Универзитет у Београду

Електротехнички факултет



Анализа социјалних мрежа

Други пројектни задатак

|  |
| --- |
| Кандидати: |
| Марина Спасић, 2021/3209  Марко Станковић, 2021/3203 |

Београд, септембар 2022.

Садржај

[Садржај 2](#_Toc114654896)

[1. Увод 3](#_Toc114654897)

[2. Чишћење података 4](#_Toc114654898)

[3. Статистичка обрада података 5](#_Toc114654899)

[4. Мрежа SNet 8](#_Toc114654900)

[4.1. Основна карактеризације мреже 8](#_Toc114654901)

[4.2. Анализа мера централности 11](#_Toc114654902)

[4.3. Детекција комуна 13](#_Toc114654903)

[5. Мрежа SNetT 16](#_Toc114654904)

[5.1. Основна карактеризација мреже 16](#_Toc114654905)

[5.2. Анализа мера централности 18](#_Toc114654906)

[5.3. Детекција комуна 18](#_Toc114654907)

[6. Мрежа SNetF 21](#_Toc114654908)

[6.1. Основна карактеризација мреже 21](#_Toc114654909)

[6.2. Анализа мера централности 24](#_Toc114654910)

[6.3. Детекција комуна 25](#_Toc114654911)

[7. Мрежа UserNet 28](#_Toc114654912)

[7.1. Основна карактеризација мреже 28](#_Toc114654913)

[7.2. Основна карактеризација мреже 32](#_Toc114654914)

[7.3. Основна карактеризација мреже 33](#_Toc114654915)

[8. Поређење SNet и SNetT мрежа 35](#_Toc114654916)

1. Увод

Циљ другог пројектног задатка из предмета Анализа социјалних мрежа за школску 2021/2022. годину је практична примена теоријског знања представљеног на предмету на примеру конкретног истраживачког проблема. Тема пројектног задатка је анализа објава и коментара корисника друштвене мреже *Reddit* током 2008. године у циљу детаљније анализе друштвених дешавања у вези са битним догађајима исте године попут светске економске кризе. Кроз задати проблем приказано је прикупљање, обрада и прелиминарна анализа сировог скупа података. Из сирових података креиран је пречишћен скуп података на основу кога су моделиране социјалне мреже које су након тога анализиране алатима за обраду и визуализацију мрежа како би се добили одговори на постављена истраживачка питања.

За прикупљање, иницијалну обраду и прелиминарну анализу сировог скупа података коришћена је библиотека *pandas* програмског језика *Python.* *Python* библиотека *NetworkX* коришћена је за анализу моделиране мреже. Док су за процес визуализације коришћени алат *Gephi* и *Python* библиотека *matplotlib*.

1. Чишћење података

У овом поглављу се описује процес чишћења примарних (сирових) података. Сирови подаци доступни су у виду .csv фајлова у архиви која је дата у склопу поставке пројектног задатка. На основу приложених фајлова креиране су две табеле: табела са подацима о објавама и табела са подацима о коментарима корисника друштвене мреже *Reddit* из 2008. године. Како би се од примарног скупа података добила пречишћена форма података предузете су следеће акције:

* Обрисани су сви редови који имају обрисаног аутора означеног вредношћу *[deleted]*
* Уклоњене су све колоне табела које нису од интереса за постављена истраживачка питања. У табели са објавама уклоњене су колоне *distinguished*, *domain*, *stickied*, *locked* и *hide\_score* док су у табели са коментарима уклоњене колоне *distinguished*, *gilded* и *controversiality*. Обрисане су и безимене колоне, тј. прва колона у сваком .csv фајлу
* Постоји један коментар који нема одговарајући идентификатор, одлучено је да се тај ред не брише јер доприноси анализи касније моделоване мреже, а сам идентификатор коментара се не користи у даљој анализи
* Ради лакшег референцирања на коментарисану објаву или коментар додата је нова колона *link* која је представља *parent\_id* без префикса који се односи на тип коментара

1. Статистичка обрада података

У овоме поглављу биће приказана резултати статистичке обраде које је тражена кроз истраживачка питања.

1. Постоји 5032 различитих субредита. Најважнији сабредити по броју корисника су *reddit.com, politics, pics, technology, funny, entertainment, science, programming, business* и *worldnew*s. Најважнији сабредити по броју коментара су *reddit.com, politics, programming, pics, science, worldnews, WTF, funny, technology* и *business*.
2. Просечан број активних корисника по сабредиту у 2008. години износи око 148,668.
3. Корисници са највећим бројем постављених објава у посматраном периоду су gst, qgyh2, rmuser, twolf1, IAmperfectlyCalm, qazamisan, charlatan, igeldard, noname99 и democracy101. Корисници са највећим бројем постављених коментара су NoMoreNicksLeft, malcontent, matts2, mexicodoug, 7oby, aletoledo, mutatron, otakucode, redditcensoredme и h0dg3s.
4. Корисници који су активни на највећем броју различитих сабредита су MrKlaatu, Escafane, omfgninja, codepoet, scientologist2, turkourjurbs, Sylveran-01, b34nz, krugerlive и tuoder.
5. Број јединствених корисника који имају бар један објављени коментар је 88.573, док је број јединствених корисника који имају бар једну објављену објаву 185.355. Ако издвојимо број јединствених корисника који су и коментарисали и објављивали добијамо 47.279 корисника. Пирсонов коефицијент корелације броја објава и броја коментара износи 0,15. На основу наведених података можемо закључити да су наведана два податка слабо корелисана (нпр. корисници који су релативно пуно објављивали објаве нису постављали пуно коментара и обрнуто – Слика 3.1).

Chart, scatter chart

Description automatically generated

Слика 3.1 График корелације између броја постављених коментара (x-оса) и броја постављених објава корисника (y-оса)

1. Објаве са највећим бројем коментара приказане су у наставку:
2. Got six weeks? Try the hundred push ups training program

Број коментара: 33.329

Сабредит: science

*Линк: reddit.com/r/science/comments/6nz1k/got\_six\_weeks\_try\_the\_hundred\_push\_ups\_training/*

2) A black community in OH goes 50 years without running water...until one day, a white family moves in. Now, guess who has the only household on the street with running water?

Број коментара: 3.657

Сабредит: WTF

*Линк: reddit.com/r/WTF/comments/78n1v/a\_black\_community\_in\_oh\_goes\_50\_years\_without/*

3) Post the Funniest Joke You Know... Upmod the Best One

Број коментара: 2.039

Сабредит: reddit.com

*Линк: reddit.com/r/reddit.com/comments/675oj/post\_the\_funniest\_joke\_you\_know\_upmod\_the\_best\_one/*

4) Obama wins the Presidency!

Број коментара: 1.934

Сабредит: politics

*Линк: reddit.com/r/politics/comments/7beo2/obama\_wins\_the\_presidency/*

5) Palin Says Her Daughter, 17, Is Pregnant

Број коментара: 1.425

Сабредит: reddit.com

*Линк: reddit.com/r/reddit.com/comments/6z2e2/palin\_says\_her\_daughter\_17\_is\_pregnant/*

1. Мрежа SNet

Мрежа SNet је тежинска неусмерена мрежа добијена анализом активности корисника на различитим сабредитима. Чворови ове мреже су сви сабредити, док грана између два чвора постоји ако и само ако постоји бар један корисник који је био активан у погледу објављивања објава и/или коментара на оба форума. Тежина гране између два чвора добијена је простим бројањем укупног броја таквих активних корисника на оба сабредита. На овај начин креирана је мрежа од 4.359 чворова и 156.959 грана.

* 1. Основна карактеризације мреже

1. Густина овако добијене мреже износи приближно 0,0165.
2. Просечна дужина путање у мрежи износи приближно 2,0986, док је дијаметар мреже једнак 5.
3. Број повезаних компоненти унутар мреже је 814. Следи приказ величина компоненти у броју чворова: постоји једна гигантска компонента величине 3524, три компоненте величине 3, 16 компонената величине 2, док су остале компоненте величине 1. Гигантска компонента је издвојена зарад прецизнијег израчунавања каснијих резултата.
4. Просечни коефицијент кластеризације мреже износи приближно 0,625. Глобални коефицијент кластеризације мреже износи приближно 0,267. Расподела локалног коефицијента кластеризације мреже по чворовима приказана је на слици 4.1.1.

Chart

Description automatically generated

Слика 4.1.1 Расподела локалног коефицијента кластеризације по чворовима мреже

Да бисмо упоредили претходно приказану расподелу потребно је да генеришемо случајну, *Erdos-Renyi* мрежу истих димензија у погледу броја чворова и приближно исти број грана (задаје се вероватноћа креирања грана у виду густине SNet мреже). Расподела тако добијене мреже приказана је на слици 4.1.2.

Chart, histogram

Description automatically generated

Слика 4.1.2 Расподела локалног коефицијента кластеризације за случајну, *Erdos-Renyi* мрежу

На основу поређења претходно две расподеле мреже јасно се види да је кластеризација SNet мреже јако изражена јер највећи број чворова SNet мреже има далеко већи коефицијент кластеризације у односу на случајну, *Erdos-Renyi* мрежу.

1. С обзиром на веома мале вредности просечне дистанце (~2) и дијаметра (5), може се закључити да се до било ког чвора може доћи у веома малом броју корака, што нас доводи до закључка да SNet мрежа показује особине малог света. Узимајући у обзир и велики степен кластеризације чворова у језгру, закључујемо да иако смо на периферији, можемо преко јегра доћи до било ког другог чвора у свега неколико корака.
2. Коефицијент асортативности на основу нетежинског степена чвора износи приближно -0,4365 док коефицијент асортативности на основу тежинског степена чвора износи приближно -0,0323. Мрежа са коефицијентом 1 исказује савршено асортативно мешање, што се више крећемо ка нули, мрежа постаје више неасортативна или неутрална, како идемо ка −1 мрежа постаје све више дисасортативна. У случају нетежинског графа мрежа је више дисасортативна, односно, чворови сличног степена се ретко везују међусобно. У случају тежинског графа мрежа је неасортативна (неутрална). Слика 4.1.3 приказује степен-степен корелацију за дату мрежу.

Chart, scatter chart

Description automatically generated

Слика 4.1.3 Степен-степен корелација за мрежу SNet

1. Мрежа испољава особине феномена клуба богатих (слике 4.3.4 и 4.3.5).
2. На сликама 4.1.4 и 4.1.5 респективно је приказана нетежинска и тежинска расподела чворова по степену. Поређењем са другим дистрибуцијама (*power law*, *exponential*, *lognormal,* итд.) добијамо да мрежа прати најбоље *truncated power law* расподелу.

Chart, scatter chart

Description automatically generated

Слика 4.1.4 Нетежинска расподела чворова по степену у мрежи

Chart, scatter chart

Description automatically generated

Слика 4.1.5 Тежинска расподела чворова по степену у мрежи

1. С обзиром да је граф неусмерен, хабови и ауторитети представљају исту ствар и у том разматрању посматраћемо само степен чвора. Најбитнији чворови у мрежи су: *reddit.com* (3143), *technology* (2221), *politics* (2211), *pics* (2171), *funny* (2123), *science* (2105), *worldnews* (2091), *entertainment* (2061), *WTF* (2044) и *programming* (2043). Ексцентричност наведених најважнијих хабова износи 3, док дијаметар мреже износи 5, што указује да се они налазе ближе језгру графа јер се налазе приближно на половини дијаметра.
   1. Анализа мера централности
2. Најважнији чворови по централности по степену (нормализован максималним могућим степеном чвора) су следећи чворови: *reddit.com* (0,892137), *technology* (0,630429), *politics* (0,627590), *pics* (0,616236), *funny* (0,602611), *science* (0,597502), *worldnews* (0,593528), *entertainment* (0,585013), *WTF* (0,580187) и *programming* (0,579903).

Најважнији чворови по централности по блискости, која указује на то колико је чвор удаљен од других чворова у мрежи. Чворови који су централнији су бољи извори ширења информација и то су: *reddit.com* (0,901946), *technology* (0,728344), *politics* (0,726092), *pics* (0,720155), *funny* (0,712581), *science* (0,710139), *worldnews* (0,708425), *entertainment* (0,704037), *WTF* (0,701374) и *programming* (0,700816).

Релациона централност говори на колико се најкраћих путања између других чворова неки чвор налази. У случају социјалне мреже, актер са високом релационом централношћу контролоше ток информација и представља кандидата за брокерски чвор. У овој мрежи најважнији чворови по овој централности су *reddit.com* (0,240906)*, technology* (0,032263)*, programming* (0,028514)*, politics* (0,028131)*, business* (0,026983)*, entertainment* (0,024673)*, pics* (0,024575)*, worldnews* (0,021995)*, funny* (0,021117)и *science* (0,020857).

1. Централност по сопственом вектору говори о томе колико је утицајан чвор у мрежи. Чвор може бити утицајан ако има велики број утицајних суседа, међутим може бити утицајан и ако има мали број утицајних чворова (та метрика је овде прикладнија). Најважнији чворови по овој централности су: *reddit.com* (0,374686), *politics* (0,312672), *pics* (0,291855), *funny* (0,279569), *science* (0,273052), *technology* (0,263950), *worldnews* (0,258876), *WTF* (0,257430), *entertainment* (0,247074) и *programming* (0,220670).
2. Кацова централност са истим параметром бета за све чворове дала је следеће рангирање: *reddit.com* (0,274342), *politics* (0,230443), *pics* (0,216083), *funny* (0,207448), *science* (0,202978), *technology* (0,196527), *worldnews* (0,192860), *WTF* (0,192017), *entertainment* (0,184611) и *programming* (0,165985). Кацова централност са бета параметром постављеним на 0,1 за све сабредите осим за *reddit.com* коме је постављен на 1 даје следеће резултате: *reddit.com* (0,374045), *politics* (0,262970), *pics* (0,245130), *funny* (0,235010), *science* (0,229711), *technology* (0,222641), *worldnews* (0,218074), *WTF* (0,216820), *entertainment* (0,208948) и *programming* (0,187157). У претходно описаним случајевима алфа параметар износио је 0,000005. Ако алфа параметар поставимо на 0,00000025 добијају се релативно слични резултати: *reddit.com* (0,167726), *politics* (0,019152), *pics* (0,018866), *funny* (0,018753), *science* (0,018700), *technology* (0,018673), *worldnews* (0,018577), *WTF* (0,018554), *entertainment* (0,018545) и *business* (0,018334).
3. Композитна мера централности узима у обзир све претходно описане мере централности и добија се према следећој формули (за Кацову централност узима се други случај):

𝑐𝑜𝑚𝑝𝑜𝑠𝑖𝑡𝑒\_𝑟𝑎𝑛𝑘=𝐷𝐶\_𝑟𝑎𝑛𝑘∗𝐶𝐶\_𝑟𝑎𝑛𝑘∗𝐵𝐶\_𝑟𝑎𝑛𝑘∗𝐸𝑉𝐶\_𝑟𝑎𝑛𝑘\**Katz\_rank.*

Најважнији актери по композитној централности су *reddit.com, politics, technology, pics, funny, science, worldnews, programming , entertainment, WTF*.

* 1. Детекција комуна

1. Приказ сопствених вредности графа лапласијана гигантске компоненте мреже дат је на сликама 4.3.1 и 4.3.2. На основу спектралне анализе процењујемо да је добар избор за број комуна 2, 3, 5, 7 и 11.

Chart

Description automatically generated with medium confidence

Слика 4.3.1 Спектар лапласијанове гигантске компоненте

Chart, scatter chart

Description automatically generated

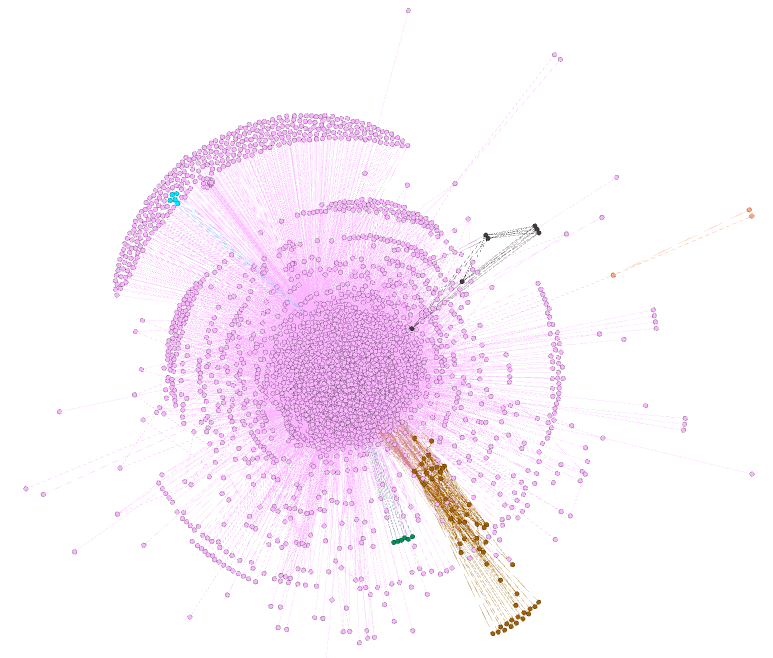
Слика 4.3.2 График првих 30 сопствених вредности графа лапласијана гигантске компоненте

1. и 22. Лувенска метода је спроведена за три вредности параметара. За параметар резолуције 0,1 креирано је 149 комуна, у којој постоји једна комуна са преко 600 чворова, и три комунер са преко 100 чворова, док су остале комуне значајно мање. За параметар резолуције 1,0 креирано је 54 комуне, где се истичу две велике комуне (преко 1600 и 1400 чворова), док су остале комуне значајно мање. За параметар резолуције 3,0 креирано је 6 комуна, где се истиче једна огромна комуна са скоро 3500 чворова, док су остале значајно мање. Резултати ове методе приказани су визуализацијом на сликама 4.3.3, 4.3.4 и 4.3.5. Визуализације најбитнијих чворова највећих комуна добијених Лувенском методом са параметром 1,0 приказане су на сликама 4.3.6 и 4.3.7.

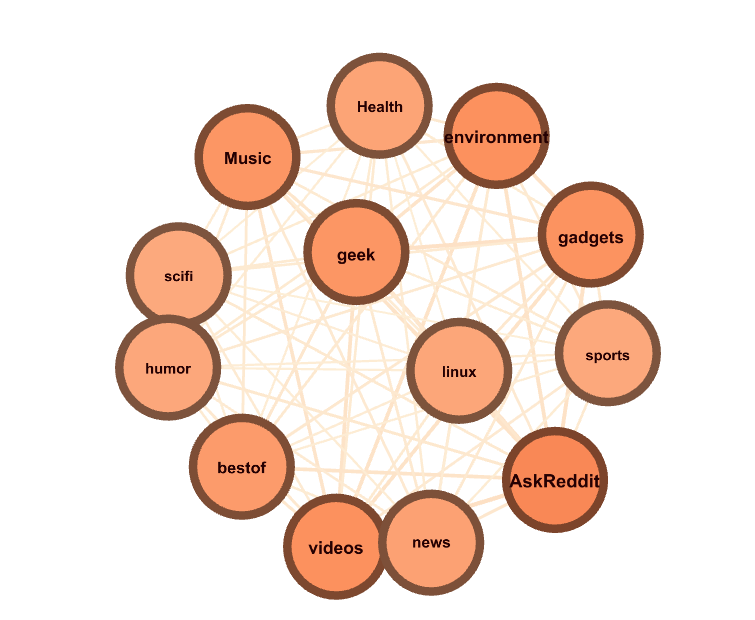
  
Слика 4.3.3 Визуелизација мреже након кластерисања Лувенском методом уз параметар резолуције 0,1



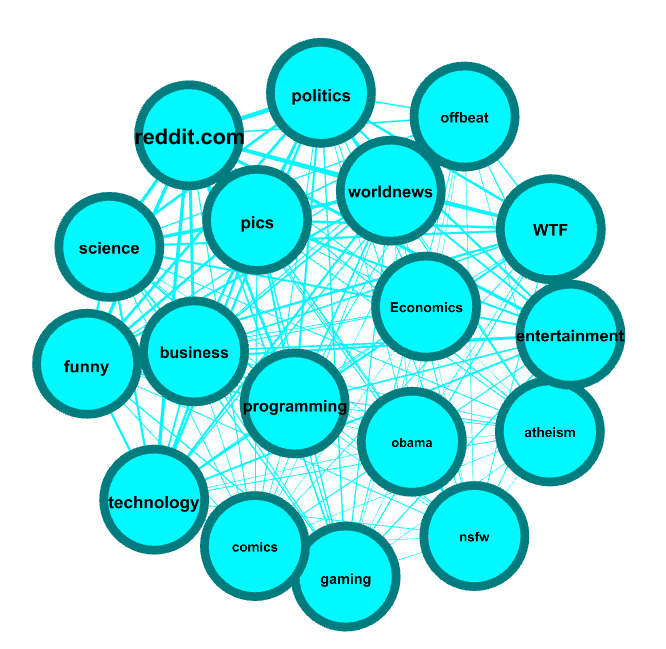
Слика 4.3.4 Визуелизација мреже након кластерисања Лувенском методом уз параметар резолуције 1,0



Слика 4.3.5 Визуелизација мреже након кластерисања Лувенском методом уз параметар резолуције 3,0



Слика 4.3.6 Мрежа најважнијих чворова прве комуне након кластерисања Лувенском методом уз параметар резолуције 1,0



Слика 4.3.7 Мрежа најважнијих чворова друге комуне након кластерисања Лувенском методом уз параметар резолуције 1,0

23. Чворови са највећом релационом централношћу су *reddit.com, technology, politics, pics, funny, science, worldnews, entertainment, WTF, programming* док се слични чворови појављују и у највећем број мостова (коришћена функција библиотеке *NetworkX*), а то су (*reddit.com*, 357), (*business*, 21), (*programming*, 15), (*Chinese*, 13), (*politics*, 9), (*technology*, 9), (*entertainment*, 8), (*worldnews*, 7), (*pics*, 7) и (*science*, 6).

1. Мрежа SNetT
   1. Основна карактеризација мреже

Мрежа SNetT има 39 чворова и 741 грану. Ова мрежа представља уску мрежу сабредита који су се по поставци задатка бавили темом светске економске кризе у 2008. години.

7 и 8. Густина, дијаметар и просечне дужина просечне путање у мрежи је 1 зато што је мрежа у потпуности повезана.

9. Број повезаних компоненти је 1. Мрежа је у потпуности повезана и постоји једна гигантска компонента од 39 чворова.

10. Просечни и глобални коефицијент кластеризације износе 1. Расподела локалног коефицијента кластеризације за случајно генерисану *Erdos-Renyi* мрежу истих димензија је иста као и расподела за посматрану мрежу и приказана је на слици 5.1.1. С обзиром на комплетно повезану мрежу, кластеризација је максимална.

Chart, histogram

Description automatically generated

Слика 5.1.1 Расподела локалног коефицијента кластеризације по чворовима мреже

11. Мрежа у потпуности исказује одлике малог света јер је сваки чвор директно спојен са свим другим чворовима у мрежи.

12. Мрежа је потпуно повезана па није могуће одредити нетежинску асортативност. Коефицијент асортативности на основу тежинског степена чвора износи приближно 0,0001. Степен-степен корелација је приказана на слици 5.1.2. Мрежа не показује одлике асортативности.

Chart

Description automatically generated

Слика 5.1.2 Степен-степен корелација за мрежу SNetT (тачка (39, 39) у горњем десном углу слике)

13. Мрежа не испољава феномен клуба богатих зато што је сваки чвор директно спојен са свим осталим чворовима.

14. Не може се фитовати ниједна расподела с обзиром да је мрежа потпуни граф. Слика 5.1.3 приказује дистрибуцију чворова мреже по тежинском степену.

Chart, scatter chart

Description automatically generated

Слика 5.1.3 Дистрибуција чворова мреже по тежинском степену

15. Сви чворови представљају хабове и ауторитете и чине језгро мреже.

* 1. Анализа мера централности

16. Сви чворови имају једнаке централности по степену, блискости и релационој централности, тако да можемо закључити да су сви чворови подједнако битни.

17. Чворове окружене мање утицајним чворовима можемо добити кроз посматрање централности по сопственом вектору. По овој метрици се издвајају следећи чворови: *reddit.com, politics, pics, funny, science, technology, worldnews, WTF, entertainment и programming.*

18. Кацова централност када сви чворови имају бета параметар једнак 1 даје следеће рангирање: *reddit.com* (0,361722), *politics* (0,304730), *pics* (0,285821), *funny* (0,274594), *science* (0,268861), *technology* (0,260549), *worldnews* (0,255887), *WTF* (0,254480), *entertainment* (0,244919) и *programming* (0,221450). Ако бета параметар поставимо чвору *reddit.com* на 1, док остале чворове поставимо на 0,1 добијамо следеће рангирање: *reddit.com* (0,453031), *politics* (0,299495), *pics* (0,278568), *funny* (0,267059), *science* (0,261091), *technology* (0,253347), *worldnews* (0,248100), *WTF* (0,246293), *entertainment* (0,237752) и *programming* (0,213624). У претходно описаним случајевима алфа параметар износио је 0,000005. Ако алфа параметар поставимо на 0,00000025 за претходно описан сценарио, добијају следећи резултати: *reddit.com* (0,842735), *politics* (0,095008), *pics* (0,093595), *funny* (0,093070), *science* (0,092815), *technology* (0,092714), *worldnews* (0,092276), *entertainment* (0,092142), *WTF* (0,092133) и *business* (0,091244).

19. Према композитној мери централности најважнији су чворови: *reddit.com, politics, pics, funny, science, technology, worldnews, WTF, entertainment* и *business*.

* 1. Детекција комуна

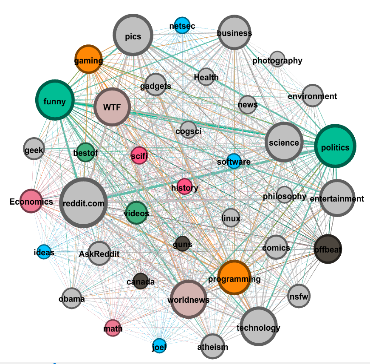
20. Приказ сопствених вредности графа лапласијана гигантске компоненте мреже дат је на слици 5.3.1. Потенцијални кандидати за број комуна у мрежи на основу спректралне анализе су: 5, 14, 17, 18, 28 и 29.

Chart, scatter chart

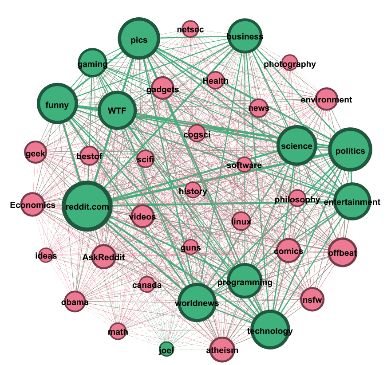
Description automatically generated

Слика 5.3.1. График целог спектра графа лапласијана гигантске компоненте мреже SNetT

1. и 22. Лувенска метода је урађена на три различите вредности параметра. За резолуцију 0,1 креирано је 17 комуна, благо се издваја један кластер са четири чвора. За резолуцију 1 креирано је 2 кластера са 26 и 13 чворова. За резолуцију 2 креиран је један кластер са 39 чворова. Резултати ове методе приказани су визуализацијом на сликама 5.3.2, 5.3.3 и 5.3.4. Визуализације најбитнијих чворова највећих комуна добијених Лувенском методом са параметром 1,0 приказане су на сликама 5.3.5 и 5.3.6.



Слика 5.3.2 Лувенска метода уз параметар резолуције 0,1

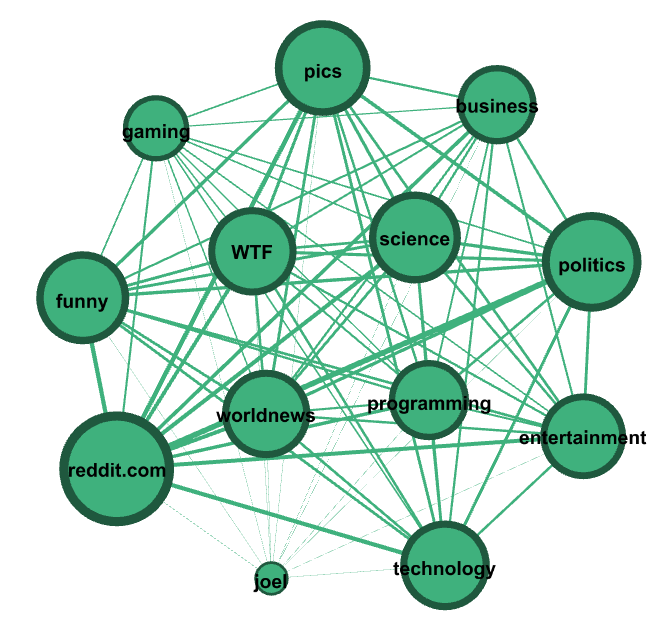


Слика 5.3.3 Лувенска метода уз параметар резолуције 1,0

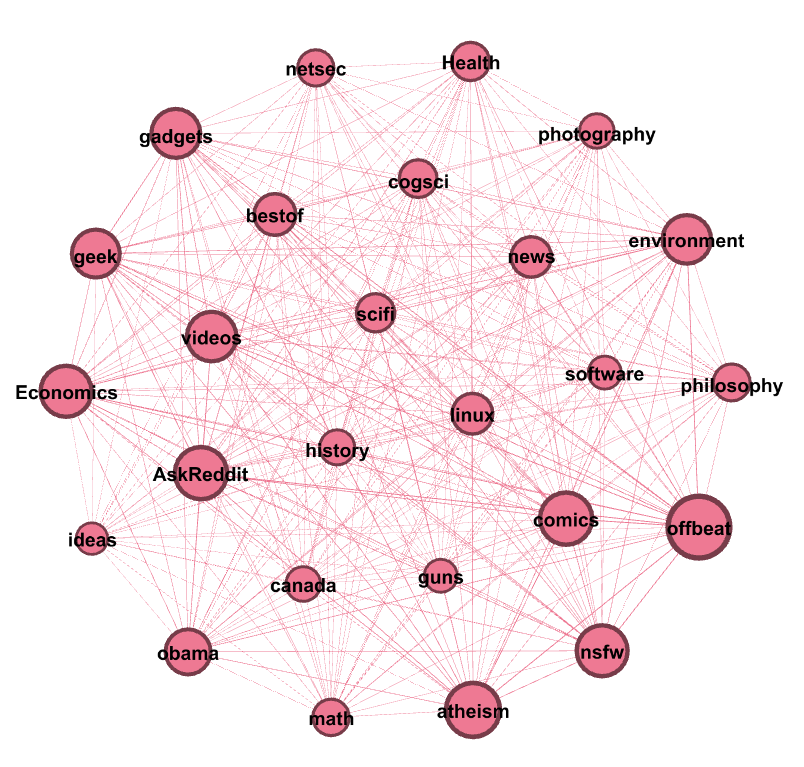
Chart, bubble chart

Description automatically generated

Слика 5.3.4 Лувенска метода уз параметар резолуције 2,0



Слика 5.3.5 Мрежа најважнијих чворова прве комуне након кластерисања Лувенском методом уз параметар резолуције 1,0



Слика 5.3.6 Мрежа најважнијих чворова друге комуне након кластерисања Лувенском методом уз параметар резолуције 1,0

23. С обзиром да је граф потпуно повезан и да сви чворови имају исту релациону централност, не могу се издвојити чворови који представљају брокере.

1. Мрежа SNetF

Мрежа SNetF је добијена је од SNet мреже тако што су уклоњене гране релативно мале тежине. У ту сврху одређен је праг вредности тежине 5, ако је тежина испод наведеног прага та грана се уклања и на тај начин добијамо мрежу SNetF. Добијена мрежа има 4359 и 30.227 грана.

* 1. Основна карактеризација мреже

7. Густина мреже износи приближно 0,00318.

8. Дијаметар мреже је 4, док је просечна дужина путање приближно 1,967.

9. Број повезаних компоненти је 3594 чворова. Постоји једна гигантска компонента величине 765 чворова, једна компонента од два чвора и остале имају само један чвор. Са овом доминантном компонентом ћемо радити у наставку анализе.

10. Просечни локални коефицијент кластеризације је око 0.853. Глобални коефицијент кластеризације је око 0,445. Расподела чвора по коефицијенту локалне кластеризације приказана је на слици 6.1.1. Расподела еквиваленте *Erdos-Renyi* мреже је приказана на слици 6.1.2.

Chart

Description automatically generated

Слика 6.1.1 Расподела чворова по коефицијенту локалне кластеризације

Chart

Description automatically generated

Слика 6.1.2 Расподела чворова по коефицијенту локалне кластеризације

На основу поређења приказаних расподела ове две мреже, можемо закључити да је кластеризација јако изражена.

11. С обзиром на веома мале вредности просечне дистанце (~2) и дијаметра (4), може се закључити да се до било ког чвора може доћи у релативно малом броју корака, што нас доводи до закључка да наша мрежа показује особине малог света.

12. Коефицијент асортативности на основу нетежинског степена чвора износи приближно -0,5945. Коефицијент асортативности на основу тежинског степена чвора износи приближно -0,0047. Степен-степен корелација чворова у мрежи дата је на слици 6.1.3. Из претходних података намеће се закључак да асортативно мешање није изражено.

Chart, scatter chart

Description automatically generated

Слика 6.1.3 Степен-степен корелација мреже SNetF

13. Мрежа испољава особине феномена клуба богатих (слике ).

14. Нетежинска и тежинска расподела чворова по степену у мрежи приказана је на сликама 6.1.4. и 6.1.5 респективно. Поређењем са другим дистрибуцијама (*power law*, *exponential*, *lognormal,* итд.) добијамо да мрежа прати најбоље *truncated power law* расподелу.

Chart, scatter chart

Description automatically generated

Слика 6.1.4 Нетежинска расподела чворова по степену у мрежи

Chart, scatter chart

Description automatically generated

Слика 6.1.5 Нетежинска расподела чворова по степену у мрежи

15. С обзиром да је граф неусмерен, хабови и ауторитети представљају исту ствар и посматра се само степен чвора. Најбитнији хабови су следећи сабредити: *reddit.com* (724), *politics* (643), *pics* (632), *science* (630), *technology* (627), *worldnews* (618), *funny* (617), *WTF* (599), *entertainment* (586) и *programming* (572). Ексцентричност најважнијих хабова износи 3, док дијаметар мреже износи 4, што указује да се они налазе негде између периферије и језгра мреже.

* 1. Анализа мера централности

16. Анализом централности по степену најважнији чворови су *reddit.com* (0,947644), *politics* (0,841623), *pics* (0,827225), *science* (0,824607), *technology* (0,820681), *worldnews* (0,808901), *funny* (0,807592), *WTF* (0,784031), *entertainment* (0,767016) и *programming* (0,748691). Анализом централности по блискости најважнији чворови су *reddit.com* (0,950249), *politics* (0,849833), *pics* (0,847947), *technology* (0,844199), *science* (0,838639), *worldnews* (0,835886), *funny* (0,834061), *WTF* (0,808466), *entertainment* (0,807611) и *programming* (0,795833). Анализом релационе централности најважнији чворови су *reddit.com* (0,191220), *politics* (0,060963), *pics* (0,054266), *technology* (0,051593), *science* (0,051344), *programming* (0,049073), *worldnews* (0,048672), *funny* (0,047436), *WTF* (0,041516) и *entertainment* (0,036209).

17. Анализом централности по сопственом вектору најважнији чворови (окружени мање утицајним чворовима) су *reddit.com* (0,374688), *politics* (0,312674), *pics* (0,291857), *funny* (0,279571), *science* (0,273053), *technology* (0,263952), *worldnews* (0,258878), *WTF* (0,257432), *entertainment* (0,247075) и *programming* (0,220671).

18. Кацова централност када сви чворови имају бета параметар једнак 1 даје следеће рангирање: *reddit.com* (0,332980)*, politics* (0,279740)*, pics* (0,262311)*, funny* (0,251832)*, science* (0,246408)*, technology* (0,238564)*, worldnews* (0,234124)*, WTF* (0,233103)*, entertainment* (0,224110)*, programming* (0,201497). Ако бета параметар поставимо чвору *reddit.com* на 1, док остале чворове поставимо на 0,1 добијамо следеће рангирање: *reddit.com* (0,411489), *politics* (0,289049), *pics* (0,269432), *funny* (0,258308), *science* (0,252484), *technology* (0,244706), *worldnews* (0,239692), *WTF* (0,238313), *entertainment* (0,229661) и *programming* (0,205704). У претходно описаним случајевима алфа параметар износио је 0,000005. Ако алфа параметар поставимо на 0,00000025 за претходно описан сценарио, добијају следећи резултати: *reddit.com* (0,342299), *politics* (0,039068), *pics* (0,038483), *funny* (0,038254), *science* (0,038146), *technology* (0,038089), *worldnews* (0,037895), *WTF* (0,037849), *entertainment* (0,037829) и *business* (0,037400).

19. Већ наведена композитна мера централности издваја следеће најважније актере: *reddit.com, politics, pics, science, technology, funny, worldnews, WTF, programming* и *entertainment*.

* 1. Детекција комуна

20. Посматрамо граф лапласијана над доминатном компонентом приказан на сликама 6.3.1 и 6.3.2. На основу спектралне анализе процењујемо да је добар избор за број комуна 1, 3, 5 и 28.

Chart

Description automatically generated

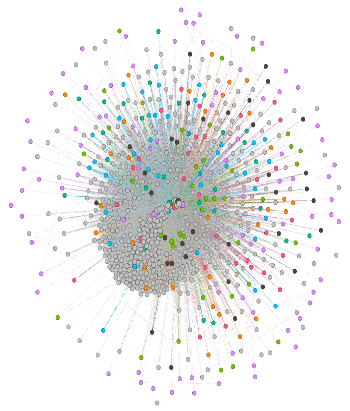
Слика 6.3.1 График целог спектра графа лапласијана гигантске компоненте мреже SNetF

Chart, scatter chart

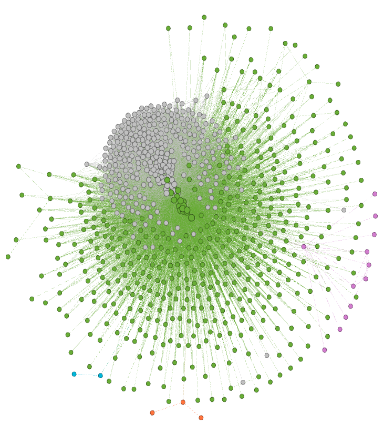
Description automatically generated

Слика 6.3.2 График првих 30 сопствених вредности графа лапласијана гигантске компоненте мреже SNetF

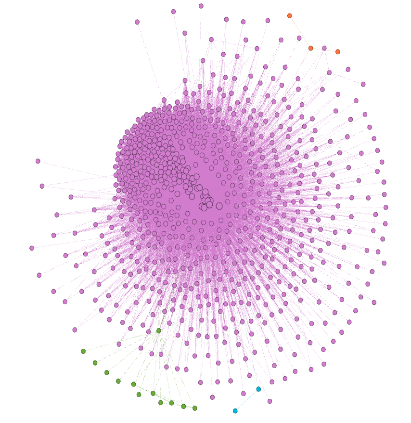
21. и 22. Лувенска метода је урађена на три различите вредности параметра. За резолуцију 0,1 креирано је 66 комуна, издваја се један кластер са 84 чвора и 5 кластера са преко 30-40 чворова. За резолуцију 1 креирано је 5 кластера, од којих се издвајају 2 кластера са око 450 и 300 чворова. За резолуцију 2 креирано је 4 кластера, од којих се издваја један са величином око 750 чворова. Резултати ове методе приказани су визуализацијом на сликама 6.3.3, 6.3.4 и 6.3.5. Визуализације најбитнијих чворова највећих комуна добијених Лувенском методом са параметром 1,0 приказане су на сликама 6.3.6 и 6.3.7.



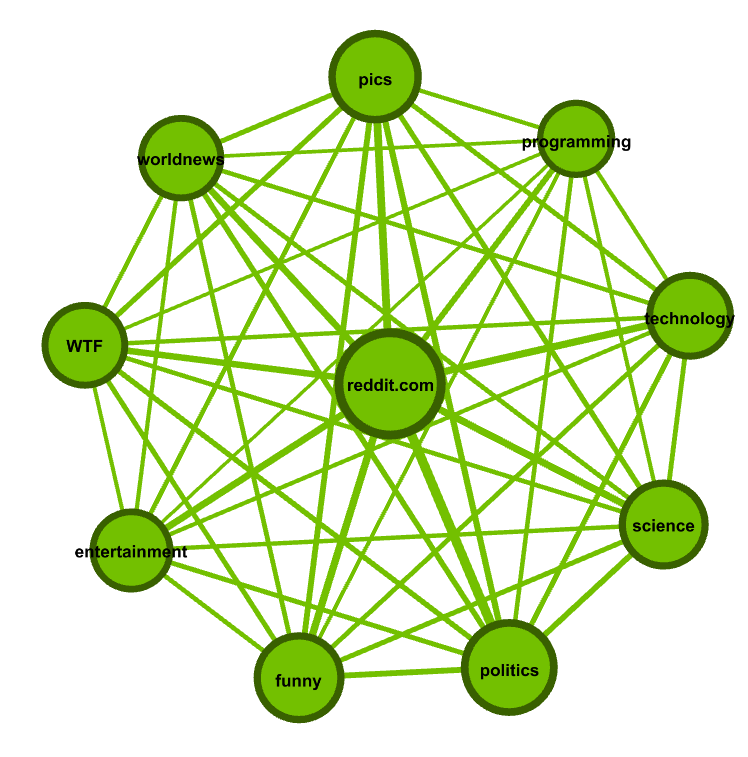
Слика 6.3.3 Лувенска метода уз параметар резолуције 0,1



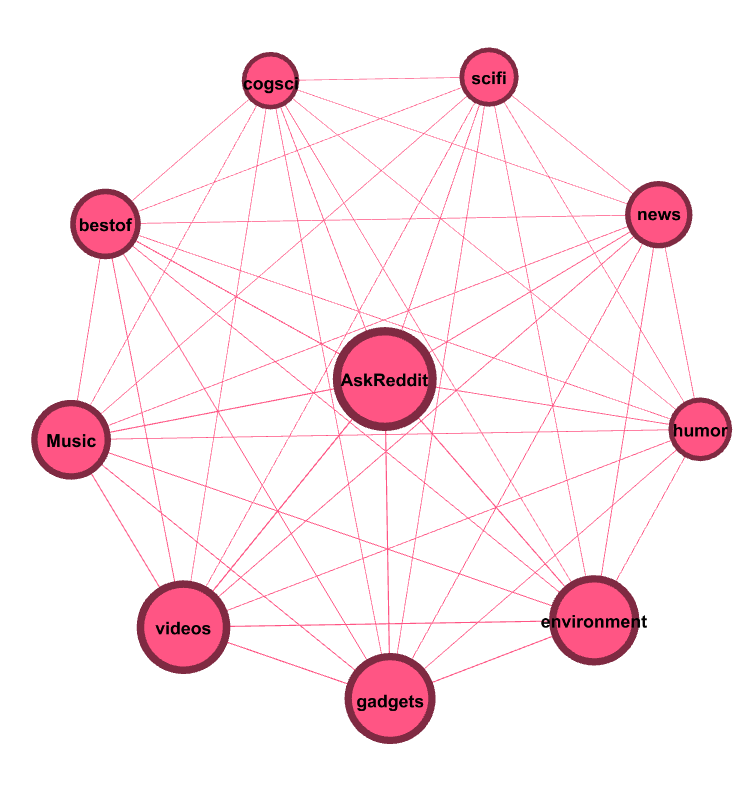
Слика 6.3.4 Лувенска метода уз параметар резолуције 1,0



Слика 6.3.5 Лувенска метода уз параметар резолуције 2,0



Слика 6.3.6 Мрежа најважнијих чворова прве комуне након кластерисања Лувенском методом уз параметар резолуције 1,0



Слика 6.3.7 Мрежа најважнијих чворова друге комуне након кластерисања Лувенском методом уз параметар резолуције 1,0

23. Чворови са највећом релационом централношћу су *reddit.com, politics, pics, technology, science, programming, worldnews, funny, WTF* и *entertainment*. Чворови који се појављују у највећем броју мостова из функције *NetworkX* су (*reddit.com*, 25), (*Chinese*, 6), (*programming*, 3), (*arabic*, 3), (*Gamer*, 2), (*ja*, 2), (*Josie*, 1), (*shoebush*, 1), (*onlycode*, 1), (*dotfiles*, 1), (*management*, 1), (*it*, 1), (*businessnews*, 1), (*Webometrics*, 1), (*Infos*, 1), (*fr*, 1), (*SinoBeta*, 1), (*sportscars*, 1), (*ChileRemix*, 1) и (*KiNov*, 1).

1. Мрежа UserNet

Мрежа UserNet представља усмерени тежински граф чији су чворови корисници друштвене мреже *Reddit*, док су везе усмерене од корисника који је оставио коментар ка кориснику којем је коментарисана објава или коментар. Тежина сваке гране одређена је бројањем постављених коментара, при чему су коментари на објаве и коментари на коментаре другог корисника сматране равноправним. Непостоје директне повратне гране, тј. не узимају се у обзир коментари корисника на сопствене објаве или коментаре. Тако добијена мрежа садржи 92.249 чворова и 2.803.021 грану.

* 1. Основна карактеризација мреже

7. Густина мреже износи око 0,00033.

8. Просечна дужина путања у мрежи износи око 3,489. Дијаметар мреже износи 13.

9. Број повезаних компоненти износи 43.651. Постоји једна гигантска компонента величине 48.502, једна компонента величине 6, по две компоненте величине 5 и 4, 12 компоненти величине 3, 44 компоненте величине 2, док су преостале компоненте сачињене од једног чвора. Издвојена је гигантска компонента као најзначајнија компонента.

10. Просечни коефицијент кластеризације износи око 0,117, док глобални коефицијент калстеризације износи око 0,0717. Расподела локалног коефицијента кластеризације по чворовима дата је на слици 7.1.1. Расподела локалног коефицијента кластеризације по чворовима приказана је за случајно генерисану *Erdos-Renyi* мрежу истих димензија и дата је на слици 7.1.2. На основу поређења ова два графика можемо закључити да је кластерисање врло изражено у посматраној мрежи.

Chart, histogram

Description automatically generated

Слика 7.1.1 График расподеле локалног степена кластеризације по чворима мреже

Chart, histogram

Description automatically generated

Слика 7.1.2 График расподеле локалног степена кластеризације по чворима за случајно генерисану *Erdos-Renyi* мрежу истих димензија

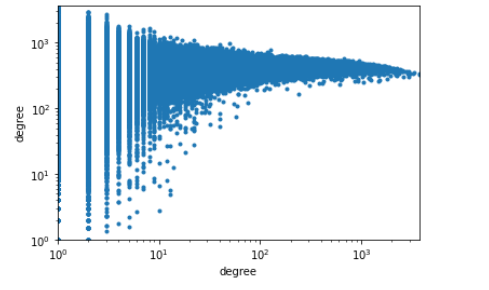
11. Мрежа исказује особине малог света због релативно малог дијаметра и просечне дужине путање, док је кластерисање јако изражено у поређењу са случајно генерисаном *Erdos-Renyi* мрежом истих димензија.

12. Коефицијент асортативности на основу нетежинског улазног степена чвора износи око 0,002151. Коефицијент асортативности на основу тежинског улазног степена чвора износи око -0,000105. Коефицијент асортативности на основу нетежинског излазног степена чвора износи око -0,038187. Коефицијент асортативности на основу тежинског излазног степена чвора износи око -0,026657. Улазни и излазни степен-степен график приказан је на сликама 7.1.3 и 7.1.4 респективно. Закључујемо да асортативно мешање није изражено.

Chart, scatter chart

Description automatically generated

Слика 7.1.3. График степен-степен корелације за улазни степен за мрежу *UserNet*



Слика 7.1.4. График степен-степен корелације за излазни степен за мрежу *UserNet*

13. На основу слика 7.3.4 и 7.3.5 закључујемо да се испољава феномен клуба богатих.

14. Графици расподеле за тежински улазни и тежински излазни степен чвора дати су на сликама 7.1.5. и 7.1.6. респективно. У оба случаја, поређењем са другим дистрибуцијама (*power law*, *exponential*, *lognormal*, итд.) добијамо да мрежа најбоље прати *truncated power law* расподелу.

Chart, scatter chart

Description automatically generated

Слика 7.1.5. График расподеле тежинског улазног степена чвора

Chart, scatter chart

Description automatically generated

Слика 7.1.6. График расподеле тежинског излазног степена чвора

15. С обзиром да је граф усмерен имамо и хабове и ауторитете. Најбитнији хабови (по улазном степену чвора) су корисници: rmuser (8433), qgyh2 (7545), noname99 (5675), alllie (5473), swampsparrow (5181), gst (4890), maxwellhill (4662), twolf1 (4052), georgewashingtonblog (3876) и 7oby (3866). Најбитнији ауторитети (по излазном степену чвора) су корисници: NoMoreNicksLeft (3747), mexicodoug (3726), 7oby (3317), h0dg3s (3239), randomb0y (3047), Poromenos (3018), mutatron (2979), nixonrichard (2976), malcontent (2939) и otakucode (2909). Ексцентричност најважнијих хабова и ауторитета износи 6-7, док дијаметар мреже износи 13, што указује да се они налазе ближе језгру мреже јер се налазе приближно на половини дијаметра.

* 1. Основна карактеризација мреже

16. Најважнији чворови по улазној централности по степену су rmuser (0,173873), qgyh2 (0,155564), noname99 (0,117008), alllie (0,112843), swampsparrow (0,106823), gst (0,100823), maxwellhill (0,096122), twolf1 (0,083545), georgewashingtonblog (0,079916) и 7oby (0,079710).

Најважнији чворови по излазној централности по степену су NoMoreNicksLeft (0,077256), mexicodoug (0,076823), 7oby (0,068390), h0dg3s (0,066782), randomb0y (0,062823), Poromenos (0,062226), mutatron (0,061421), nixonrichard (0,061360), malcontent (0,060597)и otakucode (0,059978).

Због величине мреже најважније актере по централности по блискости није могуће одредити.

Најважнији чворови по релационој централности су rmuser (0,022881), 7oby (0,022615), alllie (0,018601), qgyh2 (0,013372), NoMoreNicksLeft (0,011425), Poromenos (0,010615), malcontent (0,010069), nixonrichard (0,009931), otakucode (0,009250), mutatron (0,008516). Услед велике временске сложености, овде је употребљена техника пивота.

17. Због усмерености мреже није одређена централност по сопственом вектору.

18. Најважнији чворови по улазној Кацовој централности (параметар β=1) су rmuser (0,118230), qgyh2 (0,080567), alllie (0,051556), gst (0,045346), noname99 (0,044977), malcontent (0,039271), maxwellhill (0,037690), matts2 (0,037083), twolf1 (0,035136) и NoMoreNicksLeft (0,035128). Са различитим бета параметром за сабредит *reddit.com* и параметром α=0,0005 добијамо следеће рангирање: rmuser (0,118230), qgyh2 (0,080567), alllie (0,051556), gst (0,045346), noname99 (0,044977), malcontent (0,039271), maxwellhill (0,037690), matts2 (0,037083), twolf1 (0,035136) и NoMoreNicksLeft (0,035128).

Најважнији чворови по излазној Кацовој централности (параметар β=1) су rmuser (0,081058), qgyh2 (0,080068), alllie (0,051156), gst (0,045418), noname99 (0,044484), malcontent (0,039209), maxwellhill (0,037556), matts2 (0,037077), NoMoreNicksLeft (0,035127) и twolf1 (0,034962). Са различитим бета параметром за сабредит *reddit.com* и параметром α=0,0005 добијамо следеће рангирање: rmuser (0,050520), qgyh2 (0,009647), alllie (0,007694), gst (0,007661), noname99 (0,007327), maxwellhill (0,006818), swampsparrow (0,006798), twolf1 (0,006591), georgewashingtonblog (0,006564) и IAmperfectlyCalm (0,006350).

19. Изабрана формула за рачунање композитног ранга је:

*composite\_rank = DC\_in\_rank \* DC\_out\_rank \* BC\_rank \* Katz\_rank \* Katz\_out\_rank*

Најважнији актери по композитном рангу су rmuser, alllie, NoMoreNicksLeft, qgyh2, 7oby, malcontent, nixonrichard, Poromenos, matts2 и otakucode.

* 1. Основна карактеризација мреже

20. Спектралну и дендрограмску анализу није могуће спровести због величине мреже.

21. и 22. Лувенска метода је спроведена за три различите вредности параметара резолуције. За параметар резолуције 0,1 креирана је 1582 комунe. Преовладавају веома мале комуне, међутим постоји једна већа комуна од преко 1100 актера, 3 комуне од по 300 и 3 комуне од по 200 чворова. За параметар резолуције 1,0 креиране су 25 комуна, где се истичу 5 великих комуна (13.000, 10.000, 9.000, 8.500 и 6.000 чворова – слике 6.3.4 и 6.3.5), док су остале много мање. За параметар резолуције 2,0 креирано је 27 комуна, где се истиче једна огромна од преко 45.000 чворова, док су све остале мале. Визуелизације мреже након примене ове методе за наведене вредности приказане су сликама 7.3.1, 7.3.2 и 7.3.3. Визуализације најбитнијих чворова највећих комуна добијених Лувенском методом са параметром 1,0 приказане су на сликама 7.3.4 и 7.3.5.

A close up of a dragonfly

Description automatically generated with medium confidence

Слика 7.3.1. Визуализација након Лувенске методе са параметром 0,1

Chart, radar chart

Description automatically generated

Слика 7.3.2. Визуализација након Лувенске методе са параметром 1,0

A close-up of a feather

Description automatically generated with low confidence

Слика 7.3.3. Визуализација након Лувенске методе са параметром 2,0

Chart, radar chart

Description automatically generated

Слика 7.3.4. Визуализација најбитнијих чворова највеће комуне након Лувенске методе са параметром 1,0

Diagram

Description automatically generated

Слика 7.3.5. Визуализација најбитнијих чворова друге највеће комуне након Лувенске методе са параметром 1,0

23. Није могуће путем *Gephi* алата уочити кључне брокере због нејасне визуализације услед великог броја чворова и грана. Ипак, ако узмемо у обзир релациону централност главни посредници су rmuser (0,022881), 7oby (0,022615), alllie (0,018601), qgyh2 (0,013372), NoMoreNicksLeft (0,011425), Poromenos (0,010615), malcontent (0,010069), nixonrichard (0,009931), otakucode (0,009250), mutatron (0,008516).

1. Поређење SNet и SNetT мрежа

У овом поглављу биће разматране најбитније разлике измежу мрежа SNet и SNetT.

SNetT je потпуо повезана мрежа за разлику од SNet мреже. Ако поредимо густину мрежа можемо закључити да је SNetT мрежа је потпуно повезана па њена густина износи 1. Густина SNet мреже изности свега око 0,0165.

Мрежа SNetT има 39 чворова и 741 грану. Ова мрежа представља уску мрежу сабредита који су се по поставци задатка бавили темом светске економске кризе у 2008. години самим тим она је по својој величини значајно мања од SNet мреже. Мрежа SNetT представља подскуп SNet мреже и гледајући њену ексцентричност унутар родитељске мреже можемо закључити да се налази у самом центру мреже и представља њено језгро.

Ако сагледамо просечни тежински степен чвора, просечна активност мреже SNetT износи око 100.431, док активност остатка мреже износи око 265. Дакле, не само да су чворови SNetT мреже гушће повезани, већ они предстаљају епицентар активности мреже.