

**UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE**

LUKA KRONEGGER

**Mentorica: doc. dr. Valentina Hlebec
Somentor:izr. prof. dr. Vasja Vehovar**

**VPLIV MANJKAJOČIH PODATKOV NA
LASTNOSTI
EGOCENTRIČNIH OMREŽIJ**

DIPLOMSKO DELO

LJUBLJANA, 2005

Kazalo

1	Uvod	4
2	Socialna omrežja.....	6
2.1	Opredelitev socialnih omrežij.....	6
2.2	Zgodovina.....	7
2.3	Egocentrična socialna omrežja	9
3	Zbiranje podatkov o osebnih omrežjih	10
3.1	Načini zbiranja podatkov	10
3.2	Zbiranje podatkov in informacijska tehnologija.....	12
3.3	Metodološki problemi zbiranja podatkov	13
4	Vzorčenje alterjev v egocentričnih omrežjih.....	14
4.1	Osnovna ideja vzorčenja.....	14
4.2	Koncept vzorčenja	15
4.3	Implementacija vzorčenja in tehnične zahteve.....	15
4.4	Optimalna velikost podvzorca	16
5	Analiza in rezultati	20
5.1	Opis podatkov.....	21
5.2	Struktura omrežij	22
5.3	Druge značilnosti omrežij.....	26
5.4	Ugotovitve	34
6	Sklep	35
7	Seznam literature	37
Priloga A: intraklasna korelacija ρ in optimalna velikost omrežja glede na stroške		39
Priloga B: Struktura in druge lastnosti omrežij glede na njihovo velikost		42
Priloga C: Sintaksa R-project		46

Kazalo slik in tabel

Slika 3.1: Egocentrično omrežje je sestavljeno iz ega in povezav do njegovih alterjev	11
Slika 5.1: Razmerje med intraklasno korelacijo (ρ) in optimalno velikostjo omrežja.	18
Slika 5.1: Število omrežij glede na njihovo velikost.	22
Slika 6.1: Pogostost stikov	26
Slika 6.2: Bližina osebe	28
Slika 6.3: Pomembnost osebe.....	29
Slika 6.4: Spravljanje v slabo voljo	30
Slika 6.5: Trajanje znanstva.....	31
Slika 6.6: Oddaljenost bivanja.....	33
Tabela 5.1: Optimalna velikost vzorčenih omrežij glede na razmerje stroškov in intraklasne korelacije.	17
Tabela 6.1: Struktura omrežij – telefonska anketa	22
Tabela 6.2: Struktura omrežij – spletna anketa	24
Tabela 6.3: Pogostost stikov – Telefonska anketa.....	26
Tabela 6.4: Pogostost stikov – Spletna anketa	27
Tabela 6.5: Bližina osebe – Telefonska anketa	28
Tabela 6.6: Bližina osebe –Spletna anketa	28
Tabela 6.7: Pomembnost osebe – Telefonska anketa.....	29
Tabela 6.8: Pomembnost osebe – Spletna anketa.....	29
Tabela 6.9: Spravljanje v slabo voljo – Telefonska anketa	31
Tabela 6.10: Spravljanje v slabo voljo – Spletna anketa.....	31
Tabela 6.11: Trajanje znanstva– Telefonska anketa.....	32
Tabela 6.12: Trajanje znanstva– Spletna anketa	32
Tabela 6.13: Trajanje znanstva – Primerjava povprečij v letih	32
Tabela 6.14: Oddaljenost bivanja – Telefonska anketa.....	33
Tabela 6.15: Oddaljenost bivanja – Spletna anketa.....	33
Tabela A.1: Izračunana intraklasna korelacija po velikostih omrežij (spletna anketa)	39
Tabela A.2: Izračunana optimalna velikost omrežja glede na dano razmerje stroškov	40
Tabela A.3: Izračunana intraklasna korelacija po velikostih omrežij (telefonska anketa)..	40
Tabela A.4: Izračunana optimalna velikost omrežja glede na dano razmerje stroškov	41
Tabela B.1: Struktura omrežij glede na njihovo velikost – spletna anketa	Error! Bookmark not defined.
Tabela B.2: Struktura omrežij glede na njihovo velikost – telefonska anketa	Error! Bookmark not defined.
Tabela B.3: Lastnosti omrežij glede na njihovo velikost – spletna anketa, 1.del.....	42
Tabela B.4: Lastnosti omrežij glede na njihovo velikost – spletna anketa, 2.del.....	43
Tabela B.5: Lastnosti omrežij glede na njihovo velikost – telefonska anketa, 1.del.....	44
Tabela B.6: Lastnosti omrežij glede na njihovo velikost – telefonska anketa, 2.del.....	45

1 Uvod

Analiza osebnih oziroma egocentričnih socialnih omrežij je relativno novo področje v družboslovnem raziskovanju, zato se na vsakem koraku srečuje z novimi izzivi in idejami.

Metodološki problemi, s katerimi se srečujemo pri zbiranju podatkov o osebnih omrežjih, so poleg splošnih težav, ki so v anketnem raziskovanju že dodobra raziskane in obravnavane, vezani predvsem na obremenjenost anketirancev zaradi velikega števila vprašanj. Navedeno vpliva na neodgovarjanje na vprašanja oz. prekinitev procesa anketiranja. Z omenjenimi težavami se seveda pojavi vprašanje kvalitete, zanesljivosti odgovorov in tudi splošne veljavnosti raziskave. Pojavi se lahko tudi nevarnost visokih stroškov anketiranja. Problem neodgovorov je še posebej pereč pri novih metodah anketiranja (npr. anketiranje prek spleta – web-surveys), kjer anketiranci na anketo odgovarjajo sami, brez pomoči anketarja (self-administrated surveying).

V diplomski nalogi preučujemo inovativno rešitev, ki temelji na zmanjševanju števila vprašanj in doslej skoraj še ni bila obravnavana. To dosežemo tako, da anketirancev ne sprašujemo podrobnosti o vseh osebah navedenih v njihovem omrežju, ampak le o vzorcu oseb iz tega omrežja. Če bi npr. anketiranec pri uveljavljeni metodi zbiranja podatkov v omrežje navedel 30 oseb (alterjev), o katerih bi za vsako posebej vprašali nadaljnjih 10 vprašanj o njihovih značilnostih, bi anketiranec odgovarjal na 300 med seboj zelo podobnih vprašanj. S predlagano metodo, bi po 10 vprašanj zastavili le o vzorcu oseb iz omrežja npr. desetim, s čimer bi se število vprašanj, ki jih zastavimo anketirancu zmanjšalo iz 300 na 100 oz. na tretjino.

Cilj naloge je ugotoviti, v kakšni meri opisano zmanjševanje obsega zbranih podatkov o egocentričnih omrežjih anketirancev vpliva na kvaliteto informacije, ki nam jo ti podatki nudijo.

Celotno nalogo sestavlja teoretičen del, kjer sta v drugem in tretjem poglavju predstavljena splošna zgodovina razvoja analize socialnih omrežij ter razvoj metod zbiranja podatkov o egocentričnih omrežjih. V četrtem poglavju govorimo o splošni teoriji vzorčenja, vzorčenju alterjev v egocentričnih omrežjih, opredelimo pa tudi stroške anketiranja in izračunamo optimalno velikost podvzorcev. V petem poglavju, empiričnem delu naloge, testiramo posledice zmanjševanja količine podatkov na zbrano informacijo o omrežjih domneve na realnih podatkih in predstavimo ugotovitve.

2 Socialna omrežja

2.1 Opredelitev socialnih omrežij

Socialno omrežje predstavljajo enote (npr. osebe, podjetja, države, manjša omrežja ipd.), ki so med seboj povezane z določenimi socialnimi povezavami. Če omenimo le nekaj izmed možnih relacij, so to prijateljstvo, partnerstvo, izmenjava informacij, finančne transakcije ali odvisnost (Kogovšek, 2001:15).

Pomembna razlika med analizo socialnih omrežij in klasično družboslovno analizo je, da se pri pojasnjevanju delovanja posameznikov in družbenih pojavov ne omejuje le na demografske značilnosti enot. Analiza socialnih omrežij preučuje relacije med enotami in tako predstavlja pomemben prehod od analize splošnega individualizma k strukturalni analizi.

Eden glavnih ciljev analize socialnih omrežij je iskanje zakonitosti relacij med akterji ter pojasnjevanje strukture in delovanja omrežij. Ideja, po kateri se ravna analiza socialnih omrežij je, da način, na kakršnega je posameznik vpet v okolje, v veliki meri vpliva na življenje oz. delovanje posameznika. Princip preučevanja socialnih omrežij temelji na analizi diad, triad ter manjših in večjih skupin, omrežij in delov omrežij, na sledenju tokov informacij, materialnih in drugih dobrin.

Poleg poudarka na relacijah velja naštetih še druge značilnosti analize socialnih omrežij (Wellman, 1988:19):

- akterji in njihova dejanja so obravnavani kot medsebojno odvisne, ne pa neodvisne enote, kot je to običajno za klasično družboslovno anketno raziskovanje,
- relacije med akterji so kanali za prenos in pretok virov (nematerialnih in materialnih),
- modeli, ki temeljijo na omrežjih in ki se osredotočajo na posameznika znotraj omrežja, vidijo strukturo omrežja kot okolje, ki nudi priložnosti in omejitve za posameznikovo delovanje,
- modeli, ki temeljijo na omrežjih, strukture konceptualizirajo kot relativno trajne vzorce med posamezniki.

Za raziskovanje socialnih omrežij se v znanosti uporabljata dva pristopa, ki vsak posebej služita svojemu namenu glede na naravo omrežja, ki ga raziskujemo. Prvi raziskovalni pristop se ukvarja s popolnimi omrežji. To so tista, kjer imamo podatke o vseh enotah v omrežju in o vseh povezavah med enotami - dejansko gre za celotno populacijo. O popolnem omrežju govorimo na primer ob analizi odnosov med vsemi učenci enega razreda. Kadar raziskujemo popolno omrežje, posamezni anketiranec ponavadi odgovarja o svojih relacijah do vseh drugih članov obravnavanega omrežja. Težava pri takem zbiranju podatkov je, da je drago in težko izvedljivo. Tak pristop uporabljamo za omrežja z vnaprej jasno določenimi mejami, ki v večini primerov niso velika.

Alternativa analizi popolnih omrežij je analiza osebnih (egocentričnih) omrežij, kjer analiziramo lokalna omrežja posameznikov (egov), vključenih v vzorec. Člani osebnega omrežja posameznika se imenujejo alterji. S tem pristopom sicer ne opisujemo socialne strukture celotne populacije, ampak na podlagi reprezentativnega vzorca socialnih okolij posameznikov značilnosti posplošujemo na celotno populacijo (Marsden, 1990).

2.2 Zgodovina

Začetki analize socialnih omrežij segajo v prvo polovico prejšnjega stoletja. Ideja o analizi relacij med posamezniki se je v istem časovnem obdobju porodila na različnih področjih, ki so bila med seboj ločena.

Na področju socialne psihologije za najopaznejšega predstavnika šteje J.L. Moreno, ki je eden izmed začetnikov sociometrije. Sociometrija je veda, ki preučuje medosebne odnose znotraj manjših skupin. Njeni predstavniki so trdili, da družba ni le agregat posameznikov, ampak struktura medosebnih odnosov, ki sčasoma oblikuje večino družbenih pojavov, kakršna sta na primer ekonomija in država. Do podatkov o relacijah so prišli z vprašanji o prijateljskih odnosih znotraj skupin ter delovanje posameznikov pojasnjevali z ovirami in priložnostmi, ki so jih prijateljske vezi ponujale.

V 30. letih 20. stoletja je Moreno predstavil sociogram, glavno orodje sociometrikov. To je diagram točk in povezav med njimi, ki predstavljajo relacije med ljudmi. S pomočjo

sociograma je kot prvi opredelil voditelje in izolirane člane skupin, odkril je asimetrijo in recipročnost prijateljskih vezi. Obravnaval je tudi verige neposrednih povezav v omrežju. Ena izmed značilnih oblik omrežij, ki jih je opazoval, je bila sociometrična zvezda; posameznik v skupini, ki ga je večina članov v skupini opredelila kot prijatelja.

V istem obdobju so na področju medosebnih povezav delovali tudi antropologi. Že v 20-ih letih 20. stoletja je skupina s Harvarda preučevala vpliv delovnih pogojev in medosebnih odnosov med zaposlenimi na delovno produktivnost (Mayo v Scott, 2000). Primerjali so neformalne odnose v podjetjih z njihovo formalno strukturo. Podatke o omrežjih so predstavili z urejenimi matrikami in z analizo relacij med akterji v omrežjih opredelili značilnosti, kakršna je npr. klika. Postavili so temelje za t.i. bločno modeliranje (Homans v Scott, 2000).

Druga skupina antropologov, ki je delovala v Manchestru, je pod močnim vplivom Radcliffe-Brownove antropološke teorije preučevala spremembe na podlagi moči in konflikta. Delo te skupine je (vsaj v začetku) v veliki meri temeljilo na preučevanju afriških skupnosti, kjer formalna (institucionalna) struktura ni igrala pomembne vloge. Glavna predstavnika v tej skupini sta bila Barnes in Bott-ova. Prvi je od Radcliffe-Browna prevzel idejo omrežja medosebnih odnosov in jo vpeljal v prakso, zaradi česar velja za utemeljitelja analize socialnih omrežij. Elizabeth Bott pa je delo Barnesa povezala z delom sociometrikov. Veliko delo v tej skupini je opravil Mitchell, ki se je preučevanja omrežij lotil s pomočjo matematike. Teorijo grafov in ideje sociometrikov je prevedel v jezik sociologov. Mitchell je tudi tisti, ki je poleg strukture opredelil vsebino relacij med posamezniki. Po vsebini je ločil komunikacijske in instrumentalne vezi.

Manchesterska skupina je imela v Veliki Britaniji velik vpliv. Kljub temu je njihov pristop obveljal za omejenega na neformalne vezi v majhnih skupnostih in na analizo egocentričnih socialnih omrežij.

Dejanski preboj analize socialnih omrežij kot pomembne raziskovalne veje se je zgodil na Harvardu, kjer je v 60-ih in 70-ih letih 20. stoletja deloval Harrison White s skupino študentov. White je v analizo vključil uporabo algebralnega modeliranja skupin, kar je analizi dalo nov zagon (predvsem na področju bločnega modeliranja). Eden izmed White-ovih študentov, Mark Granovetter, je leta 1974 izdal knjigo "Getting a Job", ki velja za

temeljno delo na področju analize socialnih omrežij. V raziskavi, ki jo je opravil, je ljudi spraševal o načinu, kako so našli zaposlitev, ki jo imajo. Delo je postalo zgled mnogim drugim raziskovalcem, kar je analizo socialnih omrežij postavilo na mesto enega najpomembnejših pristopov v družboslovnem raziskovanju.

Kratek časovni pregled razvoja na splošnem področju socialnih omrežij je povzet po J. Scott Social Network Analysis: A handbook. (Scott, 2000: 7)

2.3 Egocentrična socialna omrežja

Temelje za analizo egocentričnih omrežij je torej postavil Barnes (Barnes 1954: 43). Med prve večje raziskave na področju egocentričnih omrežij sodijo Laumannova raziskava o treh najbližjih prijateljih, ki je bila izvedena v Detroitu leta 1973 (Laumann, 1973), v istem obdobju je Barry Wellman že izvedel prvi del raziskave o East Yorku, ki jo je nadaljeval v letu 1978 (Wellman, 1979; Wellman in Wortley, 1990). Leta 1982 je Claude S. Fisher izdal prvo obsežnejše delo na področju raziskav o egocentričnih omrežjih, ki temelji na raziskavi o Skupnostih severne Kalifornije (Fisher, 1982). Za največji projekt v zgodnjem obdobju raziskav o egocentričnih omrežij pa velja ameriški popis leta 1985, nad katerim sta bedela Burt (Burt, 1984) in Marsden (Marsden, 1987).

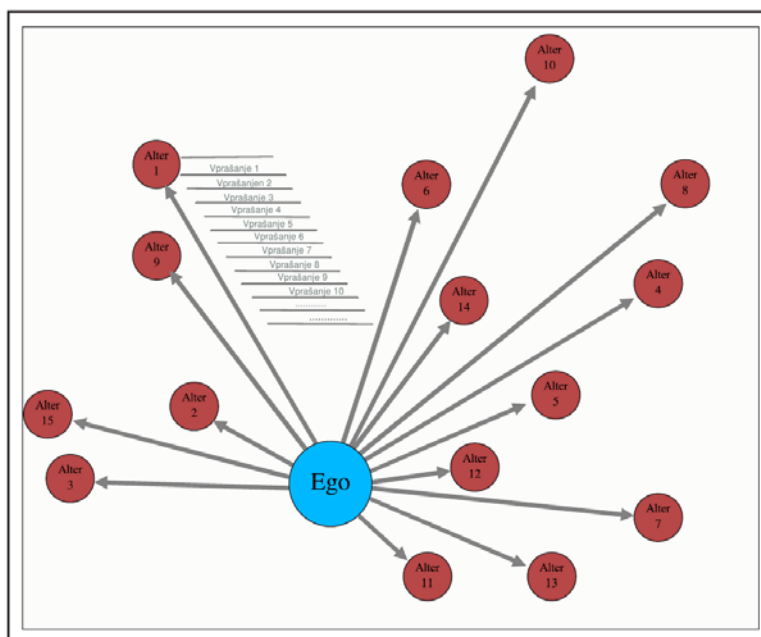
3 Zbiranje podatkov o osebnih omrežjih

Podatke o omrežjih je mogoče zbrati iz različnih virov. Na splošno so za zbiranje podatkov o omrežjih pogosto uporabljani arhivski viri, dnevniki in elektronski sledilniki, ki omogočajo obravnavo velikih količin podatkov tudi za velika omrežja, ter opazovanja in eksperimenti, ki omogočajo natančnejšo obravnavo na nivoju manjših omrežij. Te metode so uporabne za zbiranje podatkov o popolnih omrežjih, vendar za področje osebnih omrežij zaradi specifičnih metodoloških značilnosti niso primerne.

V sociologiji so najpogosteje uporabljen vir ankete in vprašalniki, na katere odgovarjajo člani omrežij sami s pomočjo anketarjev. V zadnjem času se uporabljata tudi telefonsko anketiranje in anketiranje prek interneta.

3.1 Načini zbiranja podatkov

Tipična procedura za zbiranje podatkov o egocentričnih omrežjih je, da anketiranec (ego) s pomočjo enega ali več generatorjev imen, vprašanj s katerimi našteje imena članov omrežja, opredeli člane svojega omrežja (alterje). Več podatkov o alterjih navedenih v omrežje dobimo z dodatnimi vprašanji, t.i. interpretatorji imen. Burt opredeli 3 vrste interpretatorjev imen: prvi so namenjeni zbiranju podatkov o demografskih značilnostih alterjev (starost, izobrazba...), drugi sprašujejo po značilnostih vezi med egom in alterji (frekvenca kontaktov, trajanje znanstva, vsebina relacij...) in tretji opredeljujejo intenzivnost vezi med egom in alterji (Burt, 1984).



Slika 3.1: Egocentrično omrežje je sestavljeno iz ega in povezav do njegovih alterjev

Razvoj generatorjev imen je opredeljen s prvimi raziskavami na področju osebnih omrežij (Marsden, 1990). Laumann je v raziskavi leta 1973 uporabil generator imen na podlagi socialnih vlog. Anketirance je spraševal po treh najbližjih prijateljih, podobni generatorji sprašujejo po sodelavcih, sosedih in drugih specifičnih skupinah ljudi. Druga vrsta generatorjev imen deluje na podlagi afekcij. Wellman v raziskavi o "East Yorku" sprašuje po šestih osebah, ki ne živijo v istem gospodinjstvu in so anketirancu najbližje (Wellman, 1979).

Pomemben korak na področju generatorjev imen sta storila McAllister in Fisher, ki sta prva spraševala po izmenjavah med alterji in egi. V raziskavi sta uporabila več generatorjev imen, od katerih je vsak meril specifično vrsto izmenjav v omrežju. Teme merjenih relacij so bile druženje oziroma pogovori o hobijih, pogovor o osebnih problemih in izposoja večje količine denarja. Izmed naštetih alterjev so bile izločene odrasle osebe, ki so živele v istem gospodinjstvu kot ego, dodano pa je bilo še vprašanje o "drugih, ki so za ega pomembni" (McAllister in Fisher, 1978).

Prvi generatorji so bili kritizirani predvsem zaradi omejitve števila alterjev, ki naj bi jih anketiranci navedli, saj se v realnosti velikosti osebnih omrežij med ljudmi močno razlikujejo. McAllister in Fisher sta bila tudi prva, ki anketirancev nista omejila z največjim številom alterjev, ki so jih lahko našteali. Čas anketiranja je bil skrajšan tako, da

so na dodatna vprašanja o alterjih anketiranci odgovarjali le za prvih osem navedenih alterjev (Fisher, 1982).

Burt je v ameriškem popisu leta 1985 zaradi časovne stiske pri anketiranju uporabil en sam generator, ki je spraševal po osebah, s katerimi so se anketiranci v zadnjih šestih mesecih pogovarjali o zadevah, ki so bile zanje pomembne. Tudi tu največje število navedenih alterjev ni bilo omejeno, interpretator imen pa je bil uporabljen za prvih pet oseb, navedenih v omrežje (Burt, 1984).

Nadaljnji razvoj generatorjev imen je v največji meri potekal na področju raziskav o socialni opori. Večina generatorjev imen meri omrežja na podlagi izmenjav različnih razsežnosti socialne opore, to so: instrumentalna, informacijska, emocionalna opora in druženje ter na različnih kombinacijah naštetih pristopov.

3.2 Zbiranje podatkov in informacijska tehnologija

Z razvojem informacijske tehnologije so se za zbiranje podatkov na področju osebnih omrežij odprle tudi nove možnosti anketiranja. Velika večina podatkov o osebnih omrežjih je še danes zbrana osebno, z osebnim (face to face) anketiranjem. Kljub temu, da je anketiranje prek telefona postalo pomemben dejavnik pri zbiranju podatkov v družboslovju, se na področju osebnih omrežij zaradi zahtevnosti anketiranja dolgo ni uveljavilo.

Prva zbiranja podatkov o egocentričnih omrežjih, ki so potekala prek telefonov, so bila izvedena v 90-ih letih 20. stoletja. Eno pomembnejših primerjav in analiz na področju zbiranja podatkov osebno oz. prek telefona je opravila Tina Kogovšek v svojem doktoratu (Kogovšek, 2001), kjer je preverjala zanesljivost in veljavnost merjenja podatkov z različnimi anketnimi metodami. Ugotovila je, da je telefonsko anketiranje primerno za zbiranje tovrstnih podatkov in hkrati tudi precej cenejše od osebnega anketiranja.

Hiter razvoj interneta je vzpodbudil tudi številne raziskave na osnovi analize omrežij. Prve raziskave o egocentričnih omrežjih, ki so bile izvedene prek interneta, so potekale v okviru projekta RIS na Fakulteti za družbene vede v Ljubljani v letu 2001 (RIS, 2005).

3.3 Metodološki problemi zbiranja podatkov

Že večkrat omenjena specifičnost zbiranja podatkov o osebnih omrežjih je posledica relativno zapletenega in dolgega postopka anketiranja, sestavljenega iz več sklopov. Poleg zbiranja podatkov o omrežju vsakega anketiranca s pomočjo generatorja in interpretatorja imen, so v vsako anketo vključena tudi demografska vprašanja. Gledano v celoti, vsak anketiranec odgovarja na veliko število podobnih vprašanj, kar zanj predstavlja veliko breme. Posledica napornega postopka anketiranja je tako povečana nezanesljivost odgovarjanja, neodgovarjanje na posamezna vprašanja in končno izstopanje iz procesa anketiranja. Slednje je toliko bolj značilno za raziskave, ki potekajo prek interneta, kjer anketiranci na anketna vprašanja odgovarjajo brez pomoči anketarja (samoanketiranje – "self administered surveying"). Za raziskovalce neodgovori pomenijo težave pri analizi zbranih podatkov in seveda višje stroške raziskave.

4 Vzorčenje alterjev v egocentričnih omrežjih

4.1 Osnovna ideja vzorčenja

Ob iskanju rešitev na težave pri zbiranju podatkov, opredeljene v poglavju o težavah pri zbiranju podatkov je možnih nekaj rešitev, ki imajo na kvaliteto podatkov različne vplive:

Zmanjševanje števila vprašanj v celoti, kar vodi k zavestnemu zmanjševanju zbranih informacij in nezavidljivemu položaju raziskovalca, ki je z zmanjševanjem števila spremenljivk prisiljen okrniti svoj raziskovalni model.

Deljenje anketirancev v skupine tako, da vsako skupino vprašamo le del vprašanj o njihovih alterjih. Tak način je težko sprejemljiv, saj tako dobimo povsem ločene sklope podatkov, ki jih med seboj ne moremo primerjati.

Deljenje alterjev v skupine tako, da anketiranec za posamezno skupino alterjev odgovarja le na določen sklop vprašanj.

Vzorčenje alterjev vsakega anketiranca. Ta možnost nam da primerljive podatke o strukturi omrežij vseh anketirancev. Vsi anketiranci odgovarjajo na vsa vsebinska vprašanja, pri tem pa odgovarjajo le za alterje iz svojega omrežja, ki so bili izbrani v vzorec. Število alterjev, ki je izbrano v omrežje, je lahko določeno na različne načine. Lahko je določeno fiksno, lahko pa ga določimo kot delež velikosti omrežja anketiranca.

Ker se velikosti omrežij, ki jih anketiranci navedejo praviloma med seboj razlikujejo (velikosti variirajo med ena in mejo določeno s strani raziskovalca), je treba omeniti, da anketiranci, katerih velikost omrežja je manjša od določene fiksne velikosti vzorca, odgovarjajo na vprašanja za vse navedene alterje. Če je velikost določena kot delež, mora raziskovalec določiti prag velikosti omrežij, od katerega anketiranci z manjšimi omrežji odgovarjajo na vprašanja za vse alterje.

4.2 Koncept vzorčenja

Podatki o egocentričnih omrežjih so zbrani in urejeni na dveh stopnjah. Prvo stopnjo predstavljajo podatki o anketirancih (egih), to so demografski podatki, ter odgovori na vprašanja, ki so sestavni del vsake splošne ankete. Drugo stopnjo predstavljajo podatki o posameznih članih omrežij (alterjih) vsakega ega, ki jih v procesu analize agregiramo tako, da analiziramo značilnosti celotnih omrežij na nivoju egov.

Iz vidika vzorčenja, pri uveljavljeni metodi zbiranja podatkov o egocentričnih omrežjih, ni posebnosti. Anketiranci so v vzorec izbrani z eno izmed ustaljenih metod vzorčenja. V nalogi predpostavljamo, da gre za enostavni slučajni vzorec (Kalton, Vehovar 2001: 13).

V primeru predlagane metode zbiranja podatkov, vzorčimo alterje iz omrežij anketirancev (egov), ki sami predstavljajo reprezentativen vzorec raziskovane populacije. Opraviti imamo torej z dvostopenjskim vzorčenjem, kjer iz populacije A oseb z enostavnim slučajnim vzorčenjem izberemo a anketirancev. V drugem koraku, to je na drugi stopnji vzorčenja, pa iz omrežja vsakega anketiranca z enostavnim slučajnim vzorčenjem izberemo b izmed B članov omrežja.

Izračun osnovnih statistik, za dvostopenjsko vzorčenje je predstavljeno v knjigi Vzorčenje v anketah Kaltona in Vehovarja (Kalton, Vehovar 2001: 41), podrobna razčlenitev večstopenjskega vzorčenja pa v knjigi Survey Sampling Leslia Kisha (Kish, 1965: 154).

4.3 Implementacija vzorčenja in tehnične zahteve

Iz vidika anketiranca se z vzorčenjem ne spremeni veliko. Anketiranec na podlagi generatorja imen navede vse alterje v svoje omrežje. Število teh je lahko tudi neomejeno, saj za potek raziskave velika omrežja ne predstavljajo večjih ovir. Ko anketiranec navede imena, anketar s pomočjo računalnika izvede vzorčenje. Anketiranec nato odgovarja na dodatna vprašanja o posameznih naključno izbranih članih omrežja.

Za raziskovalca tovrsten način anketiranja predpostavlja uporabo računalnika. Metoda je tako uporabna le pri osebnem anketiranju s pomočjo računalnika (CAPI), pri telefonskem anketiranju s pomočjo računalnika (CATI) in samoanketiranju prek interneta.

4.4 Optimalna velikost podvzorca

Osnovno raziskovalno vprašanje v tej nalogi je kdaj in koliko, če sploh, alterjev je primerno vzorčiti glede na dano strukturo podatkov in stroške anketiranja. Ena izmed možnosti za določitev velikosti podvzorca je na podlagi stroškov anketiranja, ki so opredeljeni s stroškovno enačbo in intraklasne korelacije ρ , ki meri stopnjo podobnosti elementov v skupinah (glej Kish, 1965: 161 in Kalton, Vehovar 2001: 44). Omenjeno metodo izračuna optimalne velikosti omrežja je mogoče uporabiti le na že zbranih podatkih. Metodo smo uporabili, ker slednje v našem primeru ni predstavljalo težav, ter zato, ker smo z nalogo skušali dokazati smiselnost vzorčenja. V primeru realne raziskave bi moral raziskovalec opraviti pilotsko raziskavo, ter na tako zbranih podatkih predhodno testirati lastnosti oz. poiskati optimalno velikost podvzorcev. Optimalna velikost se med posameznimi spremenljivkami in velikostmi omrežij razlikuje, zaradi česar jo je potrebno izračunati na vseh spremenljivkah in velikostih omrežij, ter oceniti primerno velikost vzorčenih omrežij.

Stroškovno enačbo sestavljajo skupni stroški raziskave C , ki so sestavljeni iz stroškov za posamezni element, ki je zajet v raziskavo c (variabilni stroški – glede na število alterjev navedenih v omrežje) in iz stroškov za izbrano skupino C_a , v splošnem primeru za šolanje anketarja, prevozne stroške ipd., v našem primeru gre za strošek rekrutiranja posameznega anketiranca.

$$C = aC_a + nc \tag{1}$$

Ko govorimo o stroških raziskave poleg finančnih obremenitev raziskovalcev ne smemo prezreti tudi težje izmerljivega bremena, ki ga nosijo anketiranci. Stroškovna enačba se glede na način anketiranja seveda spreminja. Pomembna razlika med telefonskim anketiranjem in spletnim samo-anketiranjem je, da se pri slednjem praktično celoten strošek C_a prenese na anketiranca samega.

Poleg stroškov je pri izračunu optimalne velikosti vzorca na drugi stopnji vzorčenja pomemben dejavnik tudi koeficient intraklasne korelacije (intraclass correlation) ρ (ro), ki meri homogenost elementov znotraj skupin.

$$\rho = \frac{\frac{BS_a^2}{S^2} - 1}{B - 1}, \quad (2)$$

kjer S_a^2 predstavlja varianco med skupinami, B število enot v posamezni skupini in S^2 varianco na celotni populaciji. Nekoliko poenostavljeno lahko tudi rečemo, da je ρ delež elementarne variance, ki jo lahko pojasnimo s pripadnostjo določenemu omrežju.

Optimalno število alterjev na omrežje glede na stroške anketiranja je potem izračunano kot

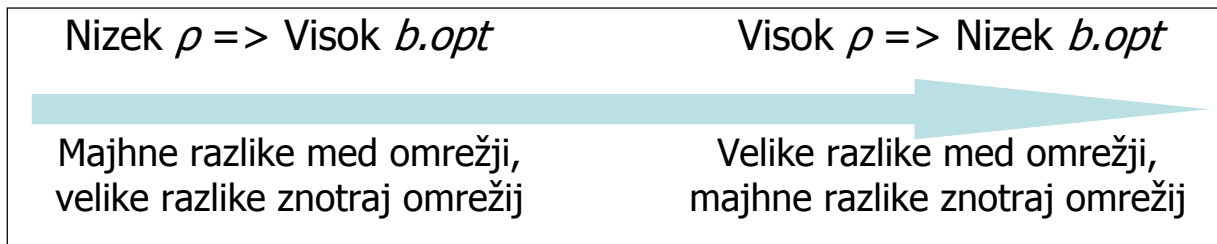
$$b.opt = \sqrt{\frac{C_a (1 - \rho)}{c \rho}} \quad (3)$$

Oglejmo si tabelo, kjer so predstavljene teoretično izračunane optimalne velikosti omrežij glede na različna razmerja variabilnih stroškov (c) in stroškov za vsako dodatno skupino (C_a), ter intraklasne korelacije (ρ).

Tabela 4.1: Optimalna velikost vzorčenih omrežij glede na razmerje stroškov in intraklasne korelacije.

C_a	c	C_a/c	ρ (intraklasna korelacija)						
			0	0,1	0,2	0,4	0,6	0,8	1
3	1	3	inf	5,2	3,46	2,12	1,41	0,87	0
3	3	1	inf	3	2	1,22	0,82	0,5	0
3	5	0,6	inf	2,32	1,55	0,95	0,63	0,39	0
5	1	5	inf	6,71	4,47	2,74	1,83	1,12	0
5	3	1,67	inf	3,87	2,58	1,58	1,05	0,65	0
5	5	1	inf	3	2	1,22	0,82	0,5	0
10	1	10	inf	9,49	6,32	3,87	2,58	1,58	0
10	3	3,33	inf	5,48	3,65	2,24	1,49	0,91	0
10	5	2	inf	4,24	2,83	1,73	1,15	0,71	0

Iz tabele je razvidno, da optimalna velikost vzorčenega omrežja glede na razmerje variabilnih stroškov (c) in stroškov za vsako dodatno skupino (C_a) z naraščanjem intraklasne korelacije (ρ) pada. Velik vpliv na optimalno velikost omrežja ima tako razmerje stroškov rekrutiranja posameznega ega in stroškov zbiranja podatkov o posameznem alterju v omrežju. Višji kot je strošek rekrutiranja, večja je optimalna velikost vzorčenega omrežja.



Slika 4.1: Razmerje med intraklasno korelacijo (ρ) in optimalno velikostjo omrežja.

Kot je razvidno iz enačbe (3) na optimalno vrednost vplivata varianca med skupinami in celotna varianca spremenljivke. Za nekatere spremenljivke, pri katerih je razmerje med S^2 in S_a^2 manjše od 1, je način določanja optimalne velikosti vzorčenega omrežja neprimeren. Dejansko gledano je zaradi strukture spremenljivk, ki so v večini primerov ordinalne in zaradi predpostavk pri izračunu variance, ocena le informativnega značaja. Izračun optimalne velikosti vzorčenega omrežja je tako narejen na vseh spremenljivkah razen na tistih, ki merijo strukturo omrežja in spol. Problem nizkega razmerja med variancama se zaradi majhnega števila podatkov občasno pojavi tudi pri drugih spremenljivkah.

Kako zmanjševanje količine podatkov vpliva na vsebinske rezultate pri vzorčenju teh spremenljivk bo razvidno iz nadaljnje analize v 5. poglavju.

Optimalna velikost omrežja za analizirane podatke, ki je uporabljena v nadaljevanju naloge, je izračunana za vsako velikost omrežja ter za vsako spremenljivko posebej in združena v povprečno vrednost. Na podatkih iz obeh raziskav je povprečna optimalna velikost omrežij pri danih stroških $C_a=5$ in $c=3$ enaka 3. Tabele vrednosti intraklasne korelacije ρ in optimalne velikosti vzorčenih omrežij so v prilogi A.

V nalogi je uporabljena fiksna velikost vzorca, kar se je med delom pokazalo za neprimerno odločitev. Rezultati so namreč pokazali, da so razlike med dodatno vzorčenimi in nevezorčenimi omrežji pri večjih omrežjih pogostejše. Sklepamo lahko, da je bilo v vzorec izbranih premalo enot, s katerimi bi lahko pojasnili značilnosti omrežij.

5 Analiza in rezultati

V nadaljevanju na podatkih iz dveh raziskav preverjamo pristop k zbiranju podatkov opisanem v naslednjem podpoglavju. Podatki za raziskavo o omrežjih Ljubljancev so zbrani leta 2000 s telefonsko anketo, podatki o omrežjih aktivnih uporabnikov interneta pa v letu 2001 s samoanketiranjem prek interneta. O omrežjih Ljubljancev so zbrani podatki za 1033, o uporabnikih interneta pa za 1009 oseb. Osnovni podatki o raziskavah so navedeni že v naslednjem podpoglavju naloge.

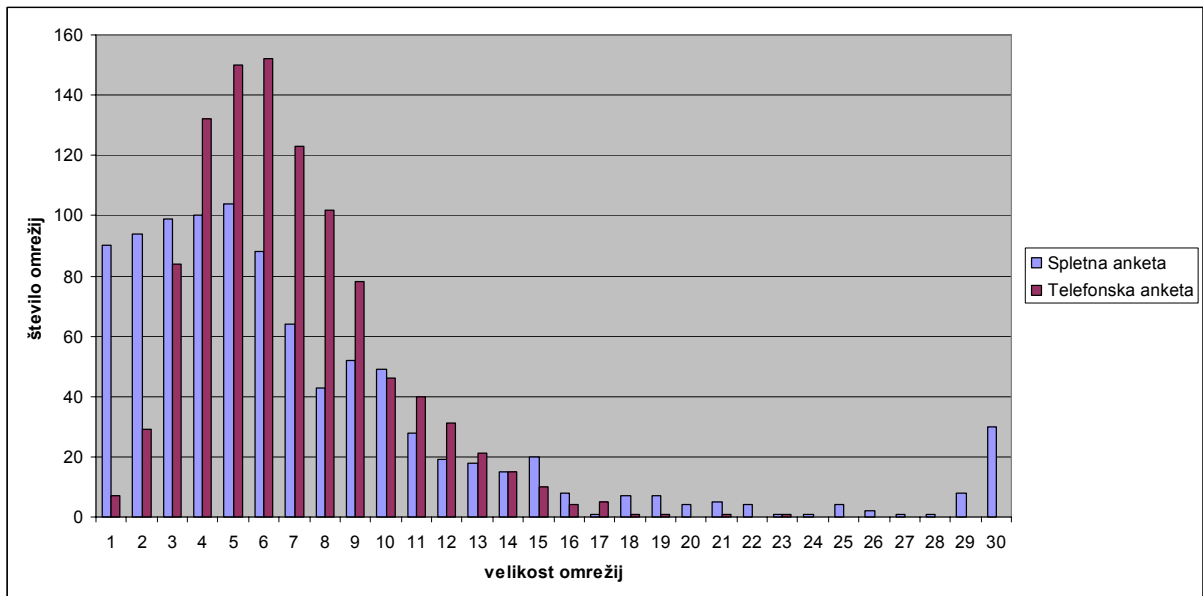
Poudariti velja, da so podatki zbrani po ustaljeni metodi zbiranja podatkov o egocentričnih omrežjih, torej, da so bili anketiranci o lastnostih članov omrežij (alterjev) vprašani za vse navedene alterje. Samo preverjanje je izvedeno tako, da iz vsakega omrežja izbrišemo podatke o alterjih, ki niso bili izbrani v naključni podvzorec. Povprečno optimalno velikost podvzorcev, 3 alterje, smo ocenili v poglavju 4.4. Vzorčene podatke nato primerjamo z nevtorčenimi.

Kriterij za primerjavo kakovosti podvtorčenih podatkov s podatki, ki niso podvtorčeni, je T-test za neodvisna vzorca. Slednji je izbran kljub precejšnjemu odstopanju spremenljivk od normalne porazdelitve, kar zaradi zadostnega števila enot ne predstavlja težav. Rezultati uporabljenega testa so bili kontrolirani z uporabo Mann-Whitney-evega neparametričnega testa razlik za dva neodvisna vzorca, ki je dal identične rezultate.

5.1 Opis podatkov

V nalogi so za preverjanje uporabnosti predlagane metode zbiranja podatkov uporabljeni podatki iz dveh raziskav. Prvi sklop podatkov je iz raziskave o omrežjih Ljubljancev, ki je bila izvedena v letu 2000. Raziskava je bila izvedena v sodelovanju Centra za metodologijo in informatiko Fakultete za družbene vede in podjetja CATI. Podatki so bili zbrani z računalniško podprto telefonsko anketo. V naključni vzorec je bilo vključenih 1033 anketirancev, od tega je 41% moških. Povprečna starost anketirancev je bila 44 let, povprečna velikost omrežja pa 6,8 alterjev. Največje omrežje je štelo 23 alterjev. Vsebinska osnova uporabljenega vprašalnika je bila socialna opora. Natančneje, uporabljenih je bilo pet generatorjev imen, ki merijo štiri razsežnosti socialne opore: materialno oporo (splošno), informacijsko oporo, druženje, emocionalno oporo in dodaten generator, ki meri finančno oporo.

Drug sklop podatkov je bil zbran v okviru spletne raziskave RIS v letu 2001. V raziskavi so sodelovali uporabniki interneta, ki so k raziskavi pristopili s t.i. samoizbiro na podlagi vabil, ki so jih prejeli prek elektronske pošte in oglasov na pomembnejših slovenskih spletnih straneh. V celotni raziskavi, ki je potekala med julijem in oktobrom 2001, je sodelovalo prek 14000 anketirancev. V analizo je vključenih 1009 aktivnih uporabnikov interneta, ki so bili naključno razdeljeni v štiri skupine, izmed katerih je vsaka dobila generator imen za eno izmed razsežnosti socialne opore. V vzorcu je 44% moških in 56% žensk. Povprečna starost anketirancev je 29,5 let, povprečna velikost omrežja v vzorcu je 7,6. Omrežij velikosti 30, kar je bila zgornja meja določena s strani raziskovalcev je v vzorcu 44 (4,6%), kar je posledica napačnega razumevanja generatorja imen. Pomembno je omeniti, da zaradi načina rekrutacije anketirancev rezultati raziskave niso primerni za posplošitev na celotno slovensko populacijo. Podrobnejši opis podatkov sledi v nadaljevanju.



Slika 5.1: Število omrežij glede na njihovo velikost.

Iz grafa na sliki **Error! Reference source not found.** Slika 5.1 lahko razberemo razporeditev števila omrežij glede na njihovo velikost. Razlike v številu članov omrežja med telefonsko in spletno anketo so posledica strukture anketirancev, splošne populacije v telefonski anketi in intenzivnih uporabnikov interneta v spletni anketi. Večje število anketirancev z velikostjo omrežja 30 v spletni anketi je posledica oblike vprašalnika (Hlebec in drugi, 2004).

5.2 Struktura omrežij

V tabeli Tabela 5.1 je predstavljena povprečna struktura omrežij Ljubljčanov. Največji delež v omrežjih predstavljajo sorodniki (44%) in prijatelji (44%). Manjši delež je sosedov (6%) in sodelavcev (3%). Pri vzorčenih podatkih, je opazna manjša, statistično neznačilna, razlika v deležu sorodnikov in prijateljev ($\pm 1\%$).

Tabela 5.1: Struktura omrežij – telefonska anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
sorodnik	44%	45%	0.89	0.37
prijatelj	44%	43%	0.23	0.82
sodelavec	3%	3%	0.23	0.82
sosed	6%	6%	0.84	0.41

Pregled strukture glede na velikost omrežij razkrije anomalije zaradi majhnega števila omrežij posameznih velikosti. V omrežjih do velikosti 15, kjer je količina podatkov za analizo zadostna, je z naraščanjem velikosti omrežij nakazano nižanje deleža sorodnikov in višanje deleža prijateljev v omrežjih. Delež sodelavcev z večanjem omrežij nekoliko narašča, s signifikantnimi skoki za velikosti omrežij 9 in 14. Delež sosedov je ne glede na velikost omrežja praktično nespremenjen. Razlike med dodatno vzorčenimi in nevezorčenimi podatki tudi pri primerjavi omrežij po velikosti ostajajo statistično neznačilne.

Tabela 5.2: Struktura omrežij glede na njihovo velikost – telefonska anketa (primerjava vzorčenih in nevezorčenih podatkov)

velikost omrežja	št. omrežij	sorodnik			prijatelj		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	68	51%	51%	1.00	36%	36%	1.00
4	121	47%	47%	0.96	42%	41%	0.76
5	145	46%	45%	0.60	43%	46%	0.50
6	149	46%	48%	0.60	44%	43%	0.58
7	123	47%	46%	0.66	42%	45%	0.38
8	102	43%	43%	0.97	44%	43%	0.83
9	78	43%	45%	0.66	40%	40%	1.00
10	46	34%	34%	0.99	50%	48%	0.71
11	40	36%	39%	0.54	50%	48%	0.70
12	30	41%	39%	0.74	46%	48%	0.68
13	20	41%	38%	0.74	50%	53%	0.65
14	15	36%	40%	0.64	41%	38%	0.71
15	10	43%	43%	0.94	44%	47%	0.80
16	4	33%	33%	0.97	34%	33%	0.94
17	5	39%	40%	0.93	44%	40%	0.81
18	1	33%	67%	0.31	53%	33%	0.55
19	1	0%	0%		71%	67%	0.90
21	1	53%	33%	0.56	41%	67%	0.44
23	1	17%	0%	0.08	78%	100%	0.04

velikost omrežja	št. omrežij	sodelavec			sosed		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	68	2%	2%	1.00	9%	9%	1.00
4	121	3%	4%	0.55	7%	8%	0.78
5	145	4%	4%	0.68	4%	4%	0.93
6	149	3%	3%	0.77	4%	4%	0.86
7	123	4%	4%	0.84	5%	5%	0.69
8	102	4%	3%	0.68	6%	8%	0.39
9	78	6%	3%	0.12	5%	6%	0.59
10	46	4%	4%	0.99	8%	11%	0.37
11	40	2%	4%	0.35	6%	5%	0.59
12	30	1%	1%	0.81	6%	6%	0.98
13	20	2%	2%	0.95	5%	7%	0.54
14	15	6%	7%	0.93	9%	11%	0.61
15	10	2%	3%	0.76	5%	3%	0.74
16	4	0%	0%		3%	8%	0.43
17	5	3%	7%	0.54	6%	7%	0.98
18	1	0%	0%		7%	0%	0.67
19	1	29%	33%	0.90	0%	0%	
21	1	0%	0%		0%	0%	
23	1	0%	0%		0%	0%	

V strukturi omrežij rednih uporabnikov interneta (Tabela 5.3: Struktura omrežij – spletna anketa) je delež sorodnikov manjši kot pri Ljubljancih (38%). Polovico omrežij sestavljajo prijatelji, delež sodelavcev je 13%, delež sosedov pa 5%. Razlike med dodatno vzorčenimi in nevtorčenimi podatki so pri spletni anketi statistično značilni.

Tabela 5.3: Struktura omrežij – spletna anketa

	nevtorčeni	vzorčeni	t-test	p (t-test)
sorodnik	38%	44%	5.70	0.00
prijatelj	50%	47%	2.10	0.04
sodelavec	13%	11%	2.43	0.01
sosed	5%	5%	0.97	0.32

Pregled razlik glede na velikost omrežij pokaže, da v manjših omrežjih uporabnikov interneta nadpovprečen delež omrežij predstavljajo sorodniki (približno 50%). Z večanjem omrežij ta delež pada. Nasprotno deleža prijateljev in sodelavcev z naraščanjem velikosti naraščata. Sosedje ne glede na velikost omrežij predstavljajo le majhen delež alterjev z izjemo omrežij velikosti med 19 in 21, kjer je delež sosedov med 15% in 22%.

Primerjava vzorčenih podatkov z nevtorčenimi glede na velikost omrežij jasno pokaže vzrok statistično značilnih razlik za vsa omrežja skupaj. Pri manjših omrežjih, ki so tudi pogostejša, razlike med vzorčenimi in nevtorčenimi podatki niso statistično značilne. Pri večjih omrežjih, ki se pojavljajo manj pogosto so razlike predvsem za nižje vrednosti pogostejše statistično značilne.

**Tabela 5.4: Struktura omrežij glede na njihovo velikost – spletna anketa
(primerjava vzorčenih in nevezorčenih podatkov)**

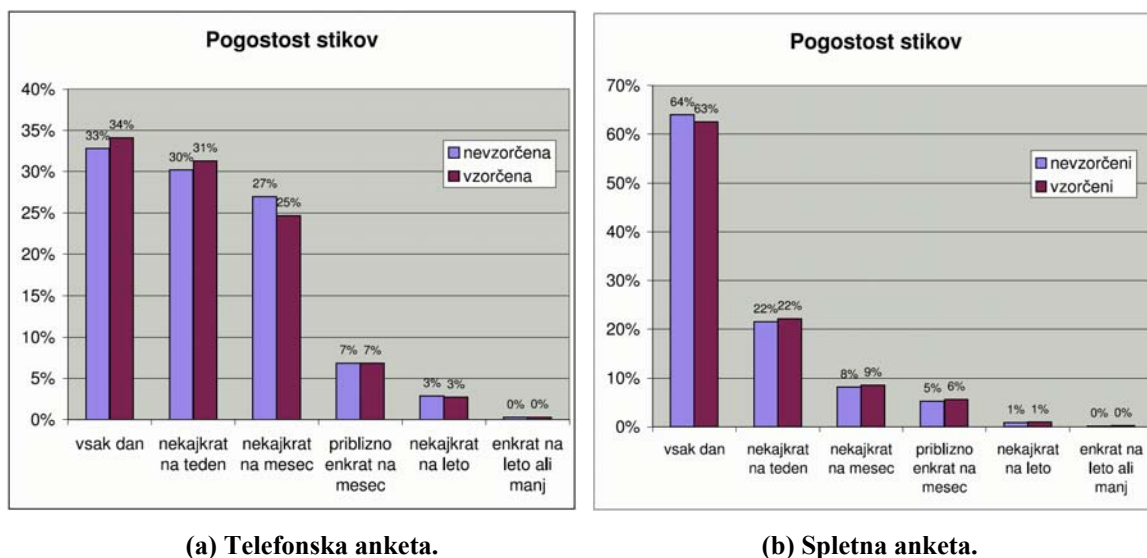
velikost omrežja	št. omrežij	sorodnik			prijatelj		
		nevzorčena	vzorčena	p (T-test)	nevzorčena	vzorčena	p (T-test)
3	98	56%	56%	0.93	42%	43%	0.94
4	96	54%	55%	0.93	41%	39%	0.62
5	101	55%	55%	0.85	42%	41%	0.78
6	92	43%	42%	0.84	53%	54%	0.87
7	61	43%	45%	0.57	49%	45%	0.42
8	46	43%	44%	0.84	43%	43%	0.95
9	51	41%	39%	0.71	52%	52%	1.00
10	37	35%	34%	0.85	57%	57%	0.99
11	30	40%	42%	0.74	52%	52%	0.96
12	25	31%	28%	0.64	52%	52%	0.98
13	20	27%	33%	0.33	56%	52%	0.59
14	15	33%	22%	0.12	55%	56%	0.98
15	19	38%	47%	0.19	57%	53%	0.59
16	9	22%	30%	0.37	66%	70%	0.67
17	2	42%	67%	0.43	24%	0%	0.00
18	6	16%	22%	0.50	74%	67%	0.52
19	7	26%	24%	0.83	65%	62%	0.78
20	4	24%	42%	0.28	54%	50%	0.78
21	4	18%	0%	0.00	58%	75%	0.24
22	5	19%	20%	0.89	64%	67%	0.84
23	2	24%	33%	0.74	62%	67%	0.88
24	2	27%	67%	0.17	12%	0%	0.54
25	4	21%	17%	0.74	56%	58%	0.89
26	2	15%	17%	0.94	38%	50%	0.60
28	1	5%	0%	0.70	53%	67%	0.67
30	44	27%	28%	0.77	47%	45%	0.74

velikost omrežja	št. omrežij	sodelavec			sosed		
		nevzorčena	vzorčena	p (T-test)	nevzorčena	vzorčena	p (T-test)
3	98	5%	5%	0.99	5%	5%	0.99
4	96	10%	9%	0.91	6%	6%	0.89
5	101	9%	8%	0.76	4%	4%	0.97
6	92	11%	10%	0.83	4%	4%	0.83
7	61	10%	16%	0.04	2%	1%	0.05
8	46	13%	13%	0.92	4%	4%	0.93
9	51	11%	12%	0.62	5%	3%	0.48
10	37	11%	12%	0.94	3%	4%	0.93
11	30	11%	10%	0.72	5%	8%	0.93
12	25	12%	16%	0.40	9%	11%	0.85
13	20	11%	13%	0.55	6%	5%	0.79
14	15	24%	29%	0.45	5%	4%	0.90
15	19	10%	7%	0.45	6%	5%	0.91
16	9	4%	0%	0.01	4%	0%	0.02
17	2	6%	0%	0.67	12%	0%	0.54
18	6	25%	28%	0.80	4%	0%	0.04
19	7	20%	24%	0.65	15%	14%	0.97
20	4	20%	8%	0.23	22%	17%	0.70
21	4	9%	8%	0.94	17%	25%	0.49
22	5	13%	27%	0.28	1%	0%	0.71
23	2	18%	0%	0.00	7%	0%	0.65
24	2	52%	0%	0.00	6%	0%	0.67
25	4	20%	8%	0.23	9%	8%	0.91
26	2	36%	17%	0.33	5%	0%	0.58
28	1	11%	0%	0.58	5%	0%	0.70
30	44	21%	23%	0.56	6%	8%	0.44

5.3 Druge značilnosti omrežij

Druge merjene značilnosti omrežij so *pogostost stikov*, moč povezave, ki je merjena z dvema indikatorjema: *blizina osebe* in *pomembnost osebe*. Negativne lastnosti vezi so merjene s spremenljivko *spravljanje v slabo voljo*. Med značilnosti omrežij sodijo še *trajanje znanstva*, *oddaljenost bivanja* in *spol* članov omrežij.

Ljubljancani, ki so sodelovali v anketi imajo stike (osebne, prek telefona, pošte ali spleta) s člani svojih omrežij v povprečju redkeje kot redni uporabniki interneta. Tretjina anketirancev spletne ankete ima stike z alterji vsakodnevno, nekoliko manj kot 60 odstotkov pa jih ima stike večkrat tedensko oz. mesečno. Med anketiranci iz spletne ankete je delež alterjev, s katerimi imajo anketiranci stike vsak dan, skoraj dve tretjini (64%), 30 odstotkov je takih, ki imajo stike večkrat tedensko oz. mesečno. Primerjava pogostosti stikov s člani omrežij glede na velikost omrežja, ki je predstavljena v prilogi B pokaže, da se povprečna pogostost stikov z velikostjo omrežja pri prebivalcih Ljubljane (telefonska anketa) nekoliko povečuje, med uporabniki interneta (spletna anketa) pa ostaja nespremenjena.



(a) Telefonska anketa.

(b) Spletna anketa.

Slika 5.2: Pogostost stikov

Tabela 5.5: Pogostost stikov – Telefonska anketa

	nevezorčeni	vzorčeni	t-test	p (t-test)
vsak dan	33%	34%	1.19	0.23
nekajkrat na teden	30%	31%	1.15	0.25
nekajkrat na mesec	27%	25%	2.39	0.02

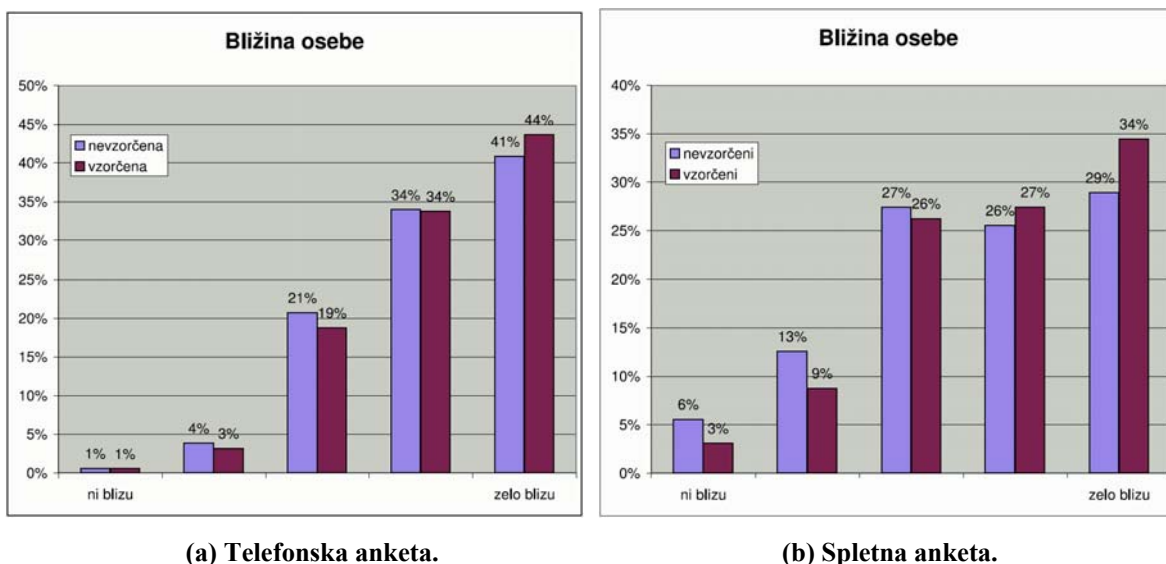
približno enkrat na mesec	7%	7%	0.00	1.00
nekajkrat na leto	3%	3%	0.40	0.69
enkrat na leto ali manj	0%	0%	0.28	0.78
n	5757	2871		

Tabela 5.6: Pogostost stikov – Spletna anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
vsak dan	64%	63%	1.22	0.22
nekajkrat na teden	22%	22%	0.61	0.54
nekajkrat na mesec	8%	9%	0.61	0.54
približno enkrat na mesec	5%	6%	0.60	0.55
nekajkrat na leto	1%	1%	0.00	1.00
enkrat na leto ali manj	0%	0%	0.64	0.52
n	7114	2334		

Razlike med dodatno vzorčenimi in nevzorčenimi podatki za vse velikosti omrežij razen za omrežje velikosti 23 v telefonski anketi, ki je eno samo, niso statistično značilne.

Bližina osebe je prvi izmed dveh indikatorjev, ki merita moč povezav med egi in alterji. Povprečna vrednost indikatorja v telefonski anketi je 4,1 in v spletni anketi 3,6, kar nakazuje, da so povezave v omrežjih Ljubljancanov močnejše kot v omrežjih uporabnikov interneta. V splošnem je povprečna vrednost indikatorja nad srednjo vrednostjo (3), torej so alterji egom blizu. V spletni anketi nekoliko izstopa srednja vrednost indikatorja, označilo jo je 27% anketirancev. Z večanjem števila članov v omrežjih¹ v obeh anketah povprečna bližina oseb v omrežjih pada.



Slika 5.3: Bližina osebe

Tabela 5.7: Bližina osebe – Telefonska anketa

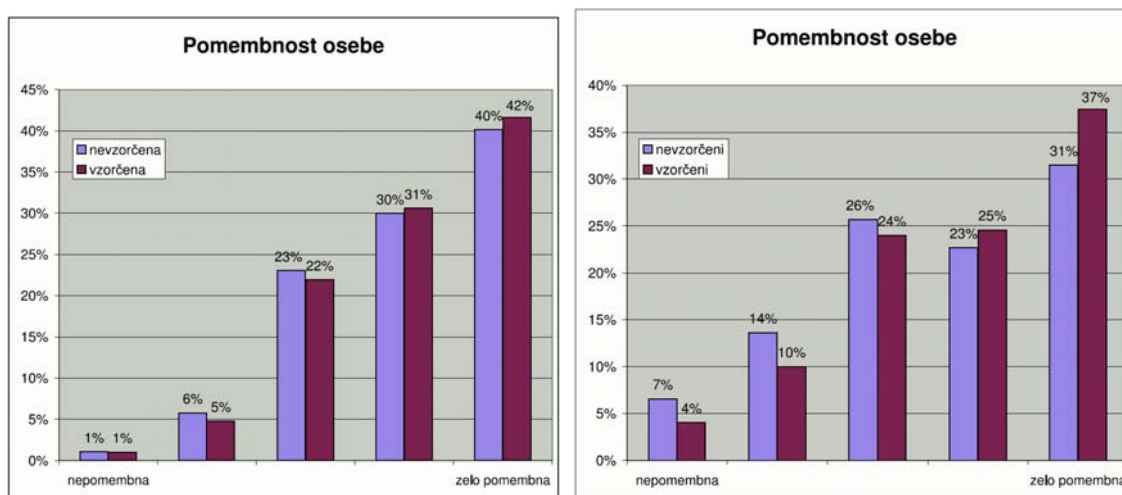
	nevzorčeni	vzorčeni	t-test	p (t-test)
ni blizu	1%	1%	0.19	0.85
	4%	3%	1.61	0.11
	21%	19%	2.05	0.04
	34%	34%	0.13	0.89
zelo blizu	41%	44%	2.39	0.02
n	5759	2871		

Tabela 5.8: Bližina osebe –Spletna anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
ni blizu	6%	3%	5.50	0.00
	13%	9%	5.34	0.00
	27%	26%	1.08	0.28
	26%	27%	1.77	0.08
zelo blizu	29%	34%	4.91	0.00
n	7062	2312		

Razlike v vrednostih te spremenljivke so med dodatno vzorčenimi in nevzorčenimi podatki statistično značilne. Pri podatkih, ki so ločeni glede na velikost omrežij so razlike značilne le pri omrežju velikosti 21 v telefonski anketi in omrežjih velikosti 23 in 24 v spletni anketi, kjer imamo le po eno oz. dve omrežji določene velikosti.

Lastnosti drugega indikatorja, ki meri moč povezav - pomembnost osebe, so praktično identične lastnostim prvega. Osebe v omrežjih rednih uporabnikov interneta so anketirancem nekoliko manj pomembne kot člani osebnih omrežij Ljubljančanov. Razlike med dodatno vzorčenimi in nevezorčenimi podatki so za vse velikosti omrežij skupaj statistično značilne, za podatke razdeljene glede na velikost omrežij pa ne.



(a) Telefonska anketa.

(b) Spletna anketa.

Slika 5.4: Pomembnost osebe

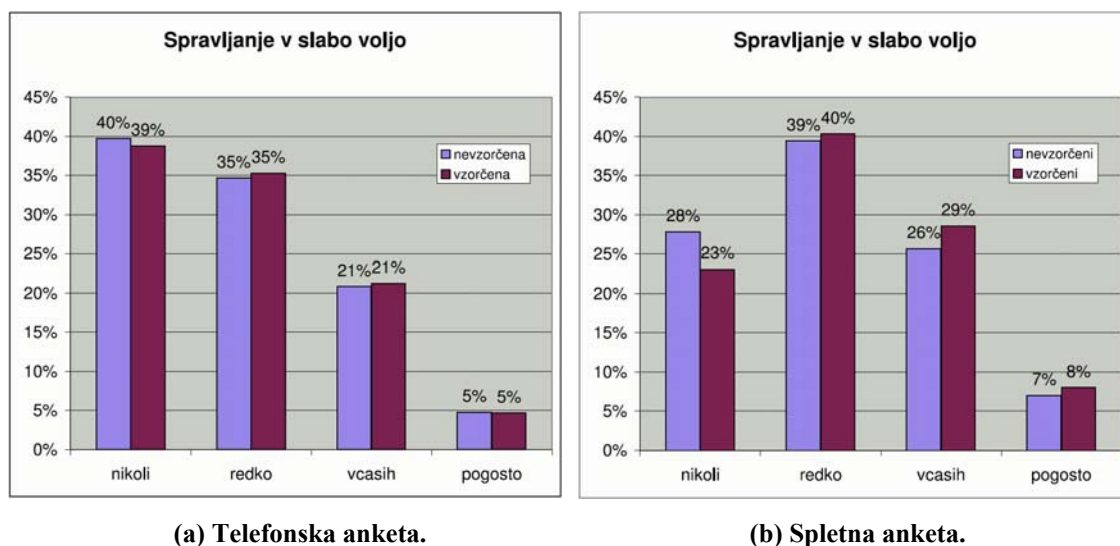
Tabela 5.9: Pomembnost osebe – Telefonska anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
nepomembna	1%	1%	0.21	0.83
	6%	5%	1.94	0.05
	23%	22%	1.07	0.29
	30%	31%	0.58	0.56
zelo pomembna	40%	42%	1.27	0.20
n	5757	2870		

Tabela 5.10: Pomembnost osebe – Spletna anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
nepomembna	7%	4%	5.07	0.00
	14%	10%	4.86	0.00
	26%	24%	1.66	0.10
	23%	25%	1.89	0.06
zelo pomembna	31%	37%	5.18	0.00
n	7058	2314		

Večina članov omrežij Ljubljancev (telefonska anketa) egov nikoli ne spravlja v slabo voljo (40%). Redko oz. včasih jih v slabo voljo spravlja skupno 56%. Takih, ki ege v slabo voljo spravlja pogosto, je 5%. V omrežjih rednih uporabnikov interneta zajetih v spletni anketi je alterjev, ki egov nikoli ne spravlja v slabo voljo manj (28%). Več je tistih, ki ege v slabo voljo spravlja redko (39%) oz. včasih (26%).



Slika 5.5: Spravljanje v slabo voljo

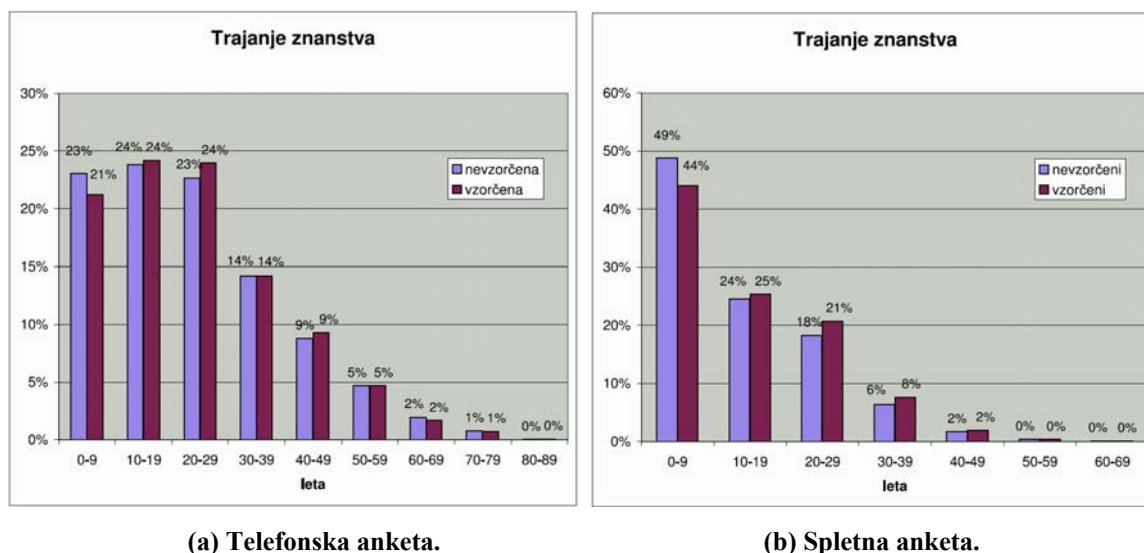
V primerjavi porazdelitve spremenljivke glede na velikosti omrežij² je nakazano nižanje povprečne vrednosti spremenljivke z večanjem velikosti omrežja kar pomeni, da je delež članov omrežij, ki ege spravlja v slabo voljo v velikih omrežjih manjše. Razlike med dodatno vzorčenimi in nevezorčenimi podatki niso statistično značilne le v telefonski anketi.

Tabela 5.11: Spravljanje v slabo voljo – Telefonska anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
nikoli	40%	39%	0.86	0.39
redko	34%	35%	0.76	0.45
včasih	21%	21%	0.28	0.78
pogosto	5%	5%	0.26	0.79
n	5749	2868		

Tabela 5.12: Spravljanje v slabo voljo – Spletna anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
nikoli	28%	23%	4.59	0.00
redko	39%	40%	0.75	0.45
včasih	26%	29%	2.69	0.01
pogosto	7%	8%	1.51	0.13
n	7056	2310		



Slika 5.6: Trajanje znanstva

Največje vsebinske razlike med raziskavama so v trajanju znanstva med anketiranci in člani njihovih omrežij. Povprečno trajanje znanstva v omrežjih Ljubljančanov je okoli 22 let. Znanstva rednih uporabnikov interneta so v povprečju stara 13 let. Zanimiva je tudi razlika v povprečnem trajanju znanstva glede na velikost omrežja. Medtem ko v telefonski anketi to z naraščanjem velikosti omrežja narašča, je povprečno trajanje znanstva v spletni anketi v velikih omrežjih krajše kot v majhnih. Pri Ljubljančanih nekoliko izstopajo omrežja velikosti 3, kjer je povprečno trajanje znanstva višje (25,1 let) kot pri drugih majhnih omrežjih.

Tabela 5.13: Trajanje znanstva– Telefonska anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
0-9	23%	21%	2.07	0.04
10-19	24%	24%	0.32	0.75
20-29	23%	24%	1.29	0.20
30-39	14%	14%	0.11	0.91
40-49	9%	9%	0.86	0.39
50-59	5%	5%	0.14	0.89
60-69	2%	2%	0.96	0.33
70-79	1%	1%	0.17	0.87
80-89	0%	0%	0.00	1.00
n	5764	2872		

Tabela 5.14: Trajanje znanstva– Spletna anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
0-9	49%	44%	3.72	0.00
10-19	24%	25%	0.75	0.45
20-29	18%	21%	2.45	0.01
30-39	6%	8%	1.73	0.08
40-49	2%	2%	0.71	0.48
50-59	0%	0%	0.28	0.78
60-69	0%	0%	0.44	0.66
n	6121	2002		

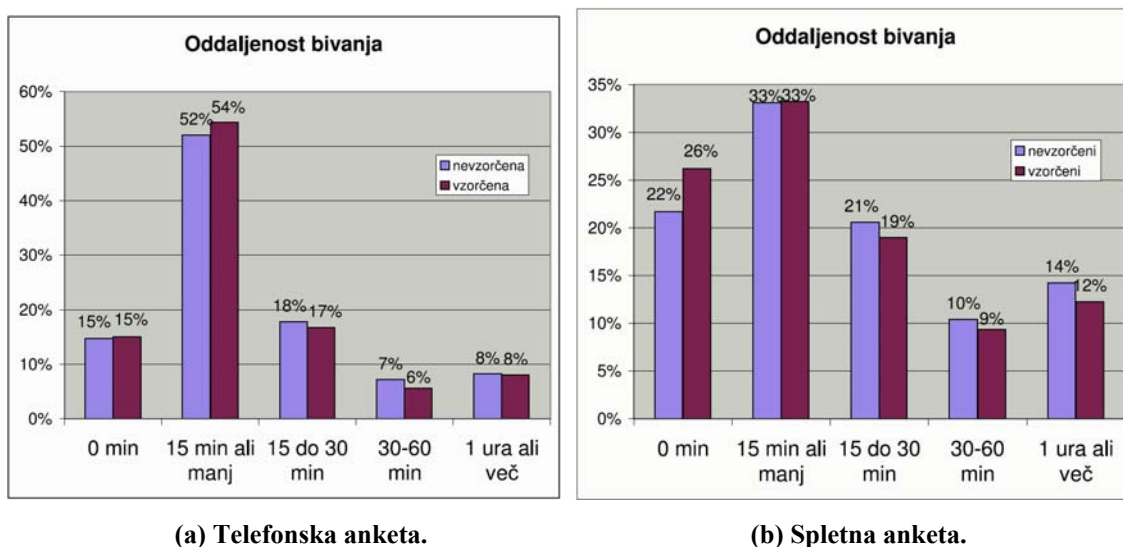
Tabela 5.15: Trajanje znanstva – Primerjava povprečij v letih

	nevzorčeni	vzorčeni	t-test	p (t-test)
Telefonska anketa	22.2	22.5	0.87	0.38
Spletna anketa	12.7	13.8	3.88	0.00

T-test razlik povprečij med dodatno vzorčenimi in nevzorčenimi podatki je pri tej spremenljivki za skupne podatke statistično značilen pri spletni anketi. Za posamezne velikosti omrežij pa le za omrežje velikosti 23 v telefonski anketi in omrežje velikosti 28 v spletni anketi, kjer so na voljo podatki le za po eno omrežje dane velikosti.

Tudi pri povprečni oddaljenosti bivanja članov omrežij od anketirancev se precej poznajo značilne razlike med merjenima populacijama. V telefonski anketi med prebivalci Ljubljane je tako več kot polovica članov omrežij anketirancev doma v krogu 15 minut vožnje z avtom. Oddaljenost bivanja 33% članov omrežij je več kot 15 minut. Struktura oddaljenosti bivanja članov omrežij pri uporabnikih interneta je precej bolj razpršena. V njihovih omrežjih člani bivajo v povprečju dlje kot v omrežjih Ljubljančanov. Takih, ki

bivajo v oddaljenosti 15 minut vožnje ali manj je tretjina. Zanimiva je primerjava članov omrežij, ki živijo v oddaljenosti 0 minut - to pomeni, da živijo v istem gospodinjstvu. V omrežjih uporabnikov interneta je odstotek takih, ki živijo v istem gospodinjstvu 22%, medtem ko je ta odstotek, v omrežjih Ljubljančanov le 15%. V omrežjih obeh populacij je odstotek takih, ki bivajo dlje v velikih omrežjih večji.



Slika 5.7: Oddaljenost bivanja

Tabela 5.16: Oddaljenost bivanja – Telefonska anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
0 min	15%	15%	0.33	0.74
15 min ali manj	52%	54%	2.19	0.03
15 do 30 min	18%	17%	1.31	0.19
30-60 min	7%	6%	2.66	0.01
1 ura ali več	8%	8%	0.27	0.79
n	5764	2874		

Tabela 5.17: Oddaljenost bivanja – Spletna anketa

	nevzorčeni	vzorčeni	t-test	p (t-test)
0 min	22%	26%	4.33	0.00
15 min ali manj	33%	33%	0.16	0.87
15 do 30 min	21%	19%	1.72	0.09
30-60 min	10%	9%	1.50	0.13
1 ura ali več	14%	12%	2.49	0.01
n	7011	2296		

Razlika med povprečnima vrednostma dodatno vzorčenih in nevtorčenih podatkov je za oddaljenost bivanja statistično značilna. Podobno kot pri ostalih spremenljivkah, so razlike manjše in neznačilne, če vzorec razdelimo glede na velikost omrežij.

5.4 Ugotovitve

Osnovna ugotovitev primerjave med vzorčenimi in nevtorčenimi podatki je, da se razlike pogosteje pojavljajo pri podatkih zbranih s spletno anketo. Razlike so opaznejše in statistično značilne predvsem pri večjih omrežjih.

Vzroki za nastale razlike so lahko različni, nedvomno pa se tudi dopolnjujejo:

- Razlike lahko pripišemo v povprečju večjim omrežjem, katerih značilnosti vzorec velikosti 3 ne zajame v celoti,
- omrežja anketirancev, ki so sodelovali v spletni anketi (O omrežjih uporabnikov interneta), so tudi vsebinsko bolj razpršena, saj gre za drugačno strukturo anketirancev,
- vpliv na vsebinsko strukturo omrežij ima lahko tudi način navajanja članov omrežij, v spletni anketi so anketiranci v omrežja naštevili zgolj na enem izmed štirih generatorjev imen, medtem ko so anketiranci v telefonski anketi omrežja opredeljevali na podlagi petih generatorjev imen,
- eden pomembnejših vzrokov je premajhno število podatkov – omrežij večjih od 15 alterjev je bilo le malo tako, kar nedvomno vpliva na reprezentativnost podvzorcev.

Vse navedene vzroke za nastale razlike je mogoče v bodoče odpraviti z različnimi načini izvedbe predlagane metode. Najenostavneje je to storiti z velikostjo podvzorca, ki sorazmerno narašča s številom navedenih članov v omrežje in manj radikalnim vzorčenjem za velikosti omrežij, ki se pojavljajo manj pogosto. Posledica omenjenih pristopov pa je ponovno povečevanje števila vprašanj za anketirance.

6 Sklep

Namen diplomske naloge je bilo preverjanje ideje o poenostavljenem zbiranju podatkov o osebnih omrežjih. Predlagana metoda naj bi potekala tako, da anketiranec potem, ko na podlagi generatorja imen navede člane omrežja na dodatna vprašanja o alterjih (interpretator imen), odgovarja le za nekaj naključno izbranih alterjev iz omrežja.

Cilj naloge je bil dokazati, da vzorčenje podatkov, ne vpliva na kvaliteto zajete informacije oz., da vzorčenje ne povzroči statistično značilnih razlik med vzorčenimi in nevzorčenimi omrežji.

Empirična analiza v nalogi je bila opravljena na podatkih dveh raziskav. Za prvo raziskavo, o egocentričnih omrežjih Ljubljančanov (FDV, CATI, 2000), so bili podatki zbrani s telefonsko anketo, za drugo, o omrežjih rednih uporabnikov interneta (RIS, 2001) pa s spletnim samoanketiranjem.

Pred samim vzorčenjem smo na podlagi intraklasne korelacije in stroškovne enačbe zbiranja podatkov, za podatke obeh raziskavi izračunali optimalno velikost podvzorca. Slednja se od spremenljivke do spremenljivke in glede na velikost omrežja razlikuje, zato smo za vzorčenje uporabili srednjo vrednost izračunanih optimalnih velikosti omrežij, tri alterje.

Uporabljena fiksna velikost podvzorca za vsa omrežja se je pri testiranju predlagane metode na podatkih izkazala kot neprimerna. Primerjava vzorčenih podatkov z nevzorčenimi je namreč pokazala, da predlagana metoda uspešno prestane test na manjših, gostejših omrežjih, kjer imamo na voljo zadostno število podatkov, medtem ko se pri večjih omrežjih, ki so tudi vsebinsko bolj razpršena pojavljajo statistično značilne razlike med vzorčenimi in nevzorčenimi podatki.

Nastali problem zaradi fiksne velikosti podvzorca je relativno enostavno rešljiv, saj lahko velikost podvzorca prilagajamo številu navedenih alterjev v omrežje. Drug vzrok za pojav razlik je težje obvladljiv, saj izvira iz majhnega števila velikih omrežij (glej Slika 5.1 **Error! Reference source not found.** **Error! Reference source not found.** **Error! Reference**

source not found.Error! Reference source not found.) iz katerih vzorčimo in v precep postavlja smiselnost podvzorčenja takih omrežij.

Ob zaključku lahko povemo, da smo z nalogo dokazali, da je predlagana metoda načeloma smiselna, izvedljiva in koristna. Izkazalo se je, da je z nekaj popravki je primerna tudi za uporabo v praksi. Pred resnejšo uporabo velja metodo še dodatno testirati na drugih realnih in simuliranih podatkih.

Konkretne rešitve v prihodnje lahko iščemo v izpopolnitvi merjenja razlik med vzorčenimi in nevzorčenimi omrežji, ter učinkovitejšem upoštevanju predvidenih stroškov pri računanju optimalnih velikosti vzorcev na prvem nivoju vzorčenja (izbira anketirancev v vzorec) in vzorčenju omrežij anketirancev. Med dodatnimi nalogami pa velja omeniti tudi analizo stroškovnih vidikov vpeljave metode v praksi, saj so izračuni pokazali, da se stroški zbiranja podatkov z uporabo metode znižajo tudi do 30 odstotkov.

7 Seznam literature

- Barnes, John (1954): "Class and Committees in a Norwegian Island Parish". *Human Relations*, 7, str. 39-58.
- Burt S., Ronald (1984): "Network Items and the General Social Survey". *Social Networks*, 6, 4, str. 293-339.
- Fisher, S., Claude (1982): *To Dwell among Friends*. The university of Chicago Press, Chicago and London.
- Hlebec, Valentina, Lozar Manfreda Katja, Vehovar, Vasja (2002): "Social Networks of Intensive Internet Users". prispevek na konferenci, <http://aoir.org/2002/programme.htm>.
- Kalton, Graham, Vehovar, Vasja (2001): *Vzorčenje v anketah*. Fakulteta za družbene vede, Ljubljana.
- Kish, Leslie (1965): *Survey Sampling*. John Wiley & Sons, Inc, New York, London, Sydney.
- Kogovšek, Tina (2001): "Ocenjevanje zanesljivosti in veljavnosti merjenja značilnosti egocentričnih socialnih omrežij: doktorska disertacija", Fakulteta za družbene vede, Ljubljana.
- Laumann, Edward (1973): *Bonds of Pluralism: The Forms and Substance of Urban Social Networks*. Wiley, New York.
- Marsden, V., Peter (1987) "Core Discussion Networks of Americans". *American Sociological Review*, 52, 1, str. 122-131.
- Marsden, V. Peter (1990): "Network data and measurement". *Annual Review of Sociology*, 16, str. 435 – 463.
- McAllister in Fisher (1978): "A Procedure for Surveying Social Networks". *Sociological Methods & Research*, 7, str. 131-148.
- RIS - Raba Internata v Sloveniji. <http://www.ris.org>.
- Scott, John (2000): *Social Network Analysis: A handbook*. Sage, London.
- Wellman, Barry (1979): "The Community Question: The Intimate Networks of East Yorkers". *American Journal of Sociology*, 84, 5, str. 1201-1231.

- Wellman, Barry (1988): "Structural Analysis: From Method and Metaphor to Theory and Substance". *Social Structures a Network Approach*, str. 19-61, Cambridge University Press, Cambridge.
- Wellman; Wortley (1990): "Different Strokes from Different Folks: Community Ties and Social Support". *American Journal of Sociology*, 96, 3, str. 558-588.

Priloga A: intraklasna korelacija ρ in izračunana optimalna velikost omrežja glede na stroške

Tabela A.1: Izračunana intraklasna korelacija po velikostih omrežij (spletna anketa)

A	B	Pogostost stikov	Bližina osebe	Pomembnost osebe	Spravljanje v slabo voljo	Trajanje znanstva	Spol	Oddaljenost bivanja
94	2	0.475	0.107	0.283	0.439	0.545	-0.275	0.192
99	3	0.390	0.053	0.088	0.099	0.302	-0.051	0.075
100	4	0.549	0.105	0.145	0.217	0.538	0.024	0.191
104	5	0.567	0.129	0.150	0.154	0.251	-0.030	0.222
88	6	0.535	0.166	0.233	0.144	0.300	-0.003	0.242
64	7	0.452	0.048	0.128	0.092	0.246	0.009	0.200
43	8	0.481	0.146	0.089	0.019	0.337	0.039	0.089
52	9	0.538	0.148	0.199	0.101	0.246	0.034	0.094
49	10	0.551	0.133	0.158	0.127	0.316	0.015	0.171
28	11	0.487	0.142	0.110	0.215	0.181	0.042	0.190
19	12	0.479	0.178	0.206	0.136	0.287	0.055	0.119
18	13	0.436	0.143	0.105	0.110	0.260	0.097	0.084
15	14	0.683	0.205	0.320	0.105	0.443	0.012	0.332
20	15	0.328	0.118	0.199	0.076	0.207	0.013	0.137
8	16	0.638	0.116	0.172	0.087	0.039	0.019	0.091
1	17	NA	NA	NA	NA	NA	NA	NA
7	18	0.655	0.221	0.173	0.269	0.499	0.110	0.050
7	19	0.204	0.071	0.039	-0.003	0.135	-0.008	0.180
4	20	0.438	-0.051	-0.015	0.039	0.295	0.116	0.228
5	21	0.156	0.112	0.139	0.020	0.070	-0.028	0.289
4	22	0.490	0.043	0.098	0.271	0.459	-0.042	0.129
1	23	NA	NA	NA	NA	NA	NA	NA
1	24	NA	NA	NA	NA	NA	NA	NA
4	25	0.431	0.334	0.022	0.000	0.098	0.038	0.265
2	26	-0.037	0.229	0.421	-0.039	0.095	-0.035	0.160
1	27	NA	NA	NA	NA	NA	NA	NA
1	28	NA	NA	NA	NA	NA	NA	NA
8	29	0.218	0.145	0.154	0.115	0.141	-0.020	0.097
30	30	0.484	0.158	0.178	0.148	0.204	0.028	0.162

Tabela A.2: Izračunana optimalna velikost omrežja glede na dano razmerje stroškov $C_a/c=5/3$ (spletna anketa)

A	B	Pogostost stikov	Bližina osebe	Pomembnost osebe	Spravljanje v slabo voljo	Trajanje znanstva	Spol	Oddaljenost bivanja
94	2	1.4	3.7	2.1	1.5	1.2	NaN	2.6
99	3	1.6	5.5	4.2	3.9	2.0	NaN	4.5
100	4	1.2	3.8	3.1	2.5	1.2	8.2	2.7
104	5	1.1	3.4	3.1	3.0	2.2	NaN	2.4
88	6	1.2	2.9	2.3	3.1	2.0	NaN	2.3
64	7	1.4	5.7	3.4	4.1	2.3	13.5	2.6
43	8	1.3	3.1	4.1	9.3	1.8	6.4	4.1
52	9	1.2	3.1	2.6	3.8	2.3	6.9	4.0
49	10	1.2	3.3	3.0	3.4	1.9	10.5	2.8
28	11	1.3	3.2	3.7	2.5	2.7	6.1	2.7
19	12	1.3	2.8	2.5	3.3	2.0	5.3	3.5
18	13	1.5	3.2	3.8	3.7	2.2	3.9	4.3
15	14	0.9	2.5	1.9	3.8	1.4	11.9	1.8
20	15	1.8	3.5	2.6	4.5	2.5	11.2	3.2
8	16	1.0	3.6	2.8	4.2	6.4	9.3	4.1
1	17	NA	NA	NA	NA	NA	NA	NA
7	18	0.9	2.4	2.8	2.1	1.3	3.7	5.6
7	19	2.5	4.7	6.4	NaN	3.3	NaN	2.8
4	20	1.5	NaN	NaN	6.4	2.0	3.6	2.4
5	21	3.0	3.6	3.2	9.0	4.7	NaN	2.0
4	22	1.3	6.1	3.9	2.1	1.4	NaN	3.4
1	23	NA	NA	NA	NA	NA	NA	NA
1	24	NA	NA	NA	NA	NA	NA	NA
4	25	1.5	1.8	8.5	NaN	3.9	6.5	2.1
2	26	NaN	2.4	1.5	NaN	4.0	NaN	3.0
1	27	NA	NA	NA	NA	NA	NA	NA
1	28	NA	NA	NA	NA	NA	NA	NA
8	29	2.4	3.1	3.0	3.6	3.2	NaN	3.9
30	30	1.3	3.0	2.8	3.1	2.5	7.5	2.9
Povprečna opt. velikost		1.5	3.5	3.4	3.9	2.5	7.6	3.2

Tabela A.3: Izračunana intraklasna korelacija po velikostih omrežij (telefonska anketa)

A	B	Pogostost stikov	Bližina osebe	Pomembnost osebe	Spravljanje v slabo voljo	Trajanje znanstva	Spol	Oddaljenost bivanja
58	2	0.307	0.428	0.483	0.312	0.546	-0.145	0.097
120	3	0.173	0.160	0.257	0.357	0.479	-0.018	0.064
192	4	0.113	0.172	0.210	0.241	0.357	-0.015	0.082
178	5	0.211	0.138	0.175	0.269	0.452	0.031	0.157
158	6	0.132	0.166	0.218	0.209	0.370	0.022	0.140
105	7	0.123	0.248	0.278	0.175	0.410	0.054	0.167
72	8	0.124	0.148	0.156	0.228	0.401	-0.008	0.134
57	9	0.190	0.182	0.162	0.267	0.446	0.019	0.161
26	10	0.126	0.191	0.081	0.143	0.411	0.047	0.150
19	11	0.154	0.145	0.140	0.102	0.539	0.037	0.037
9	12	0.085	0.279	0.237	0.162	0.482	0.017	0.262
4	13	0.188	-0.033	0.018	0.016	0.488	0.010	-0.048
10	14	0.198	0.224	0.178	0.260	0.384	0.047	0.085
5	15	0.168	0.203	0.145	0.420	0.333	0.017	0.083
2	16	0.066	0.033	0.093	0.137	0.077	-0.062	0.076
2	17	0.051	-0.049	-0.054	-0.023	0.099	-0.058	0.014
1	18	NA	NA	NA	NA	NA	NA	NA

**Tabela A.4: Izračunana optimalna velikost omrežja glede na dano razmerje stroškov $Ca/c=5/3$
(telefonska anketa)**

A	B	Pogostost stikov	Bližina osebe	Pomembnost osebe	Spravljanje v slabo voljo	Trajanje znanstva	Spol	Oddaljenost bivanja
58	2	1.9	1.5	1.3	1.9	1.2	NaN	3.9
120	3	2.8	3.0	2.2	1.7	1.3	NaN	4.9
192	4	3.6	2.8	2.5	2.3	1.7	NaN	4.3
178	5	2.5	3.2	2.8	2.1	1.4	7.2	3.0
158	6	3.3	2.9	2.4	2.5	1.7	8.6	3.2
105	7	3.4	2.2	2.1	2.8	1.6	5.4	2.9
72	8	3.4	3.1	3.0	2.4	1.6	NaN	3.3
57	9	2.7	2.7	2.9	2.1	1.4	9.4	2.9
26	10	3.4	2.7	4.3	3.2	1.5	5.8	3.1
19	11	3.0	3.1	3.2	3.8	1.2	6.6	6.6
9	12	4.2	2.1	2.3	2.9	1.3	9.8	2.2
4	13	2.7	NaN	9.6	10.2	1.3	13.2	NaN
10	14	2.6	2.4	2.8	2.2	1.6	5.8	4.2
5	15	2.9	2.6	3.1	1.5	1.8	9.8	4.3
2	16	4.9	7.0	4.0	3.2	4.5	NaN	4.5
2	17	5.6	NaN	NaN	NaN	3.9	NaN	11.0
1	18	NA	NA	NA	NA	NA	NA	NA
Povprečna opt. velikost		3.3	3.0	3.2	3.0	1.8	8.2	4.3

Priloga B: Lastnosti omrežij glede na njihovo velikost

Tabela B.1: Lastnosti omrežij glede na njihovo velikost – spletna anketa, 1.del
(primerjava vzorčenih in nevezorčenih podatkov)

velikost omrežja	št. omrežij	Pogostost stikov			Bližina osebe		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	98	1.6	1.6	1.00	4.1	4.1	1.00
4	96	1.7	1.7	0.79	4.0	4.0	0.92
5	101	1.7	1.6	0.83	4.1	4.1	0.98
6	92	1.7	1.7	0.75	3.9	3.9	0.55
7	61	1.6	1.6	0.70	3.9	4.0	0.20
8	46	1.5	1.4	0.37	3.8	3.8	0.69
9	51	1.6	1.7	0.72	3.7	3.5	0.23
10	37	1.6	1.6	0.66	3.7	3.6	0.43
11	30	1.8	1.7	0.59	3.7	3.7	0.74
12	25	1.7	1.7	0.75	4.0	3.9	0.86
13	20	1.5	1.5	0.91	3.5	3.7	0.31
14	15	1.8	1.8	1.00	3.6	3.3	0.08
15	19	1.3	1.4	0.32	3.4	3.6	0.14
16	9	1.9	2.0	0.67	3.3	3.4	0.56
17	2	1.2	1.3	0.54	3.8	4.7	0.10
18	6	2.0	2.1	0.67	3.0	3.1	0.76
19	7	1.4	1.2	0.27	3.2	3.1	0.92
20	4	1.7	1.6	0.72	3.0	3.4	0.30
21	4	1.6	1.8	0.62	3.2	2.9	0.57
22	5	1.3	1.1	0.06	3.3	3.1	0.70
23	2	2.3	1.3	0.07	3.4	5.0	0.00
24	2	1.1	1.0	0.67	2.7	5.0	0.00
25	4	1.7	1.8	0.66	2.7	2.7	0.90
26	2	1.7	1.8	0.69	3.1	3.2	0.95
28	1	1.8	1.3	0.33	2.7	3.0	0.82
30	44	1.4	1.4	0.84	3.0	3.0	0.64

velikost omrežja	št. omrežij	Pomembnost osebe			Spravljanje v slabo voljo		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	98	4.2	4.2	1.00	2.5	2.5	1.00
4	96	4.0	4.0	0.93	2.3	2.2	0.77
5	101	4.1	4.1	0.50	2.3	2.3	0.91
6	92	3.9	3.9	0.90	2.2	2.3	0.82
7	61	3.9	3.9	0.31	2.3	2.2	0.80
8	46	3.8	3.9	0.46	2.1	2.1	0.91
9	51	3.7	3.6	0.36	2.2	2.2	0.75
10	37	3.6	3.5	0.65	2.2	2.2	0.93
11	30	3.7	3.7	0.60	2.2	2.3	0.78
12	25	3.9	3.8	0.70	2.1	2.1	0.73
13	20	3.5	3.6	0.58	2.1	2.4	0.09
14	15	3.5	3.2	0.12	2.1	2.0	0.81
15	19	3.4	3.5	0.38	1.9	1.8	0.69
16	9	3.4	3.5	0.57	1.8	1.6	0.13
17	2	3.6	4.3	0.20	2.0	1.7	0.45
18	6	2.9	2.9	0.87	1.9	1.8	0.64
19	7	3.1	3.2	0.59	2.0	1.9	0.65
20	4	3.0	3.3	0.27	2.0	2.3	0.39
21	4	3.0	2.8	0.63	1.8	1.5	0.25
22	5	3.0	3.0	0.96	1.9	1.8	0.75
23	2	3.3	5.0	0.00	2.3	2.3	0.92
24	2	2.7	5.0	0.00	1.4	1.3	0.85
25	4	3.1	3.4	0.58	1.9	2.0	0.79
26	2	2.8	3.2	0.57	2.0	1.5	0.26
28	1	2.4	2.5	0.91	2.3	2.0	0.36
30	44	3.0	2.9	0.39	1.9	2.0	0.26

**Tabela B.2: Lastnosti omrežij glede na njihovo velikost – spletna anketa, 2.del
(primerjava vzorčenih in nevezorčenih podatkov)**

velikost omrežja	št. omrežij	Trajanje znanstva			Spol (delež moških)		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	98	15.2	15.2	1.00	0.4	0.4	1.00
4	96	16.7	16.9	0.85	0.4	0.4	0.96
5	101	16.3	16.5	0.79	0.4	0.4	0.85
6	92	13.6	13.4	0.86	0.5	0.4	0.51
7	61	13.5	13.3	0.82	0.4	0.4	0.96
8	46	12.8	11.9	0.42	0.5	0.5	0.74
9	51	13.3	12.7	0.53	0.5	0.5	0.92
10	37	12.3	12.2	0.92	0.4	0.4	0.34
11	30	12.2	13.3	0.45	0.4	0.5	0.33
12	25	12.7	11.9	0.57	0.5	0.5	0.82
13	20	11.1	11.8	0.61	0.4	0.4	0.76
14	15	11.5	10.4	0.57	0.5	0.7	0.00
15	19	10.2	9.4	0.54	0.5	0.4	0.11
16	9	8.6	8.8	0.89	0.5	0.5	0.86
17	2	9.5	10.7	0.81	0.5	0.5	0.97
18	6	9.4	10.0	0.76	0.4	0.4	0.64
19	7	11.6	14.3	0.29	0.5	0.4	0.65
20	4	11.6	13.4	0.49	0.4	0.2	0.04
21	4	12.2	10.4	0.62	0.6	0.7	0.65
22	5	13.2	14.2	0.80	0.5	0.5	0.92
23	2	11.0	8.7	0.61	0.5	0.3	0.73
24	2	8.5	15.0	0.20	0.6	0.3	0.43
25	4	12.0	14.5	0.46	0.5	0.8	0.15
26	2	5.7	2.5	0.20	0.4	0.2	0.30
28	1	8.9	20.0	0.00	0.3	0.7	0.26
30	44	11.1	11.3	0.82	0.5	0.4	0.05

velikost omrežja	št. omrežij	Oddaljenost bivanja		
		nevezorčena	vzorčena	p (T-test)
3	98	2.1	2.1	1.00
4	96	2.4	2.4	0.88
5	101	2.3	2.3	0.80
6	92	2.4	2.5	0.49
7	61	2.5	2.5	0.85
8	46	2.5	2.3	0.15
9	51	2.4	2.4	0.96
10	37	2.8	2.8	0.56
11	30	2.6	2.5	0.53
12	25	2.7	2.8	0.48
13	20	2.8	2.9	0.62
14	15	3.1	3.1	0.99
15	19	2.5	2.3	0.21
16	9	2.9	3.0	0.77
17	2	2.6	2.3	0.68
18	6	3.0	3.1	0.81
19	7	2.9	2.9	0.99
20	4	2.5	2.4	0.94
21	4	2.9	2.8	0.91
22	5	2.8	2.7	0.82
23	2	2.3	1.7	0.27
24	2	3.4	5.0	0.00
25	4	2.9	3.0	0.89
26	2	3.1	3.0	0.87
28	1	3.1	3.5	0.72
30	44	3.0	2.8	0.13

**Tabela B.3: Lastnosti omrežij glede na njihovo velikost – telefonska anketa, 1.del
(primerjava vzorčenih in nevezorčenih podatkov)**

velikost omrežja	št. omrežij	Pogostost stikov			Bližina osebe		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	68	2.1	2.1	1.00	4.2	4.2	1.00
4	121	1.9	1.9	0.54	4.3	4.3	0.96
5	145	2.1	2.1	0.77	4.2	4.2	0.56
6	149	2.1	2.0	0.52	4.2	4.3	0.46
7	123	2.1	2.2	0.60	4.1	4.1	0.84
8	102	2.1	2.1	0.98	4.1	4.1	0.98
9	78	2.3	2.3	0.72	4.1	4.2	0.17
10	46	2.5	2.6	0.42	3.9	3.9	0.58
11	40	2.3	2.4	0.60	4.0	4.1	0.53
12	30	2.2	2.2	0.88	4.0	4.0	0.85
13	20	2.1	2.0	0.70	4.2	4.2	0.73
14	15	2.4	2.4	0.96	4.0	4.1	0.39
15	10	2.7	2.5	0.41	4.1	4.1	0.76
16	4	2.8	3.0	0.69	3.3	3.3	0.84
17	5	2.4	2.5	0.71	3.9	3.8	0.72
18	1	2.3	2.7	0.43	3.8	4.7	0.29
19	1	1.3	1.3	0.90	4.7	4.7	0.90
21	1	2.1	2.0	0.85	3.4	3.0	0.03
23	1	2.0	3.3	0.05	3.4	3.3	0.80

velikost omrežja	št. omrežij	Pomembnost osebe			Spravljanje v slabo voljo		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	68	4.1	4.1	1.00	1.9	1.9	1.00
4	121	4.2	4.2	0.85	2.1	2.1	0.72
5	145	4.1	4.1	0.71	1.9	1.8	0.65
6	149	4.1	4.2	0.45	2.0	2.0	0.68
7	123	4.1	4.0	0.64	2.0	2.0	0.70
8	102	4.0	4.0	0.72	1.9	1.9	0.69
9	78	4.0	4.1	0.32	1.8	1.8	0.65
10	46	3.8	3.8	0.83	2.0	2.1	0.78
11	40	3.9	3.9	0.64	1.9	1.9	0.77
12	30	4.0	4.0	0.53	1.8	1.8	0.92
13	20	4.1	4.1	0.82	1.8	1.8	0.69
14	15	3.9	4.1	0.34	1.8	1.8	0.80
15	10	3.9	4.0	0.81	2.1	2.1	0.82
16	4	3.2	3.2	0.97	1.3	1.3	0.86
17	5	3.8	3.7	0.64	1.8	1.8	0.77
18	1	3.7	4.7	0.26	2.0	2.0	1.00
19	1	5.0	5.0		1.0	1.0	
21	1	3.5	3.3	0.66	1.5	1.3	0.76
23	1	3.2	2.7	0.43	1.8	1.0	0.00

**Tabela B.4: Lastnosti omrežij glede na njihovo velikost – telefonska anketa, 2.del
(primerjava vzorčenih in nevezorčenih podatkov)**

velikost omrežja	št. omrežij	Trajanje znanstva			Spol (delež moških)		
		nevezorčena	vzorčena	p (T-test)	nevezorčena	vzorčena	p (T-test)
3	68	25.1	25.1	1.00	0.6	0.6	1.00
4	121	22.2	22.3	0.93	0.5	0.5	0.65
5	145	22.9	22.5	0.72	0.5	0.5	0.48
6	149	20.6	21.3	0.40	0.5	0.5	0.62
7	123	22.1	23.3	0.22	0.5	0.5	0.93
8	102	21.0	21.3	0.74	0.5	0.5	0.52
9	78	24.2	25.1	0.45	0.5	0.5	0.30
10	46	20.8	21.4	0.73	0.5	0.5	0.90
11	40	20.6	20.5	0.96	0.5	0.5	0.63
12	30	19.7	20.9	0.50	0.5	0.5	0.90
13	20	24.7	23.9	0.73	0.5	0.5	0.63
14	15	24.4	21.8	0.31	0.5	0.5	0.78
15	10	26.0	24.1	0.55	0.4	0.3	0.28
16	4	29.8	27.7	0.71	0.5	0.6	0.44
17	5	25.0	28.7	0.49	0.5	0.3	0.35
18	1	10.3	17.3	0.12	0.4	0.3	0.87
19	1	13.7	14.0	0.97	1.0	1.0	
21	1	12.4	14.3	0.75	0.7	0.7	0.90
23	1	7.9	4.7	0.05	0.7	0.7	0.85

velikost omrežja	št. omrežij	Oddaljenost bivanja		
		nevezorčena	vzorčena	p (T-test)
3	68	2.2	2.2	1.00
4	121	2.2	2.1	0.70
5	145	2.3	2.3	0.90
6	149	2.4	2.4	0.89
7	123	2.4	2.4	0.90
8	102	2.5	2.4	0.81
9	78	2.5	2.6	0.95
10	46	2.6	2.6	0.79
11	40	2.4	2.4	0.99
12	30	2.5	2.4	0.42
13	20	2.5	2.3	0.24
14	15	2.5	2.5	0.78
15	10	2.8	2.8	0.96
16	4	2.7	2.3	0.24
17	5	2.7	2.8	0.67
18	1	2.3	2.3	0.89
19	1	2.4	2.7	0.54
21	1	3.5	3.7	0.79
23	1	2.6	3.0	0.02

Priloga C: Sintaksa R-project

```
rm(list=ls())

#####
#       Branje baze, dihromiziranje strukturne spremenljivke!       #
#####

alt<-read.table("lj_alter1.dat",header=TRUE, sep="\t", dec=",")
alt$TIEID<-NA;
alt<-(alt[c(3,2,1, 14:29)])

      alt[20]<-ifelse((alt[,10]==1),1,0)

      alt[21]<-ifelse((alt[,10]==2),1,0)
      alt[22]<-ifelse((alt[,10]==3),1,0)
      alt[23]<-ifelse((alt[,10]==4),1,0)
      alt[24]<-ifelse((alt[,10]==5),1,0)
      alt[25]<-ifelse((alt[,10]==6),1,0)
      alt[26]<-ifelse((alt[,10]==7),1,0)
      alt[27]<-ifelse((alt[,10]==8),1,0)
      alt[28]<-ifelse((alt[,10]==9),1,0)
      alt[29]<-ifelse((alt[,10]==10),1,0)
      alt[30]<-ifelse((alt[,10]==11),1,0)
alt<-alt[c(1:3,10,4,6,7,9,15,16,19, 20:30)]

      colnames(alt)<-c("ego_id", "alt_no", "alt_id", "NG", "V1", "V2", "V3", "V4", "V5", "V6",
"V7",'partner', 'starsi', 'bratje_sestre','otrok', 'drug_sorodnik', 'sodelavec', 'soclan', 'sosed', 'prijatelj',
'svetovalec', 'drugo' )
      alt[23:25]<-NA

tabela.b.opt<-numeric(0)
tabela.roh<-numeric(0)
tabela.podatki.S<-numeric(0)
tabela.podatki.Sa<-numeric(0)
tabela.podatki.Sb<-numeric(0)
tabela.podatki.deff<-numeric(0)

#####
#       !!!!!!! RAZČLENITEV BAZE PO VELIKOSTI OMREŽIJ!!!!!!       #
#####
#input je alt

for (d in 2:30){
so.B<-0;niso.B<-0
baza<-function (alt,d) {if(identical(length(alt[,2]),as.integer(d))) alt}

roar<-by(alt , alt[,1],baza , d )
alt1<-numeric(0);
for (ro in (1:length(roar))) { if(!identical(NULL,roar[[ro]])) (alt1<-rbind (alt1, as.matrix(roar[[ro]])) ) }
alt1<-data.frame(alt1); rownames(alt1)<-1:nrow(alt1)
rm(ro, baza, roar)

alt1<-as.matrix(alt1)
mode(alt1)<-"double"
#####
#       !! konec RAZČLENITEV BAZE PO VELIKOSTI OMREŽIJ!!!       #
```

```
#####
#output je alt1

je.mu<-by(alt1[,5:22],alt1[,1],function (x) {cbind(colMeans(x, na.rm=T))})
je.var<-by(alt1[,5:22],alt1[,1],function (x) {cbind(colVar(x))})

baz.mu<-numeric(0)
baz.var<-numeric(0)

for (i in 1:length(je.mu)) {baz.mu<-rbind (baz.mu, t(je.mu[[i]]))}
for (i in 1:length(je.var)) {baz.var<-rbind (baz.var, t(je.var[[i]]))}

B<-d
A<-nrow(baz.mu)#Kaj storiti z missingi
N<-nrow(alt1)#Kaj storiti z missingi

Sa<-apply (baz.mu, 2,var ,na.rm=T)
Sb<-apply (baz.var, 2,sum ,na.rm=T) /A
S<-colVar (alt1[,5:22])

#####
#                Izračun def, ro in b.opt                #
#                BODI POZOREN NA STROŠKE Ca in c                #
#####

Ca<-5
mali.c<-3

deff<-B*Sa*((A-1)/A)/S
rororo<-((deff-1)/(B-1))
b.opt<-sqrt((Ca/mali.c)*((1-rororo)/rororo))

tabela.roh<-rbind(tabela.roh, cbind (A,B, t(rororo)))
tabela.b.opt<-rbind(tabela.b.opt, cbind (A,B, t(b.opt)))
tabela.podatki.S<-rbind(tabela.podatki.S, cbind (A,B,N,t(S)))
tabela.podatki.Sa<-rbind(tabela.podatki.Sa, cbind (A,B,N,t(Sa)))
tabela.podatki.Sb<-rbind(tabela.podatki.Sb, cbind (A,B,N,t(Sb)))
tabela.podatki.deff<-rbind(tabela.podatki.deff, cbind (A,B,N,t(deff)))

print(B);print(date())}
tabela<-cbind( tabela.b.opt,tabela.roh)

write.table(tabela, file = "roh_b_opt"
, append = FALSE, quote = FALSE, sep = "\t",
eol = "\n", na = "NA", dec = ",", row.names = T,
col.names = TRUE )
```

```
#####
# !!!!!!! KOMPLET SKOLOP SRS VZORCENJE!!!!!! #
#####
#input je alt1, velikost omrežja

vel.vzorca<-3

#SRS funkcija

SRS<- function(alt1,n) {f<-as.vector(sample (alt1[,2],n))

for (r in 1:n)
{alt1<-t(apply(alt1, 1,function (a,fr){if(identical (as.integer(a[2]),as.integer(fr)))a[3]<-a[2];a},f[r]))}

# _____

#Vzorcenje

alt2<-by(alt1 , alt1[,1], SRS, vel.vzorca)

#Zasukas bazo in naredis data.frame
alt3<-numeric(0)
for (ro in (1:length(alt2))) alt3<-rbind (alt3, as.matrix(alt2[[ro]]))

alt3<-as.data.frame(alt3)

#####
#!!!!!! KONEC KOMPLET SKOLOPA SRS VZORCENJE!!!!!! #
#####
#output je alt3

#####
#iz baze vzemi samo v vzorec izbrane enote

alt4<-by(alt3[,1:25] , alt3[,1], function(alt3) {by(alt3, is.na(alt3[,3]), function (x) x<-x )[[1]])}
alt5<-numeric(0)
for (joj in 1:dim(alt4)) alt5<-rbind(alt5,alt4[[joj]])

nova.baza<-rbind(nova.baza, alt5)
#####
# IZRAČUNI #
#####

je.var<-by(alt3[,1:22] , alt3[,1], function(alt3) {by(alt3, is.na(alt3[,3]), colVar)[[1]])}
je.mu<-by(alt3[,1:22] , alt3[,1], function(alt3) {by(alt3, is.na(alt3[,3]), colMeans, na.rm=T)[[1]])}

JE.mu<-by(alt1[,5:22] ,alt1[,1],function (x) {(colMeans(x, na.rm=T))})
JE.var<-by(alt1[,5:22] ,alt1[,1],function (x) {(colVar(x))})

BAZ.mu<-numeric(0)
BAZ.var<-numeric(0)
baz.mu<-numeric(0)
baz.var<-numeric(0)

for (i in 1:length(JE.mu)) {BAZ.mu<-rbind (BAZ.mu, JE.mu[[i]])}
for (i in 1:length(JE.var)) {BAZ.var<-rbind (BAZ.var, JE.var[[i]])}
```



```
for (i in 1:length(je.mu)) {baz.mu<-rbind (baz.mu, je.mu[[i]]}  
for (i in 1:length(je.var)) {baz.var<-rbind (baz.var, je.var[[i]]}
```

```
baz.var<-baz.var[,5:22]  
baz.mu<-baz.mu[,5:22]
```

```
b<-vel.vzorca  
B<-d  
a<-nrow(baz.mu)  
A<-nrow(alt1)/d  
N<-nrow(alt1)
```

```
Sa<-apply (BAZ.mu, 2,var ,na.rm=T)  
Sb<-apply (BAZ.var, 2,sum ,na.rm=T) /a
```

```
sa<-apply (baz.mu, 2,var ,na.rm=T)  
sb<-apply (baz.var, 2,sum ,na.rm=T) /a  
S<-colVar (alt1[,5:22])
```

```
eSa<-sa-(1-b/B)*sb/b
```

```
var.cenilke<-sa/a- 1/A*(sa-(1-b/B)*sb/b)  
Var.cenilke<- (1-a/A)*Sa/a+(1-b/B)*Sb/(a*b)
```

```
tabela.V.srs<-rbind(tabela.V.srs, cbind (A,B,a,b, t(Var.cenilke)))  
tabela.v.srs<-rbind(tabela.v.srs, cbind (A,B,a,b, t(var.cenilke)))  
tabela.x.srs<-rbind(tabela.x.srs, cbind (A,B,a,b, t(colMeans(baz.mu,na.rm=T))))  
tabela.sa.var.srs<-rbind(tabela.sa.var.srs, cbind (A,B,a,b, t(sa/a)))
```

```
print(B);print(date())}
```

```
#tabela<-cbind( tabela.V,tabela.v)
```

