utf-8")] = {}
    slovarMain[tweet.user.screen_name.encode("utf-8")]["infoUser"] = tweet.user.__getstate__ ()
    slovarMain[tweet.user.screen_name.encode("utf-8")]["infoTweet"] = tweet.__getstate__ ()

    if len(imena)%17500 == 0:

        print "\nTrenutna dolzina seznama", len(imena)
        ser = api.rate_limit_status()["resources"]["search"]["/search/tweets"]
        st = datetime.datetime.fromtimestamp(ser["reset"]).strftime("%d-%m-%Y %H:%M:%S")
        print "" "\nLimit za search tweetov je %s klicev na 15 minut, kjer 15 tweetov pomeni 1 klic. Skupaj to pomeni 2700 do 18000 tweetov
na 15 minut.\nPreostalih klicev je %s. Naslednji API search reset %s.\n"" % (ser["limit"], ser["remaining"], st) # Prikaz limita za search
tweetov

        if ser["remaining"]<120:
            time.sleep(900)
            print "" "\nPojdimo naprej!\n"

imena_ociscena = list(set(imena)) # Znebimo se dvojnikov

dvojniki = collections.Counter() # Dvojniki oz seznam imen s pojavitvami v prvotnem seznamu
for dv in imena:
    dvojniki[dv] += 1
```



```

#print ""Prvotni seznam prikaznih imen dolzine: %i\n\n%r\n\n
#Ociscen seznam prikaznih imen dolzine: %i\n\n%r\n\n
#Seznam vseh imen s pojavitvami, kjer vidimo dvojnike:\n\n%r\n\n"" % (len(imena), imena, len(imena_ociscena), imena_ociscena,
dvojniki) # Prikaz seznamov imen

print "Originalen seznam dolzine: %i\nOciscen seznam dolzine: %i\n" % (len(imena), len(imena_ociscena))

#ser = api.rate_limit_status()["resources"]["search"]["/search/tweets"]
#st = datetime.datetime.fromtimestamp(ser["reset"]).strftime("%d-%m-%Y %H:%M:%S")
#print ""Limit za search tweetov je %s klincev na 15 minut, kjer 15 tweetov pomeni 1 klic. Skupaj to pomeni 2700 tweetov na 15 minut.
#Preostalih klincev je %s. Naslednji API search reset %s.\n"" % (ser["limit"], ser["remaining"], st) # Prikaz limita za search tweetov

slovarImena["Imena"] = imena
slovarImena["Ociscena imena"] = imena_ociscena

pickle.dump(slovarImena, open("Imena12122013.p", "wb"))
pickle.dump(slovarMain, open("Main12122013.p", "wb"))

print len(imena_ociscena)
print len(slovarMain.keys())

kon = datetime.datetime.now()

print "Zacetek: %s\nKonec: %s" % (zac, kon)

#slovarMain = pickle.load(open("Main25112013.p", "rb"))

```

PRILOGA B: KODA – ZBIRANJE FOLLOWEES

```

import time
import datetime
import tweepy
#from tweepy import OAuthHandler
import collections
#from collections import Counter
import itertools
import pickle
import random
#-----
kljuci = {"app1": ["*****", "*****", "*****_*****", "*****"],
"app2": ["*****", "*****", "*****_*****", "*****"],
"app3": ["*****", "*****", "*****_*****", "*****"],
"app4": ["*****", "*****", "*****_*****", "*****"],
"app5": ["*****", "*****", "*****_*****", "*****"],
"app6": ["*****", "*****", "*****_*****", "*****"],
"app7": ["*****", "*****", "*****_*****", "*****"],
"app8": ["*****", "*****", "*****_*****", "*****"],
"app9": ["*****", "*****", "*****_*****", "*****"],
"app10": ["*****", "*****", "*****_*****", "*****"]}
#-----

slovarMain = pickle.load(open("slovarMainSest1211_1212.p", "rb"))

slovarFriends = {}
seznamUserjevError = []

seznam = slovarMain.keys()

nrError = 0
runda = 1
cifra = 1
kombo = "app" + str(cifra)

auth = tweepy.OAuthHandler(kljuci[kombo][0], kljuci[kombo][1])
auth.set_access_token(kljuci[kombo][2], kljuci[kombo][3])
api = tweepy.API(auth,
retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503])
)

```

```

zac = datetime.datetime.now()
print "Zacetek: %s\n" %zac

for screenName in seznam: #seznam15

    try:
        if api.rate_limit_status()["resources"]["friends"]["/friends/ids"]["remaining"] > 0:
            if api.rate_limit_status()["resources"]["friends"]["/friends/ids"]["remaining"] >
slovarMain[screenName]["infoUser"]["friends_count"]/5000:

                print "Oznaka kombinacije: %s\n" %kombo

                idsTemp = []

                #print api1.rate_limit_status()["resources"]["friends"]["/friends/ids"], "\n"
                print "Uporabnik:", screenName, "sledil", slovarMain[screenName]["infoUser"]["friends_count"], "osebam.\n"

                for page in tweepy.Cursor(api.friends_ids, screen_name = screenName).pages(): # friends_ids nam vrne do 5000 user id-jev na
"stran" za friends nasege userja (opomba: metoda vraca sezname)
                    idsTemp.extend(page)

                slovarFriends[screenName] = {}
                slovarFriends[screenName]["Friends"] = {}
                slovarFriends[screenName]["Friends"]["IDs"] = idsTemp

                time.sleep(9)

    else:
        cifra += 1

        if cifra > 10:
            cifra = 1
            vmesni = datetime.datetime.now()
            print vmesni
            print "Runda %s:\nPretece do sedaj: %s" %(runda, vmesni-zac)
            print "Stevilo obravnavanih userjev: %s\n" %len(slovarFriends.keys())
            print random.choice(slovarFriends.items()), "\n"
            runda += 1

            kombo = "app" + str(cifra)
            auth = tweepy.OAuthHandler(kljuci[kombo][0], kljuci[kombo][1])
            auth.set_access_token(kljuci[kombo][2], kljuci[kombo][3])
            api = tweepy.API(auth,
                retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503])
            )

        else:
            kombo = "app" + str(cifra)
            auth = tweepy.OAuthHandler(kljuci[kombo][0], kljuci[kombo][1])
            auth.set_access_token(kljuci[kombo][2], kljuci[kombo][3])
            api = tweepy.API(auth,
                retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503])
            )

        print "Oznaka kombinacije: %s\n" %kombo

        idsTemp = []

        #print api1.rate_limit_status()["resources"]["friends"]["/friends/ids"], "\n"
        print "Uporabnik:", screenName, "sledil", slovarMain[screenName]["infoUser"]["friends_count"], "osebam.\n"

        for page in tweepy.Cursor(api.friends_ids, screen_name = screenName).pages(): # friends_ids nam vrne do 5000 user id-jev na
"stran" za friends nasege userja (opomba: metoda vraca sezname)
            idsTemp.extend(page)

        slovarFriends[screenName] = {}
        slovarFriends[screenName]["Friends"] = {}
        slovarFriends[screenName]["Friends"]["IDs"] = idsTemp

        time.sleep(9)

    except tweepy.TweepError, e:
        seznamUserjevError.append(screenName)
        nrError += 1
        print "Napaka st.: %s\nNapaka zaradi %s" %(nrError, e.reason)

```

```

#pickle.dump(slovarFriends, open("FriendsIDS25112013.p", "wb"))
#pickle.dump(seznamUserjevError, open("UserjiErrorPonoviIDS25112013.p", "wb"))
pass

pickle.dump(slovarFriends, open("FriendsIDSslovarMainSest1211_1212.p", "wb"))
pickle.dump(seznamUserjevError, open("UserjiErrorPonoviIDS1211_1212.p", "wb"))
kon = datetime.datetime.now()
print "Konec:", kon
print "Od zacetka do konca:", kon-zac
print len(slovarFriends.keys())

```

PRILOGA C: KODA – ZBIRANJE FOLLOWERS

```

import time
import datetime
import tweepy
#from tweepy import OAuthHandler
import collections
#from collections import Counter
import itertools
import pickle
import random
#-----
kljuci = {"app1": ["*****", "*****", "*****_*****", "*****"],
"app2": ["*****", "*****", "*****_*****", "*****"],
"app3": ["*****", "*****", "*****_*****", "*****"],
"app4": ["*****", "*****", "*****_*****", "*****"],
"app5": ["*****", "*****", "*****_*****", "*****"],
"app6": ["*****", "*****", "*****_*****", "*****"],
"app7": ["*****", "*****", "*****_*****", "*****"],
"app8": ["*****", "*****", "*****_*****", "*****"],
"app9": ["*****", "*****", "*****_*****", "*****"],
"app10": ["*****", "*****", "*****_*****", "*****"]}
#-----

slovarMain = pickle.load(open("slovarMainSest1211_1212.p", "rb"))

slovarFollowerji = {}
seznamUserjevError = []

seznam = slovarMain.keys()

nrError = 0
runda = 1
cifra = 1
kombo = "app" + str(cifra)

auth = tweepy.OAuthHandler(kljuci[kombo][0], kljuci[kombo][1])
auth.set_access_token(kljuci[kombo][2], kljuci[kombo][3])
api = tweepy.API(auth,
    retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503])
)

zac = datetime.datetime.now()
print "Zacetek: %s\n" %zac

for screenName in seznam: #seznam15

    try:
        if api.rate_limit_status()["resources"]["followers"]["followers/ids"]["remaining"] > 0:
            if api.rate_limit_status()["resources"]["followers"]["followers/ids"]["remaining"] >
slovarMain[screenName]["infoUser"]["followers_count"]/5000:

                print "Oznaka kombinacije: %s\n" %kombo

                idsTemp = []

                #print api.rate_limit_status()["resources"]["followers"]["followers/ids"], "\n"
                print "Uporabniku:", screenName, "sledi", slovarMain[screenName]["infoUser"]["followers_count"], "oseb.\n"

                for page in tweepy.Cursor(api.followers_ids, screen_name = screenName).pages(): # follower_ids nam vrne do 5000 user id-jev na
"stran" za followerje nasege userja (opomba: metoda vraca sezname)

```

```

        idsTemp.extend(page)

        slovarFollowerji[screenName] = {}
        slovarFollowerji[screenName]["Followerji"] = {}
        slovarFollowerji[screenName]["Followerji"]["IDs"] = idsTemp

        time.sleep(9)

    else:
        cifra += 1

        if cifra > 10:
            cifra = 1
            vmesni = datetime.datetime.now()
            print vmesni
            print "Runda %s:\nPretečeno do sedaj: %s" %(runda, vmesni-zac)
            print "Stevilo obravnavanih userjev: %s\n" %len(slovarFollowerji.keys())
            print random.choice(slovarFollowerji.items()), "\n"
            runda += 1

            kombo = "app" + str(cifra)
            auth = tweepy.OAuthHandler(kljuci[kombo][0], kljuci[kombo][1])
            auth.set_access_token(kljuci[kombo][2], kljuci[kombo][3])
            api = tweepy.API(auth,
                            retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503])
                            )

        else:
            kombo = "app" + str(cifra)
            auth = tweepy.OAuthHandler(kljuci[kombo][0], kljuci[kombo][1])
            auth.set_access_token(kljuci[kombo][2], kljuci[kombo][3])
            api = tweepy.API(auth,
                            retry_count=3, retry_delay=5, retry_errors=set([401, 404, 500, 503])
                            )

        print "Oznaka kombinacije: %s\n" %kombo

        idsTemp = []

        #print api1.rate_limit_status()["resources"]["followers"]["followers/ids"], "\n"
        print "Uporabniku:", screenName, "sledi", slovarMain[screenName]["infoUser"]["followers_count"], "oseb.\n"

        for page in tweepy.Cursor(api.followers_ids, screen_name = screenName).pages(): # follower_ids nam vrne do 5000 user id-jev na
"stran" za followerje nasege userja (opomba: metoda vraca sezname)
            idsTemp.extend(page)

            slovarFollowerji[screenName] = {}
            slovarFollowerji[screenName]["Followerji"] = {}
            slovarFollowerji[screenName]["Followerji"]["IDs"] = idsTemp

            time.sleep(9)

except tweepy.TweepError, e:
    seznamUserjevError.append(screenName)
    nrError += 1
    print "Napaka st.: %s\nNapaka zaradi %s" %(nrError, e.reason)
    #pickle.dump(slovarFollowerji, open("FollowerjiIDS25112013.p", "wb"))
    #pickle.dump(seznamUserjevError, open("UserjiErrorPonoviIDS25112013.p", "wb"))
    pass

pickle.dump(slovarFollowerji, open("FollowerjiIDSslovarMainSest1211_1212.p", "wb"))
pickle.dump(seznamUserjevError, open("UserjiErrorPonoviIDS1211_1212.p", "wb"))
kon = datetime.datetime.now()
print "Konec:", kon
print "Od zacetka do konca:", kon-zac
print len(slovarFollowerji.keys())

```

PRILOGA Č: PAJEK – OSNOVNE LASTNOSTI OMREŽJA

Number of vertices (n): 13807

	Arcs	Edges
Total number of lines	552389	0
Number of loops	0	0
Number of multiple lines	0	0

Density1 [loops allowed] = 0.00289765
 Density2 [no loops allowed] = 0.00289786
 Average Degree = 80.01578909

PRILOGA D: PAJEK – STOPNJE TOČK

Stopnje točk razvrščene po skupni stopnji (največjih 50 vrednosti)

Enota	Vhodne stopnje točk	Izhodne stopnje točk	Skupne stopnje točk
Val202	2165	1220	3385
vladaRS	1696	1612	3308
rtvslo	1873	839	2712
praprotnix	1495	802	2297
SpletnaMladina	1877	321	2198
RTV_Slovenija	1574	550	2124
toplovodar	1483	569	2052
TinaMaze	1937	0	1937
Dnevnik_si	1694	210	1904
Delo	1778	124	1902
petrasovdat	1349	551	1900
crnkovic	1662	193	1855
Simobil	1236	589	1825
andrejkaroli	1538	271	1809
nejcjemec	1030	758	1788
kinosiska	1028	757	1785
vanfranco	748	1009	1757
NeuroVirtu	856	897	1753
LJ_Taksist	1714	38	1752
zzTurk	1541	207	1748
BorutPahor	1472	263	1735
TelekomSlo	998	697	1695
mn3njalik	933	759	1692
racnovice	599	1069	1668

IEBergant	1611	51	1662
24UR	1100	558	1658
DC43	1369	285	1654
finance_si	1505	87	1592
had	1406	178	1584
matjazsircelj	950	584	1534
MatijaStepisnik	1118	415	1533
kricac	1379	146	1525
markopotrc	1279	243	1522
KatarinaKresal	1186	335	1521
StudioCity_	1318	202	1520
jasna	812	692	1504
multikultivator	868	634	1502
mojcakleva	675	798	1473
policija_si	1393	78	1471
pengovsky	1225	237	1462
STA_novice	1359	70	1429
bosstjanz	517	851	1368
steinbuch	1032	315	1347
Vikend	709	623	1332
TopSlo	797	533	1330
ursicdomen	558	771	1329
alivea	1033	287	1320
TeamSlovenia	1096	224	1320
radio_kaos	523	788	1311
radiostudent	1046	255	1301

PRILOGA E: PAJEK – MERE SREDIŠČ. IN POMEMBNOСТИ

Vhodna dostopnost, izhodna dostopnost in vmesnost (50 vrednosti glede na enote stopenj točk)

Enota	Vhodna dostopnost	Izhodna dostopnost	Vmesnost
Val202	0.233562	0.21626	0.018549
vladaRS	0.227085	0.225985	0.032241
rtvslo	0.235712	0.214034	0.021037
praprotnix	0.227757	0.209948	0.005634
SpletnaMladina	0.229768	0.198383	0.003191
RTV_Slovenija	0.227046	0.205019	0.006063
toplovodar	0.225376	0.210197	0.007862
TinaMaze	0.246182	0	0
Dnevnik_si	0.227986	0.195953	0.001543
Delo	0.233906	0.191036	0.001644
petrasovdat	0.222115	0.204725	0.002302
crnkovic	0.227783	0.193204	0.001283
Simobil	0.222078	0.204725	0.008167
andrejkaroli	0.22312	0.198744	0.001168
nejcjemec	0.220247	0.213578	0.004032
kinosiska	0.219911	0.207669	0.004161
vanfranco	0.216811	0.213027	0.004065
NeuroVirtu	0.217477	0.218109	0.013244
LJ_Taksist	0.22811	0.183953	0.000773
zzTurk	0.22813	0.195078	0.001897
BorutPahor	0.22935	0.194735	0.002493
TelekomSlo	0.229204	0.206525	0.012043
mn3njalnik	0.217751	0.215095	0.007493
racnovice	0.211874	0.222423	0.009212

IEBergant	0.224916	0.185698	0.000349
24UR	0.218772	0.205058	0.003598
DC43	0.224744	0.197723	0.001552
finance_si	0.220983	0.190502	0.001775
had	0.226083	0.195793	0.001123
matjazsircelj	0.219383	0.209752	0.003089
MatijaStepisnik	0.221718	0.206916	0.002523
kricac	0.220033	0.191866	0.00042
markopotrc	0.225935	0.197844	0.001943
KatarinaKresal	0.224382	0.198134	0.002664
StudioCity_	0.221057	0.195741	0.00082
jasna	0.217697	0.208333	0.003175
multikultivator	0.21954	0.21009	0.003257
mojcakleva	0.214868	0.214734	0.004328
policija_si	0.222389	0.190594	0.000482
pengovsky	0.223378	0.197897	0.000903
STA_novice	0.221774	0.188078	0.00034
bosstjanz	0.2126	0.214976	0.001794
steinbuch	0.219893	0.199043	0.00126
Vikend	0.215422	0.210078	0.004483
TopSlo	0.219043	0.209716	0.004355
ursicdomen	0.2128	0.208773	0.001525
alivea	0.2216	0.201846	0.00173
TeamSlovenia	0.230402	0.191633	0.004993
radio_kaos	0.213941	0.214591	0.003892
radiostudent	0.219443	0.19736	0.001796

PRILOGA F: PAJEK – OBMOČJE VPLIVA IN BLIŽINA

Območje vpliva in bližina (50 vrednosti glede na enote stopenj točk)

Enota	Območje vpliva norm.	Bližina
Val202	0,758149	0,233556
vladaRS	0,758149	0,22708
rtvslo	0,758149	0,235707
praprotnix	0,758149	0,227751
SpletnaMladina	0,758149	0,229763
RTV_Slovenija	0,758149	0,227041
toplovodar	0,758149	0,225371
TinaMaze	0,762205	0,246177
Dnevnik_si	0,758149	0,22798
Delo	0,758149	0,2339
petrasovdat	0,758149	0,22211
crnkovic	0,758149	0,227778
Simobil	0,758149	0,222073
andrejkaroli	0,758149	0,223115
nejcjemec	0,758149	0,220242
kinosiska	0,758149	0,219906
vanfranco	0,758149	0,216806
NeuroVirtu	0,758149	0,217472
LJ_Taksist	0,758149	0,228105
zzTurk	0,758149	0,228125
BorutPahor	0,758149	0,229344
TelekomSlo	0,758149	0,229199
mn3njalik	0,758149	0,217746
racnovice	0,758149	0,211869

IEBergant	0,758149	0,224911
24UR	0,758149	0,218767
DC43	0,758149	0,224739
finance_si	0,758149	0,220978
had	0,758149	0,226077
matjazsircelj	0,758149	0,219377
MatijaStepisnik	0,758149	0,221713
kricac	0,758149	0,220028
markopotrc	0,758149	0,225929
KatarinaKresal	0,758149	0,224377
StudioCity_	0,758149	0,221052
jasna	0,758149	0,217692
multikultivator	0,758149	0,219535
mojcakleva	0,758149	0,214863
policija_si	0,758149	0,222384
pengovsky	0,758149	0,223373
STA_novice	0,758149	0,221768
bosstjanz	0,758149	0,212595
steinbuch	0,758149	0,219888
Vikend	0,758149	0,215417
TopSlo	0,758149	0,219038
ursicdomen	0,758149	0,212795
alivea	0,758149	0,221595
TeamSlovenia	0,758149	0,230396
radio_kaos	0,758149	0,213936
radiostudent	0,758149	0,219438

PRILOGA G: PAJEK – KAZALA IN OPISI

Kazala in opisi (50 vrednosti glede na enote stopenj točk)

Enota	Kazala	Opisi
Val202	0.074709	0.114686
vladaRS	0.069848	0.095214
rtvslo	0.054576	0.098575
praprotnix	0.06333	0.097809
SpletnaMladina	0.033468	0.102821
RTV_Slovenija	0.043012	0.087002
toplovodar	0.042319	0.087342
TinaMaze	0	0.081839
Dnevnik_si	0.02762	0.096476
Delo	0.02039	0.100647
petrasovdat	0.053157	0.092553
crnkovic	0.023478	0.094769
Simobil	0.030999	0.061725
andrejkaroli	0.034686	0.092177
nejcjemec	0.057639	0.073624
kinosiska	0.05737	0.059474
vanfranco	0.069367	0.058054
NeuroVirtu	0.056907	0.064199
LJ_Taksist	0.007233	0.084141
zzTurk	0.020786	0.090839
BorutPahor	0.030744	0.081017
TelekomSlo	0.043979	0.055333
mn3njalik	0.058787	0.06182
racnovice	0.051607	0.034848

IEBergant	0.010297	0.093632
24UR	0.035767	0.069737
DC43	0.03441	0.091164
finance_si	0.010953	0.085731
had	0.017849	0.085708
matjazsircelj	0.043635	0.068227
MatijaStepisnik	0.044472	0.081819
kricac	0.020687	0.092216
markopotrc	0.026635	0.08151
KatarinaKresal	0.034023	0.074547
StudioCity_	0.028512	0.082095
jasna	0.053289	0.056782
multikultivator	0.048462	0.065165
mojcakleva	0.065496	0.055368
policija_si	0.012926	0.079095
pengovsky	0.029633	0.086544
STA_novice	0.009768	0.081185
bosstjanz	0.066588	0.040584
steinbuch	0.03085	0.073785
Vikend	0.034841	0.05082
TopSlo	0.026828	0.053125
ursicdomen	0.065237	0.045883
alivea	0.026737	0.069622
TeamSlovenia	0.015752	0.05368
radio_kaos	0.059579	0.042068
radiostudent	0.025722	0.062761