

УНИВЕРЗИТЕТ У БЕОГРАДУ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ



АНАЛИЗА СОЦИЈАЛНИХ МРЕЖА

Пројектни задатак

Студенти:

Ана Стакић 2020/3103

Страхиња Стефановић 2020/3104

Школска година:

2021/2022.

Београд, јануар 2022.

Садржај

САДРЖАЈ	2
1. ЧИШЋЕЊЕ ПОДАТАКА	3
2. МОДЕЛОВАЊЕ МРЕЖА	5
2.1. SUBREDDIT NETWORK – SNET.....	5
2.2. FILTERED SUBREDDIT NETWORK – SNETF	6
2.3. TARGETED SUBREDDIT NETWORK – SNETT.....	8
2.4. USER NETWORK – USERNET.....	9
3. СТАТИСТИЧКА ОБРАДА ПОДАТАКА	10
4. ОСНОВНА АНАЛИЗА МОДЕЛОВАНИХ МРЕЖА	13
5. АНАЛИЗА МЕРА ЦЕНТРАЛНОСТИ	28
6. ДЕТЕКЦИЈА КОМУНА	37
7. ПОРЕЂЕЊЕ SNET И SNETT МРЕЖА	47
СПИСАК СЛИКА	49
СПИСАК ТАБЕЛА	50

1. ЧИШЋЕЊЕ ПОДАТАКА

Најпре је било потребно податке о објавама и коментарима подељене у по 12 засебних табела за сваки појединачни месец у 2008. години груписати у две засебне табеле. Прва садржи податке о свим објавама у тој години, а друга податке о свим коментарима. Приликом спајања табела, додата је и још једна колона *month* целобројног типа која памти редни број месеца у којем је креирана одговарајућа објава, односно коментар. Ова колона је додата да би се касније, евентуално, извршила нека анализа датих података заснована на времену настанка.

Након тога, из добијених табела су избачени редови са инвалидним идентификатором и та колона је преименована у *submission_id*, односно *comment_id*, како би постојала јединствена колона *id* са индексом реда. Такође, из табела су избачене колоне које нису од интереса за даљу анализу. Из табеле објава избачене су: *created_utc*, *url*, *permlink*, *domain*, *distinguished*. Колона *distinguished*, иако у теорији занимљива за анализу зато што показује да ли је аутор објаве био админ одговарајућег сабредита, се није показала као корисна зато што је веома мали узорак заиста и имао неку валидну вредност те колоне (мање од 10 редова). Из тог разлога је и она уклоњена приликом стварања секундарног скупа података. Из табеле коментара су избрисане следеће колоне: *created_utc*, *distinguished*, *controversiality*. Колона *controversiality* је избачена из истог претходно поменутог разлога.

Током прегледа скупа података примећено је да се неретко у колони *author* налази вредност *[deleted]* што је вероватно последица ажурирања овог скупа података у блиској прошлости, како објаву или коментар може поставити само постојећи регистровани корисник ове платформе. Узимајући у обзир да је за даљу анализу и моделовање тражених мрежа (како мреже корисника, тако и мрежа сабредита) нужен податак о аутору, одлучено је да се сви редови са овом вредношћу уклоне. Такође, пошто сваки коментар има линк ка објави за коју је везан, било је потребно избрисати и коментаре који сада указују на објаве које више не постоје у табели објава, а које имају *[deleted]* аутора.

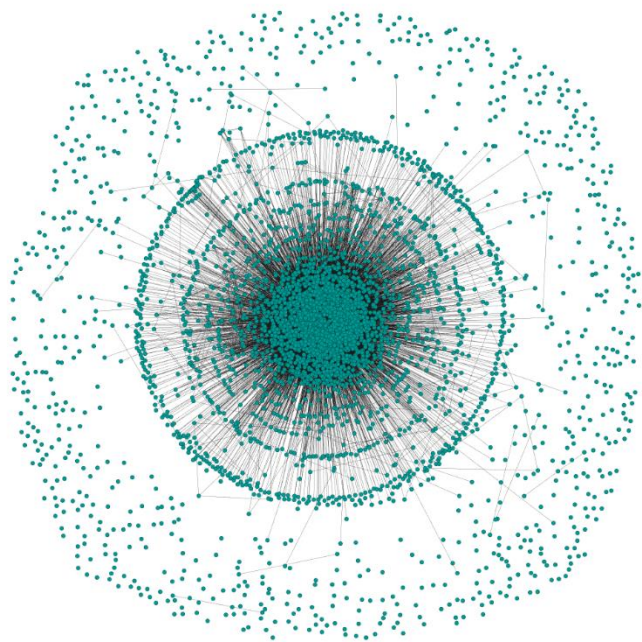
Табела 1-1 Утицај чишћења на величину скупа података

	Примарни скуп података	Секундарни скуп података
Објаве	$2.519.853 \times 16$	$2.044.810 \times 12$
Коментари	$7.242.871 \times 12$	$3.783.039 \times 10$

2. МОДЕЛОВАЊЕ МРЕЖА

2.1. Subreddit network – SNet

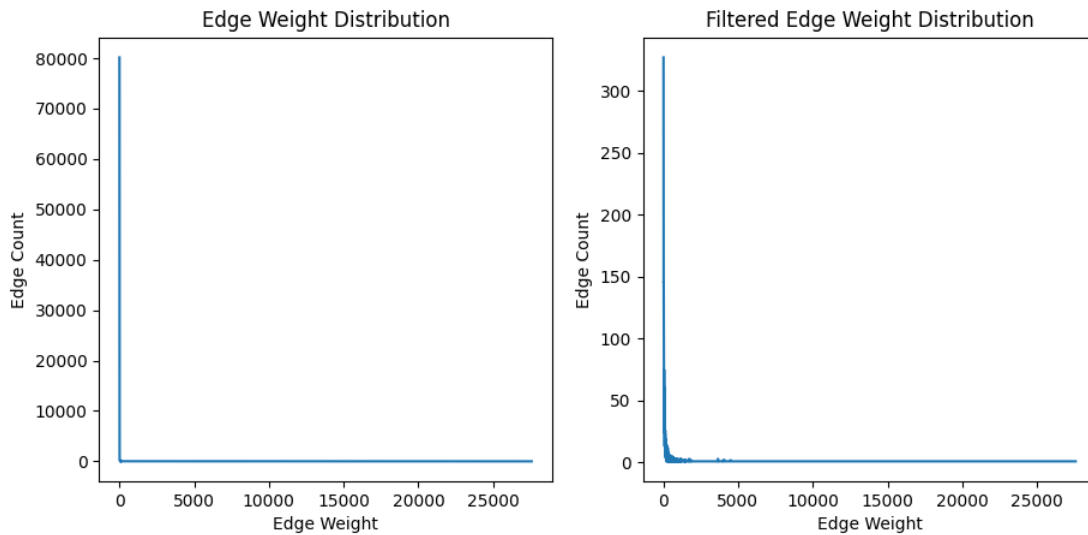
Мрежа сабредита представља тежински неусмерени граф у којем су чворови сабредити, а веза између чворова се успоставља уколико постоје корисници који су били активни на оба сабредита. Тежине грана су добијене простом агрегацијом бројањем. Ова мрежа је формирана на основу комплетног скупа података.



Слика 2.1 Мрежа SNet

2.2. Filtered subreddit network – SNetF

Мрежа добијена уклањањем грана са тежином испод одређеног прага – $w_threshold$. Најпре је било потребно визуелизовати расподелу тежина грана у SNet мрежи.

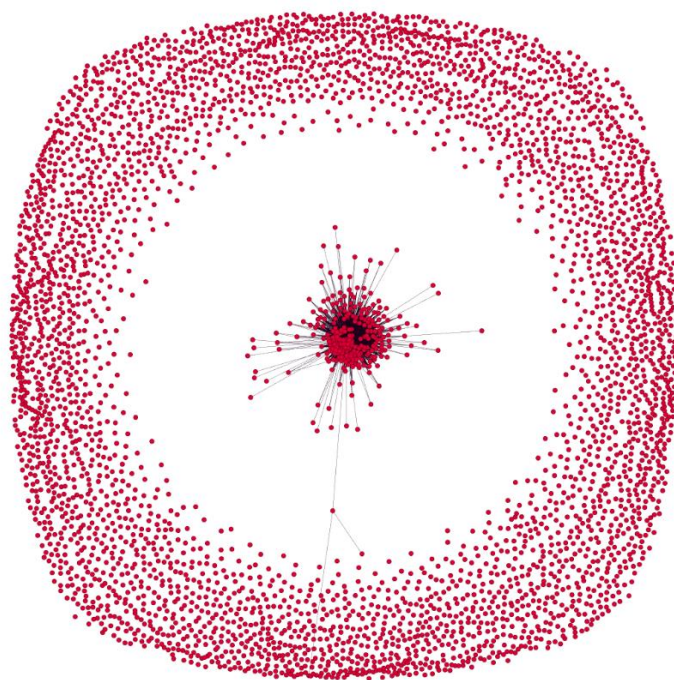


Слика 2.2 Расподела тежина грана

Са графика се може видети да највећи проценат грана има минималну тежину (један), док са повећањем тежине број грана драстично опада све до тзв. *outlier*-а који представљају гране огромне тежине. Из претходно наведеног може се рећи да тежине грана, као и тежине чворова прате *power law* расподелу. За потребе генерисања SNetF мреже, као тежински праг изабрана је вредност 20 што драстично редукује број грана у мрежи.

Табела 2-1 Промена броја грана приликом филтрације

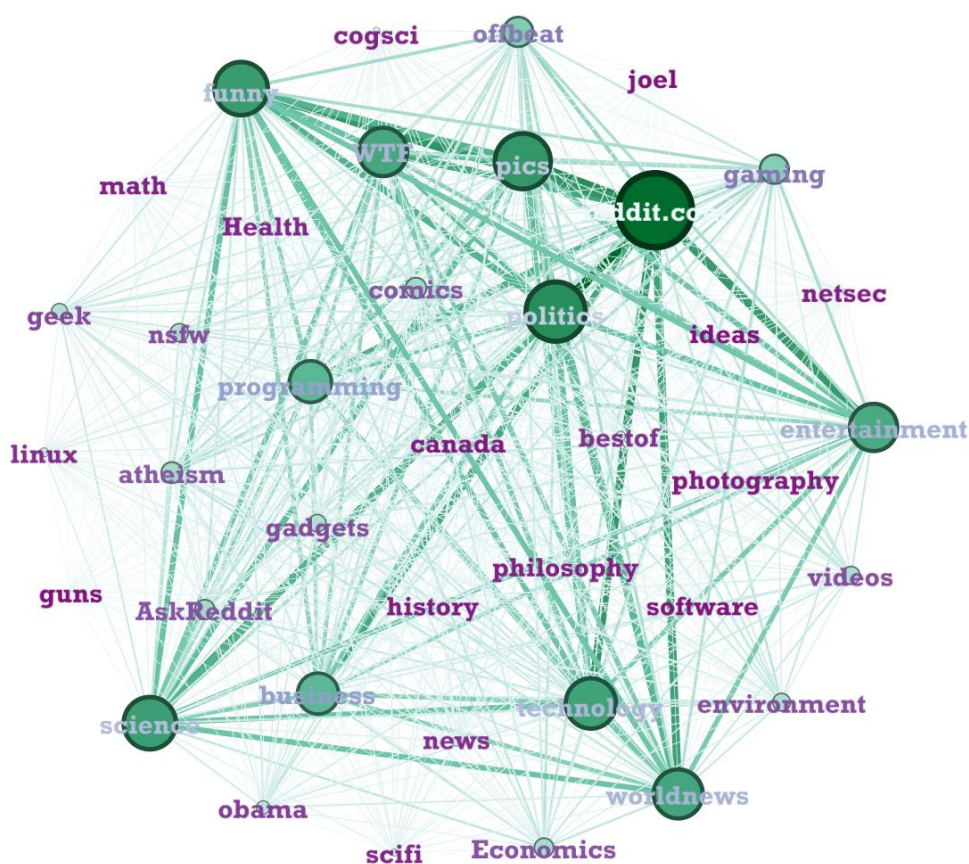
	Пре филтрирања	После филтрирања
Број грана	137.264	9.913



Слика 2.3 Мрежа SNetF

2.3. Targeted subreddit network – SNetT

Još jedna podmreža prvobitne SNet mreže s tim da se sada uklanjaју svi sabrediti koji ne pripadaју задатом скупу sabredita чија тематика је блиска теми економске кризе. Укупан број sabredita је **4191**, док подскуп који се посматра броји **39** sabredita.

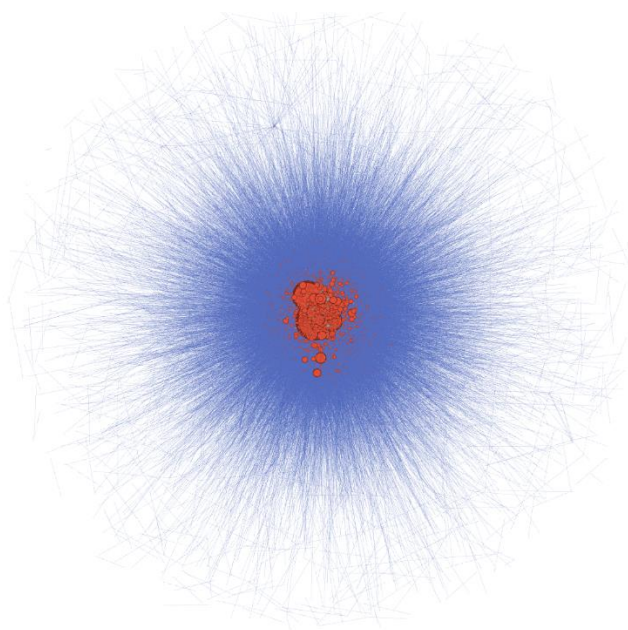


Слика 2.4 Мрежа SNetT

2.4. User network – UserNet

Мрежа интеракције између корисника *Reddit* платформе у којој су корисници повезани уколико су коментарисали једни другима објаве или коментаре.

Приликом креирања мреже, корисници су представљени као чворови графа, док усмерене гране представљају интеракцију између корисника. Два типа интеракције – коментар објаве и коментар коментара, тумачени су и третирани на исти начин. Због огромне количине како корисника, тако и њихових интеракција, у мрежу су укључени само они парови корисника између којих интеракције, односно гране, постоје у оба смера.



Слика 2.5 Мрежа UserNet

3. СТАТИСТИЧКА ОБРАДА ПОДАТАКА

- 1) Колико постоји различитих сабредита који се појављују у посматраном периоду? Који су најважнији по броју корисника, а који по броју коментара?

Укупан број сабредита: **4191**

Табела 3-1 Анализа сабредита

Subreddit	Number of Authors	Subreddit	Number of Comments
reddit.com	159563	reddit.com	860727
politics	35769	politics	625028
pics	27537	programming	294857
technology	26701	pics	219045
funny	26308	science	192784
entertainment	25204	worldnews	186210
programming	24243	WTF	140085
science	24044	funny	136459
business	23954	technology	120440
worldnews	23240	entertainment	112399

- 2) Какав је просечан број забележених корисника активних у посматраном периоду по сабредиту? Корисник се сматра активним на сабредиту ако је забележен барем један коментар или објава тог корисника.

Просечан број корисника по сабредиту: **144.4063**

- 3) Ко су корисници са највећим бројем објава, а ко корисници са највећим бројем коментара?

Табела 3-2 Анализа корисника

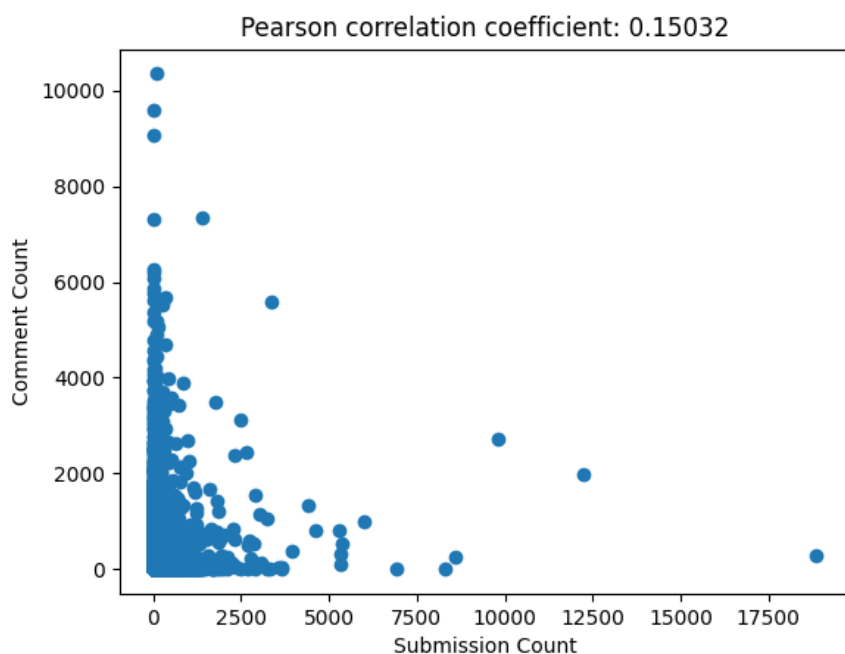
Author	Number of Submissions	Author	Number of Comments
gst	18870	NoMoreNicksLeft	10345
qgyh2	12238	malcontent	9598
rmuser	9822	matts2	9063
twolf1	8597	7oby	7343
IAmperfectlyCalm	8308	mexicodoug	7313
qazamisan	6927	aletoledo	6255
charlatan	5998	otakucode	6200
igeldard	5373	redditcensoredme	6090
noname99	5334	h0dg3s	5875
democracy101	5332	Poromenos	5770

- 4) Који корисници су активни на највећем броју сабредита? На колико су сабредита активни?

Табела 3-3 Најактивнији корисници

Author	Number of Subreddits
Escafane	152
MrKlaatu	147
omfgninja	113
scientologist2	106
codepoet	106
krugerville	102
turkourjurs	102
tuoder	101
b34nz	99
bobcat	99

- 5) Како су корелисани бројеви објава и бројеви коментара корисника? Одредити Пирсонов коефицијент корелације и извршити визуелизацију.



Слика 3.1 Пирсонов коефицијент корелације

- 6) Које објаве поседују највећи број коментара и на којим су сабредитима постављене? Приказати податке о тим објавама, укључујући то на којем су сабредиту постављене и шта им је садржај (ако је поље објаве *over 18* постављено на *false*).

Табела 3-4 Објаве са највећим бројем коментара

Author	Subreddit	*num_comments	Score	**Number of Comments
zekel	science	33329	1621	1866
matiasklein	reddit.com	2039	1098	1500
willjohnston	politics	1934	8538	1375
rpi22	reddit.com	1377	669	1297
TheRealStick	politics	1425	788	1076
Schlichten	worldnews	1335	1589	1063
georgeb	programming	1269	1904	1057
nucleophile	reddit.com	1425	1517	1036
IM_A_REPTILIAN	politics	1346	3361	1028
cup	reddit.com	1342	269	1004

*num_comments – број коментара на објави из истоимене колоне почетног сета података где су највероватније урачунати и коментари настали након 2008. године

**Number of Comments – сума коментара на објави, након филтрирања сета података

4. ОСНОВНА АНАЛИЗА МОДЕЛОВАНИХ МРЕЖА

7) Колика је густина мреже?

8) Колике су просечне дистанце у оквиру мреже и дијаметар мреже?

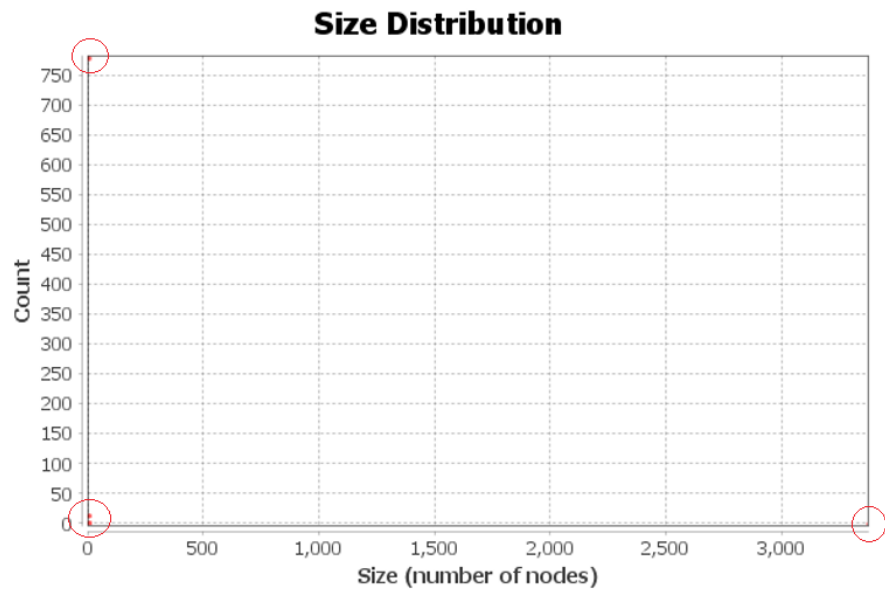
Табела 4-1 Густина и дијаметар мреже

	SNet	SNetF	SNetT	UserNet
Густина мреже	0.016	0.001	1.0	~ 0.0
Просечна дистанца	2.1065	1.84	1.0	4.196
Дијаметар мреже	5.0	4.0	1.0	14.0

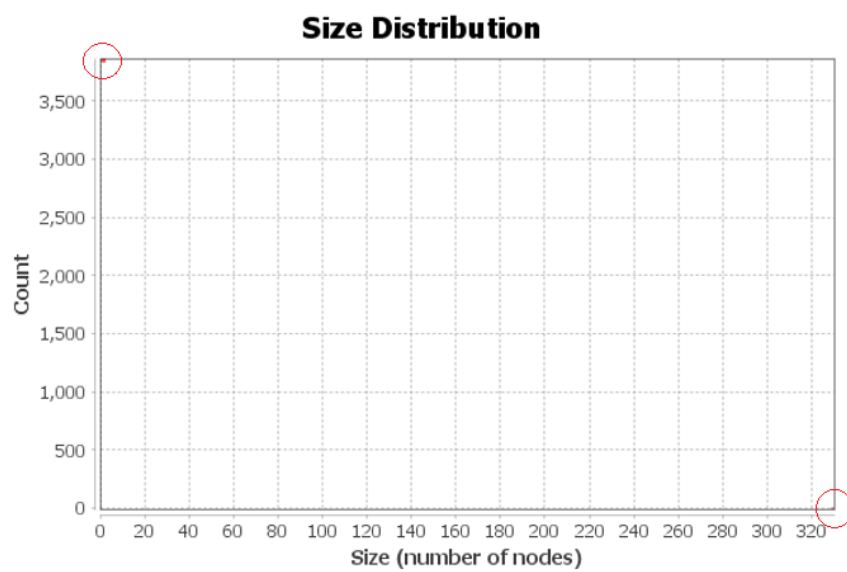
9) У којој мери је мрежа повезана и централизована? Навести број и величине повезаних компонената и проценити да ли постоји гигантска компонента.

Табела 4-2 Број повезаних компонената мреже

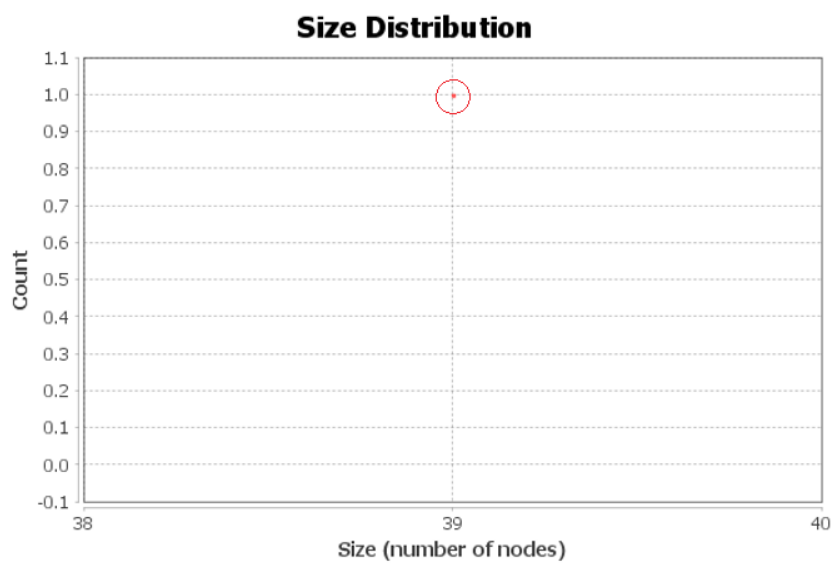
	SNet	SNetF	SNetT	UserNet
Број повезаних компонената	799	3863	1	201



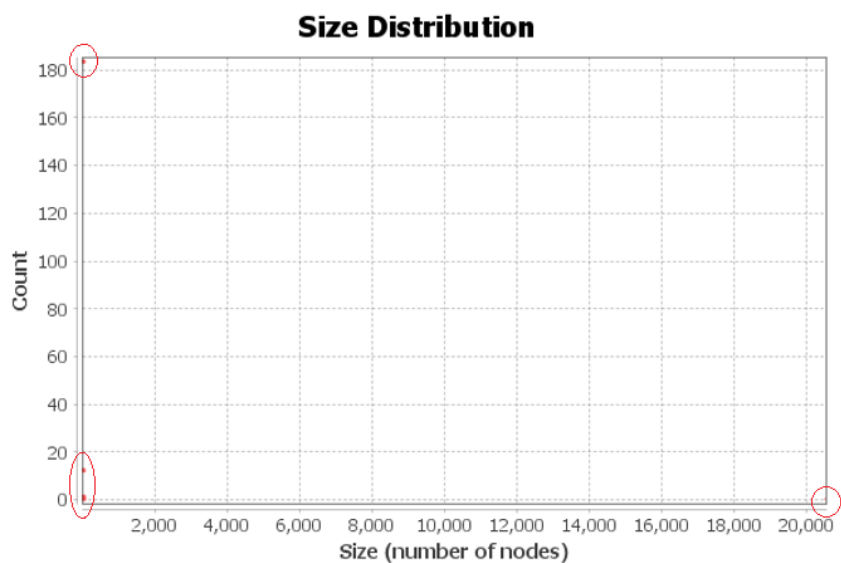
Слика 4.1 Величина компонентата мреже SNet



Слика 4.2 Величина компонентата мреже SNetF



Слика 4.3 Величина компонената мреже SNetT



Слика 4.4 Величина компонената мреже UserNet

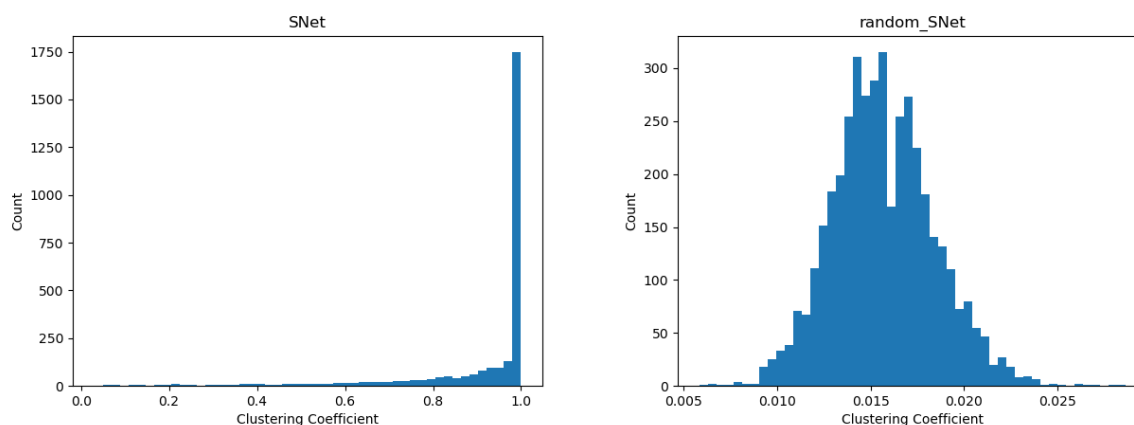
Све мреже осим мреже SNetT поседују једну гигантску компоненту. Мрежа SNetT је специфична по томе што представља комплетан граф, односно има једну повезану компоненту максималне величине.

- 10) Колики је просечни, а колики глобални коефицијент кластеризације мреже? Каква је расподела локалног коефицијента кластеризације њених чворова? Да ли је кластерисање изражено или не? Одговор дати упоређивањем са случајно генерисаном *Erdos-Renyi* мрежом истих димензија.

Табела 4-3 Коефицијент кластеризације мреже

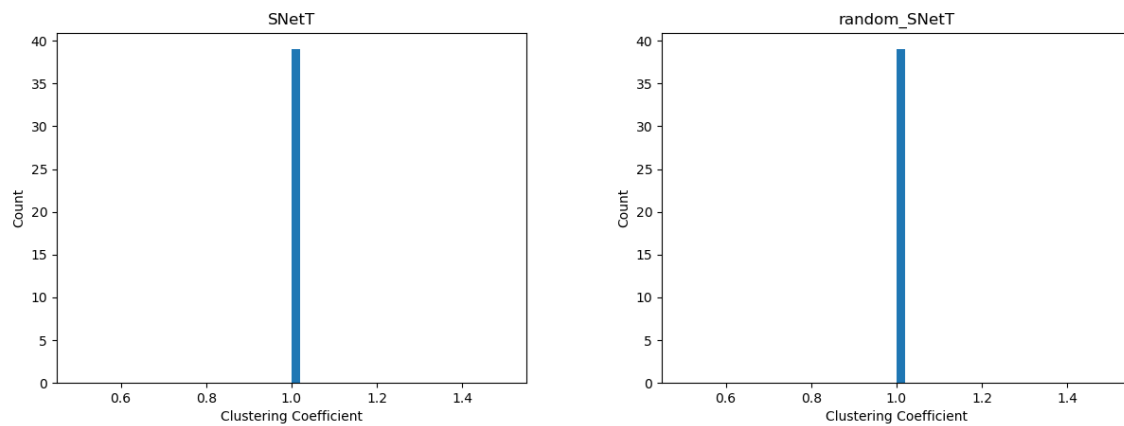
	Просечни	Глобални	Просечни	Глобални
	Са утицајем тежина грана		Без утицаја тежина грана	
SNet	0.000244	0.001199	0.618883	1.0
SNetF	0.000446	0.013474	0.066099	1.0
SNetT	0.0621021	0.1411455	1.0	1.0
UserNet	0.000320	0.094210	0.028932	1.0

У наставку следе визуелизовани прикази расподела коефицијената кластеризације за сваку од мрежа, као и за случајно генерисане *Erdos-Renyi* мреже истих димензија.



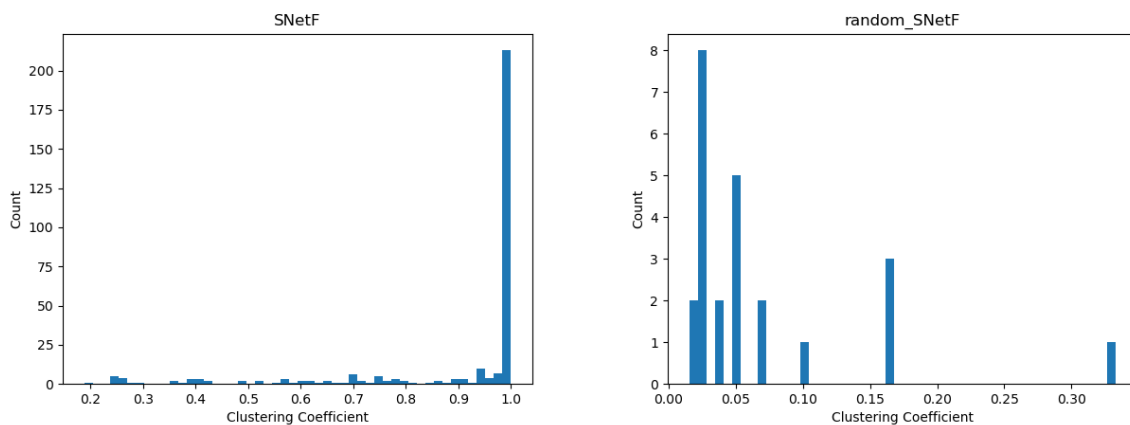
Слика 4.5 Расподела коефицијента кластеризације мреже SNet и случајне мреже истих димензија

Са левог графика се може видети да највећи број чворова има максималан коефицијент кластеризације (један) у оригиналној мрежи. Тај скуп чворова може представљати један добро повезани кластер унутар саме мреже, централни скуп са **Слика 2.1**, па самим тим кластерисање ове мреже јесте изражено.



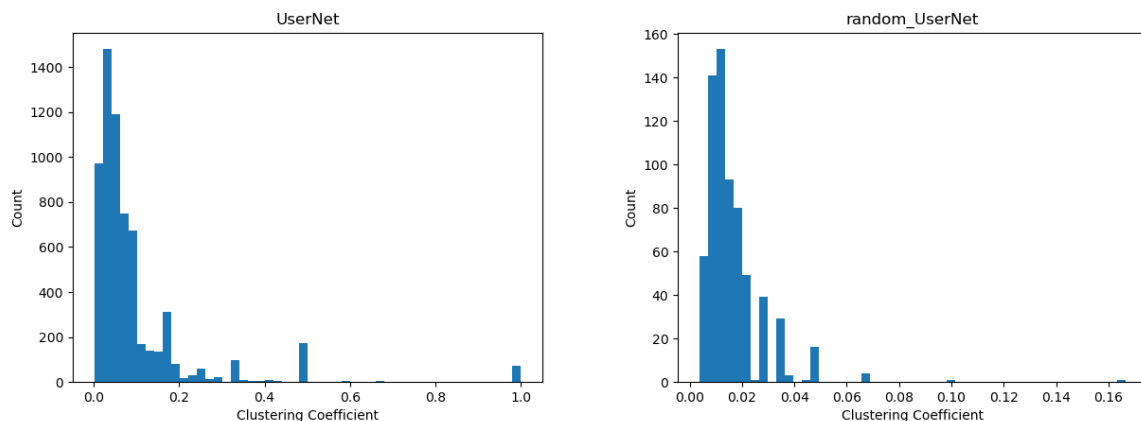
Слика 4.6 Расподела коефицијента кластеризације мреже SNetT и случајне мреже истих димензија

Приликом креирања случајне мреже, прослеђена су два параметра – број чворова у мрежи и вероватноћа да између два чвора постоји грана. Управо учествовање ове вероватноће, која за потпуно повезани мрежни граф SNetT има вредност 1 (100%), доводи до тога да и оригинална и случајно генерисана мрежа изгледају идентично. Самим тим и њихова расподела коефицијента кластеризације је потпуно иста и сви чворови поседују коефицијент 1 (један), а мрежа представља тачно један кластер.



Слика 4.7 Расподела коефицијента кластеризације мреже SNetF и случајне мреже истих димензија

Слично као код расподеле коефицијента кластеризације SNet мреже, са левог графика се може видети да највећи број чворова има максималан коефицијент (један) у оригиналној мрежи. У примеру ове мреже, скок тог дела графика још је израженији у односу на сличан скок код SNet мреже, разлог тога видљив је и као изолованији подскуп чворова на **Слика 2.3**, а самим тим кластерисање и ове мреже јесте изражено.



Слика 4.8 Расподела коефицијента кластеризације мреже UserNet и случајне мреже истих димензија

Расподела коефицијента кластеризације за мрежу корисника представља доминацију чворова са јако малим вредностима коефицијента (између 0 и 0.2), што говори о слабијој повезаности чворова, него што је то случај у осталим мрежама. Са друге стране, уочава се велика сличност расподеле оригиналне и случајно генерисане мреже. На основу тога можемо закључити да највероватније и не постоји неки систематични начин повезивања корисника, односно да је углавном насумичан.

11) На основу одговора на питања 8 и 10, проценити да ли мрежа исказује особине малог света.

Да би мрежа исказивала особине малог света треба да има малу просечну удаљеност било која два чвора у односу на величину мреже и висок степен кластеризације.

	Просечна дистанца	Просечан степен кластеризације	Особина малог света
SNet	2.1065	0.618883	ДА
SNetF	1.84	0.066099	НЕ (низак степен кластеризације)
SNetT	1.0	1.0	ДА
UserNet	4.196	0.028932	НЕ (низак степен кластеризације)

12) Извршити асортативну анализу по степену чвора и дати одговор да ли је изражено асортативно мешање. У случају да је мрежа усмерена, анализу извршити и по улазном и по излазном степену чвора. Приложити и визуелизацију.

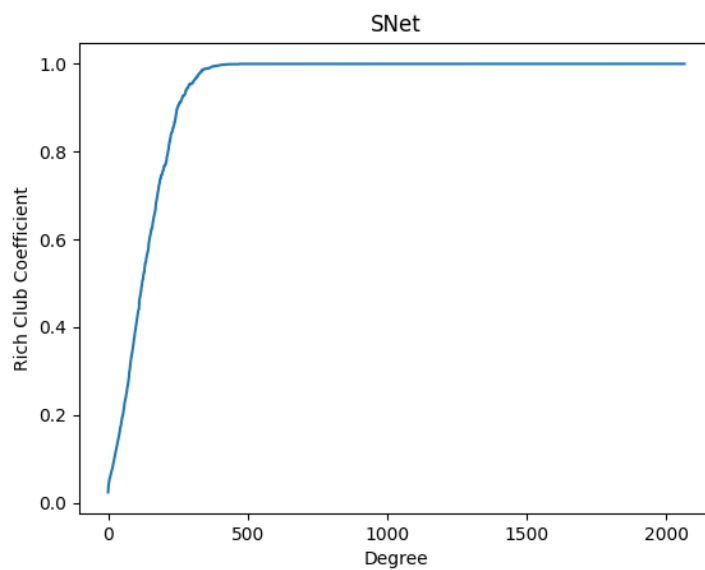
Табела 4-4 Асортативност по степену чвора

	Асортативност по степену чвора		Тежинска асортативност по степену чвора	
	Улазни	Излазни	Улазни	Излазни
SNet	-0.43037		-0.03707	
SNetF	-0.61644		0.01636	
SNetT	NaN		0.00017	
UserNet	0.00190	-0.00020	0.02562	-0.00869

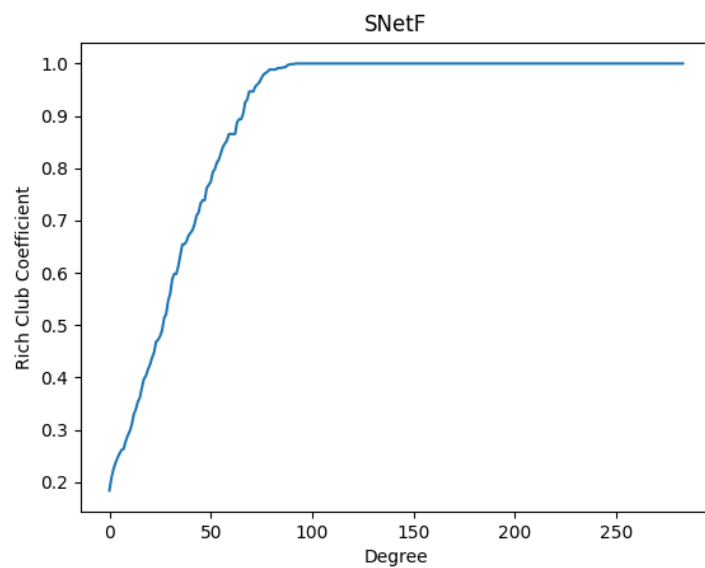
Може се приметити да у већем броју случаја асортативно мешање није изражено, осим у случају мрежа SNet i SNetF када се тежина грана не узме у обзир. Тада нам коефицијент асортативности говори да се чворови са већим степеном (чворови груписани у језгрима на **Слика 2.1** и **Слика 2.3**) најпре повезују са чворовима мањег степена (чворови на периферији мреже).

Обележена NaN вредност у табели је последица чињенице да је мрежа SNetT потпун граф. Наиме, уколико посматрамо математичку дефиницију асортативности можемо увидети да је то у ствари нормализована корелација степена два чвора. Уколико су степени свих чворова исти, као што је то овде случај, онда је корелација једнака нули, као и варијанса, па добијамо $0/0 = NaN$. Уколико су степени скоро једнаки, као што је то случај код тежинске асортативности, онда је корелација скоро једнака нули, варијанса различита од нуле, па добијамо *врло_мала_вредност/ненулта_вредност = врло_мала_вредност*. Овај резултат нам у ствари говори о слободи избора, у SNetT мрежи чворови немају избора него да се повежу са свим чворовима, што није случај у осталим мрежама где је распон степена чвора доста шири.

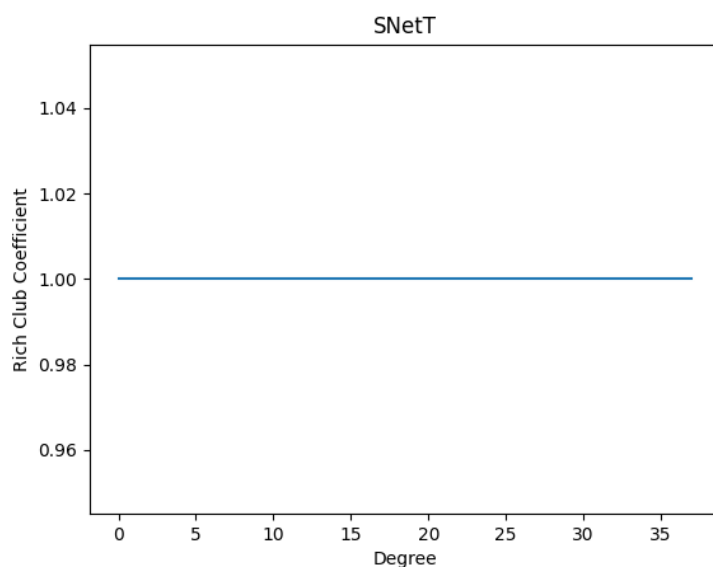
13) Да ли мрежа испољава феномен клуба богатих (енг. *rich club phenomenon*)?



Слика 4.9 Коефицијент клуба богатих мреже SNet



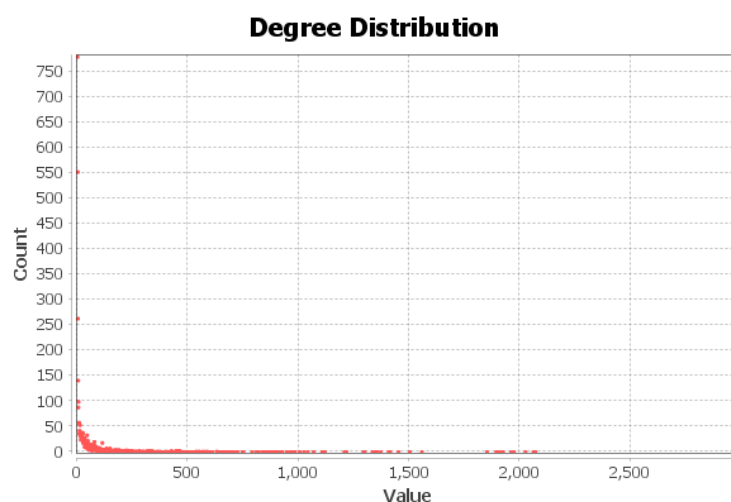
Слика 4.10 Коефицијент клуба богатих мреже SNetF



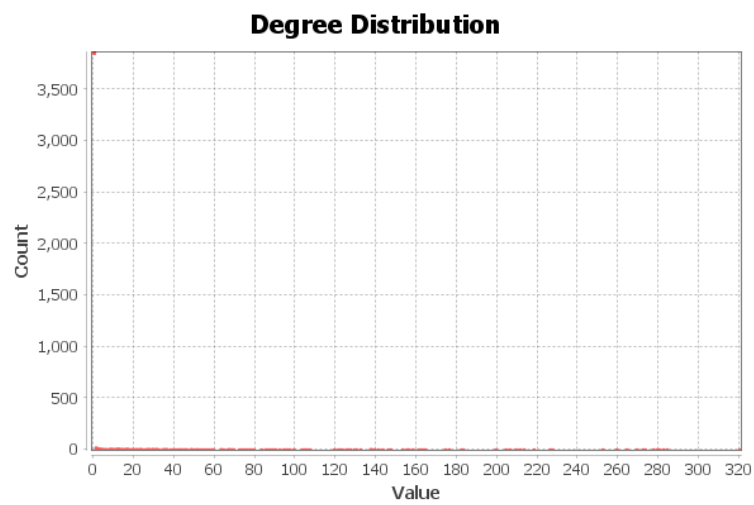
Слика 4.11 Коефицијент клуба богатих мреже SNetT

Изузев мреже SNetT која је потпун граф и где је очекивана вредност коефицијента максимална, може се приметити да је и у осталим мрежама сабредита присутан феномен клуба богатих почев од одређене вредности степена чвора. То су опет они сабредити који се налазе у језгру мреже и нови корисници који приступе платформи ће највероватније постати чланови тих сабредита (*The rich get richer*).

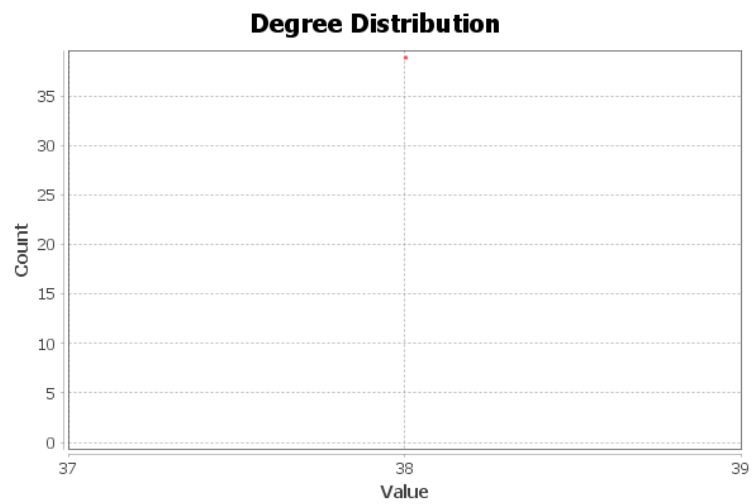
14) Каква је дистрибуција чворова по степену и да ли прати *power law* расподелу?



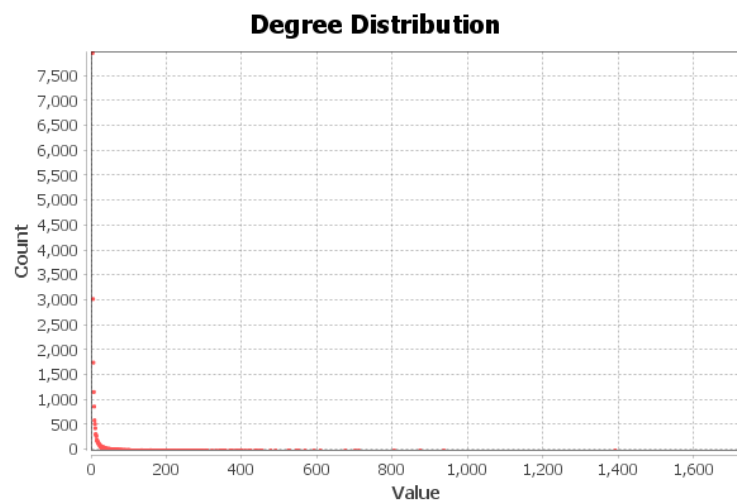
Слика 4.12 Дистрибуција чворова по степену мреже SNet

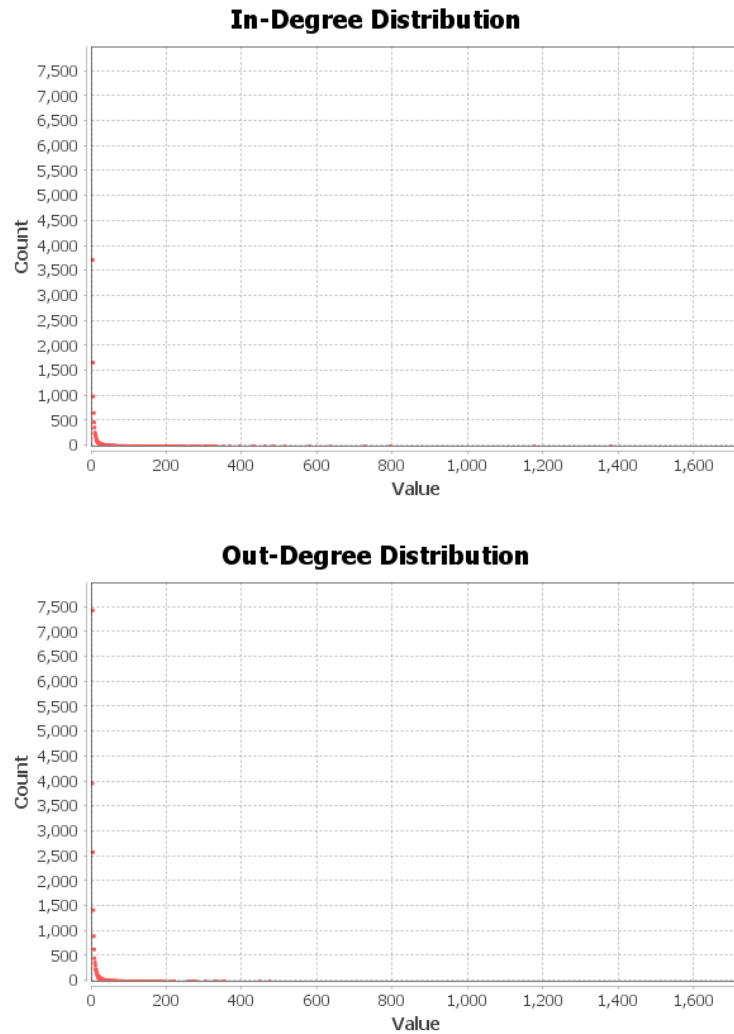


Слика 4.13 Дистрибуција чворова по степену мреже SNetF



Слика 4.14 Дистрибуција чворова по степену мреже SNetT



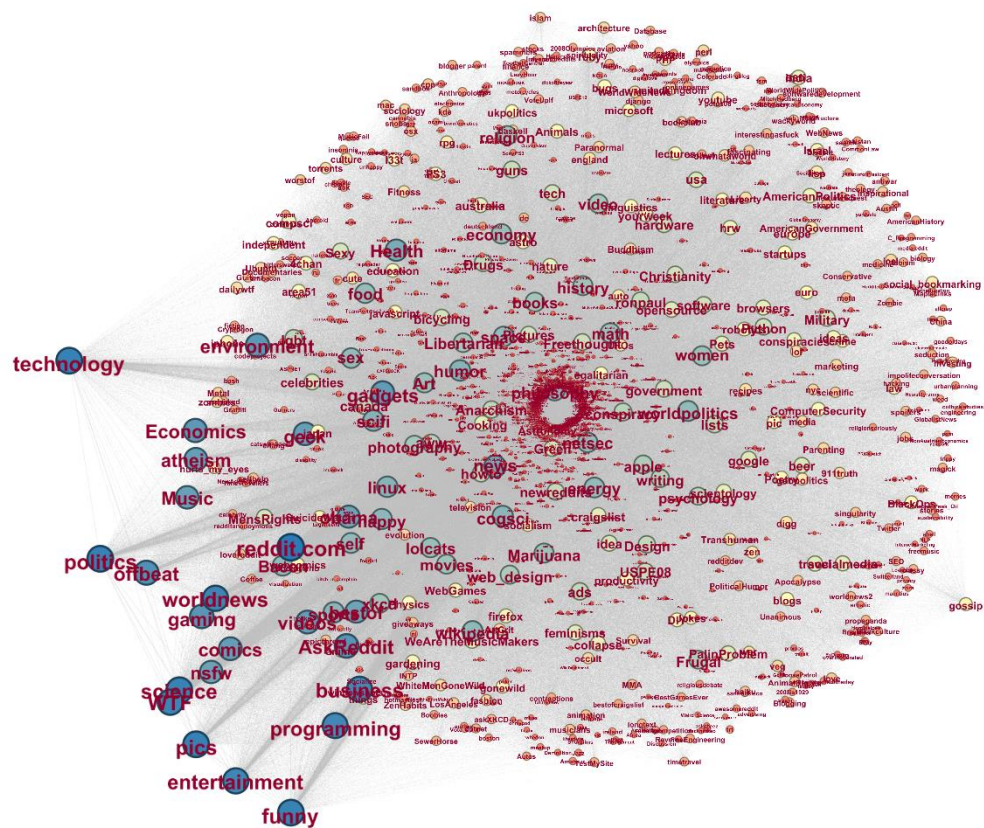


Слика 4.15 Дистрибуција чворова по степену мреже UserNet

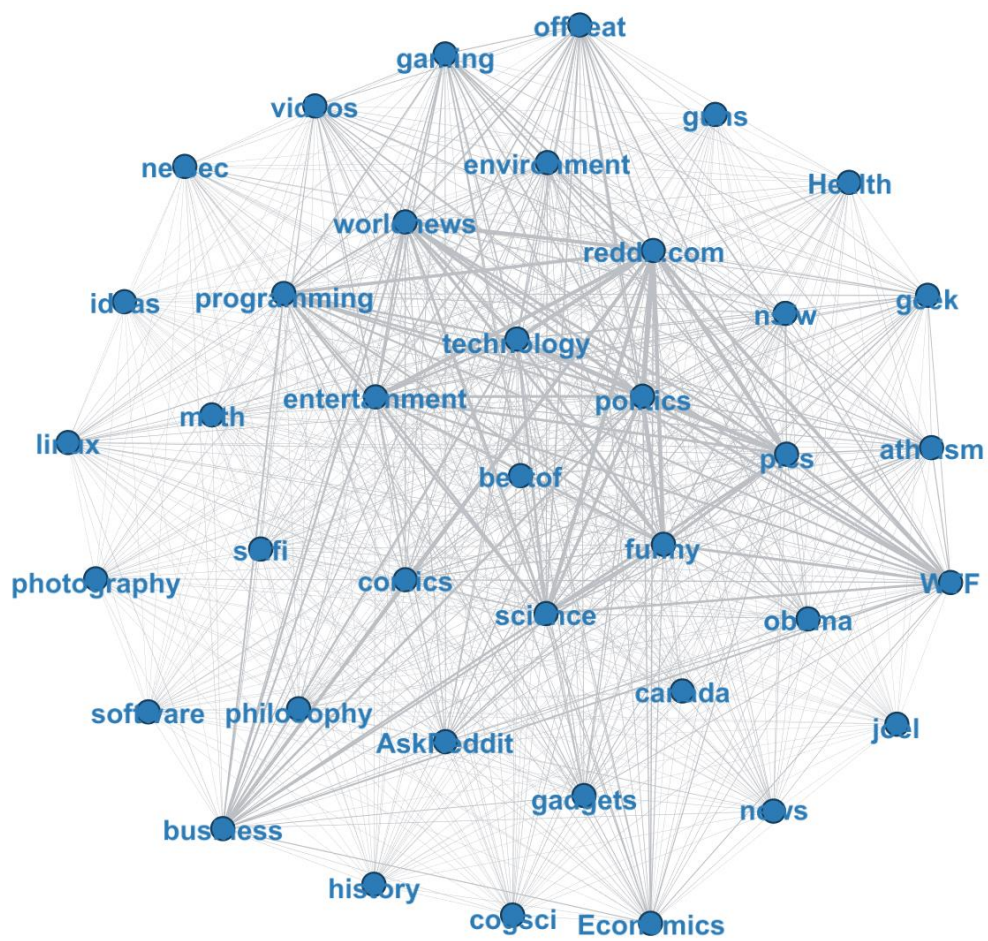
Табела 4-5 Преглед scale-free особине мрежа

	<i>Scale-free</i> мрежа
SNet	ДА (<i>truncated power law</i>)
SNetF	НЕ
SNetT	НЕ
UserNet	НЕ (вредност параметра p није довољно мала)

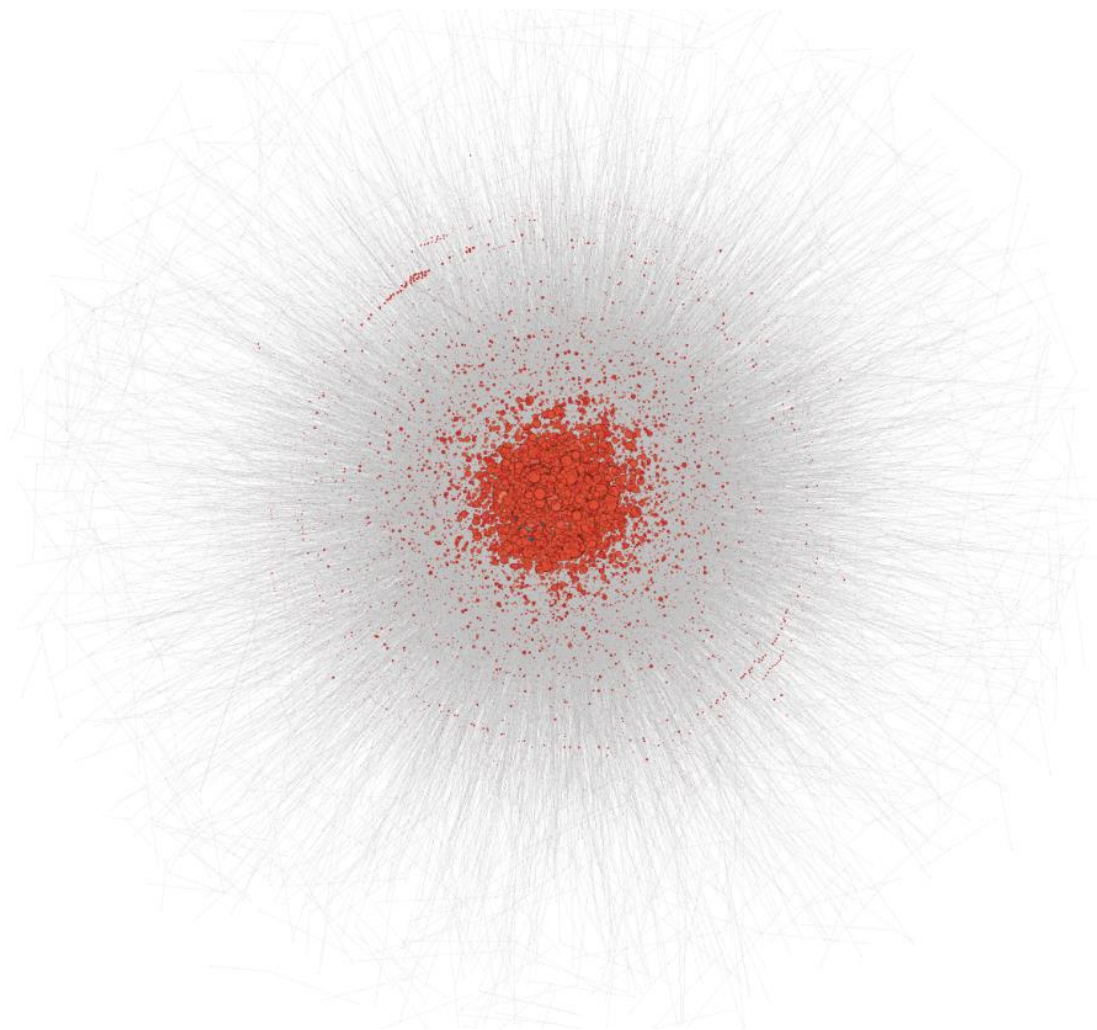
15) Одредити најважније хабове и ауторитете у мрежи. Како су они распоређени и уграђени у мрежу, да ли су на периферији или у језгру мреже?



Слика 4.16 Хабови мреже SNet



Слика 4.18 Хабови мреже SNetT



Слика 4.19 Хабови мреже UserNet

5. АНАЛИЗА МЕРА ЦЕНТРАЛНОСТИ

- 16) Спровести анализе централности по степену, блискости и релационој централности. Дати преглед најважнијих актера по свакој од њих.

Табела 5-1 Преглед најважнијих актера мреже SNet према централности

	DC		CC		BC
reddit.com	0.714558	reddit.com	0.72303	reddit.com	0.247257
technology	0.493317	technology	0.57864	technology	0.03196
politics	0.491169	politics	0.576917	programming	0.029464
pics	0.482578	pics	0.572532	politics	0.028106
funny	0.470167	funny	0.565961	business	0.026988
science	0.467064	science	0.56443	pics	0.02507
entertainment	0.458473	entertainment	0.560464	entertainment	0.024711
worldnews	0.457518	worldnews	0.560117	worldnews	0.021666
programming	0.453461	programming	0.557354	funny	0.020998
WTF	0.451074	WTF	0.556554	science	0.020522

Табела 5-2 Преглед најважнијих актера мреже SNetF према централности

	DC		CC		BC
reddit.com	0.076372	reddit.com	0.076418	reddit.com	0.00124
politics	0.06778	politics	0.068653	programming	0.000397
pics	0.067303	pics	0.068288	pics	0.000309
funny	0.066826	technology	0.067748	politics	0.000293
technology	0.066587	funny	0.067748	funny	0.000258
science	0.06611	science	0.067216	technology	0.000252
worldnews	0.065155	worldnews	0.066692	science	0.000233
WTF	0.064916	WTF	0.066347	WTF	0.00023
entertainment	0.0642	entertainment	0.066006	worldnews	0.000221
programming	0.063007	programming	0.065168	entertainment	0.000199

Табела 5-3 Преглед најважнијих актера мреже SNetT према централности

	DC		CC		BC
reddit.com	1	reddit.com	1	reddit.com	0
scifi	1	scifi	1	scifi	0
gadgets	1	gadgets	1	gadgets	0
nsfw	1	nsfw	1	nsfw	0
news	1	news	1	news	0
environment	1	environment	1	environment	0
atheism	1	atheism	1	atheism	0
canada	1	canada	1	canada	0
Economics	1	Economics	1	Economics	0
bestof	1	bestof	1	bestof	0

Табела 5-4 Преглед најважнијих актера мреже UserNet према централности

	DC		CC		BC
7oby	0.082342833	7oby	0.306775163	7oby	0.039196139
allie	0.066303881	allie	0.298119079	allie	0.023206397
rmuser	0.044584467	tsteele93	0.28500795	rmuser	0.012760779
tsteele93	0.041577164	rmuser	0.2848434	NoMoreNicksLeft	0.011660147
deuteros	0.03828345	qgyh2	0.279841833	mutatron	0.01128532
bobcat	0.033796363	bobcat	0.279393965	qgyh2	0.011136087
NoMoreNicksLeft	0.033319013	deuteros	0.278058921	tsteele93	0.010237132
qgyh2	0.032077903	Aerik	0.274640136	deuteros	0.009183418
mutatron	0.028927395	brainburger	0.273699474	bobcat	0.00876283
Aerik	0.028115901	AMerrickanGirl	0.272857059	revoman	0.008363546

17) Ко су најважнији актери по централности по сопственом вектору? Шта нам то говори о њима?

Централност по сопственом вектору је варијанта централности по степену која у обзир узима и суседство посматраног чвора па се уводе и појмови **утицаја** и **моћи**. Чвор је утицајнији уколико његови суседи исто имају велики број суседа (послате поруке брже стижу до великог броја актера). Вредности које се виде у следећим табелама су управо показатељи утицаја најважнијих актера у мрежама које се посматрају.

Табела 5-5 Централност по сопственом вектору
мреже SNet

	EVC
reddit.com	0.381007
politics	0.314568
pics	0.291901
funny	0.280356
science	0.273919
technology	0.265258
worldnews	0.259681
WTF	0.254369
entertainment	0.251715
programming	0.220964

Табела 5-6 Централност по сопственом вектору
мреже SNetF

	EVC
reddit.com	0.38102883
politics	0.314585862
pics	0.291916494
funny	0.280370786
science	0.273933578
technology	0.265271486
worldnews	0.259694672
WTF	0.25438246
entertainment	0.251727913
programming	0.220974409

Табела 5-7 Централност по сопственом вектору
мреже SNetT

	EVC
reddit.com	0.385713394
politics	0.31828122
pics	0.295063285
funny	0.283340973
science	0.276882272
technology	0.268179974
worldnews	0.262552493
WTF	0.256882934
entertainment	0.254221437
programming	0.223580564

Табела 5-8 Централност по сопственом вектору
мреже UserNet

	EVC
rmuser	0.519411366
qgyh2	0.472877403
alllie	0.4298271
noname99	0.231885008
georgewashingtonblog	0.164664647
7oby	0.158856471
J-Ro	0.118801799
Aerik	0.097444908
tsteele93	0.090397666
glengyron	0.081531568

- 18) Рангирати чворове по Кацовој централности (енг. *Katz centrality*) са варијацијом параметара. При рачунању Кацове централности, експериментисати са додељивањем другачије вредности параметра β за сабредит који се у приложеним CSV фајловима идентификује вредношћу колоне *subreddit* једнаком *reddit.com*. Дати преглед најважнијих актера у случају да је β исто за све сабредите и у случају да је β наведеног сабредита значајно веће.

Табела 5-9 Katz централност мреже SNet

	Katz
reddit.com	0.158408657
politics	0.132766949
pics	0.124498968
funny	0.120102638
science	0.117791082
technology	0.114438667
worldnews	0.112221374
WTF	0.110421676
entertainment	0.109213275
programming	0.097525221

Табела 5-10 Модификована Katz централност мреже SNet

	Katz
reddit.com	0.676970355
politics	0.258119175
pics	0.233637885
funny	0.223710272
science	0.218074649
technology	0.213456068
worldnews	0.206643988
entertainment	0.203024551
WTF	0.201098879
business	0.178295802

Табела 5-11 Katz централност мреже
SNetF

	Katz
reddit.com	0.155917022
politics	0.130733559
pics	0.122603372
funny	0.11827912
science	0.11599702
technology	0.112682631
worldnews	0.110518083
WTF	0.108759516
entertainment	0.107564027
programming	0.096050046

Табела 5-12 Модификована Katz централност мреже
SNetF

	Katz
reddit.com	0.677036116
politics	0.25811284
pics	0.233630354
funny	0.223702787
science	0.218066972
technology	0.213448964
worldnews	0.206636659
entertainment	0.203017944
WTF	0.201091349
business	0.178289807

Табела 5-13 Katz централност мреже

SNetT

	Katz
reddit.com	0.349150548
politics	0.294002268
pics	0.275858594
funny	0.266384618
science	0.26145428
technology	0.254270987
worldnews	0.249623674
WTF	0.245427162
entertainment	0.242898295
programming	0.218446513

Табела 5-14 Модификована Katz централност мреже

SNetT

	Katz
reddit.com	0.697572754
politics	0.254933904
pics	0.230115041
funny	0.22024902
science	0.21468633
technology	0.210360266
worldnews	0.203462022
entertainment	0.199956438
WTF	0.19771756
business	0.175884733

Табела 5-15 Katz централност мреже UserNet

	Katz
allie	0.007511999
rmuser	0.007466094
qgyh2	0.00723921
7oby	0.007189754
georgewashingtonblog	0.007040982
tsteele93	0.007036377
J-Ro	0.007026781
deuteros	0.007026194
dons	0.007015367
bobcat	0.007013088

19) На основу претходна три питања предложити и конструисати хеуристику (композитну меру централности) за проналажење најважнијих актера и пронаћи их. Обратите пажњу на тип мреже који се анализира (усмерена или неусмерена) и, сходно томе, прилагодити колико различите мрежне метрике утичу на хеуристику.

За израчунавање композитне централности коришћена је следећа формула:

$$composite_rank = DC_rank * CC_rank * BC_rank * EVC_rank$$

Табела 5-16 Композитна централност мреже SNet

	Composite Rank
reddit.com	1
technology	48
politics	72
pics	288
funny	900
science	1800
programming	2430
entertainment	3087
worldnews	3584
business	6655

Табела 5-17 Композитна централност мреже SNetF

	Composite Rank
reddit.com	1
politics	32
pics	81
funny	360
technology	810
science	1260
programming	2000
worldnews	3087
WTF	4096
entertainment	7290

Табела 5-18 Композитна централност мреже SNetT

	Composite Rank
reddit.com	8000
politics	16000
pics	24000
funny	32000
science	40000
technology	48000
worldnews	56000
WTF	64000
entertainment	72000
programming	80000

Табела 5-19 Композитна централност мреже UserNet

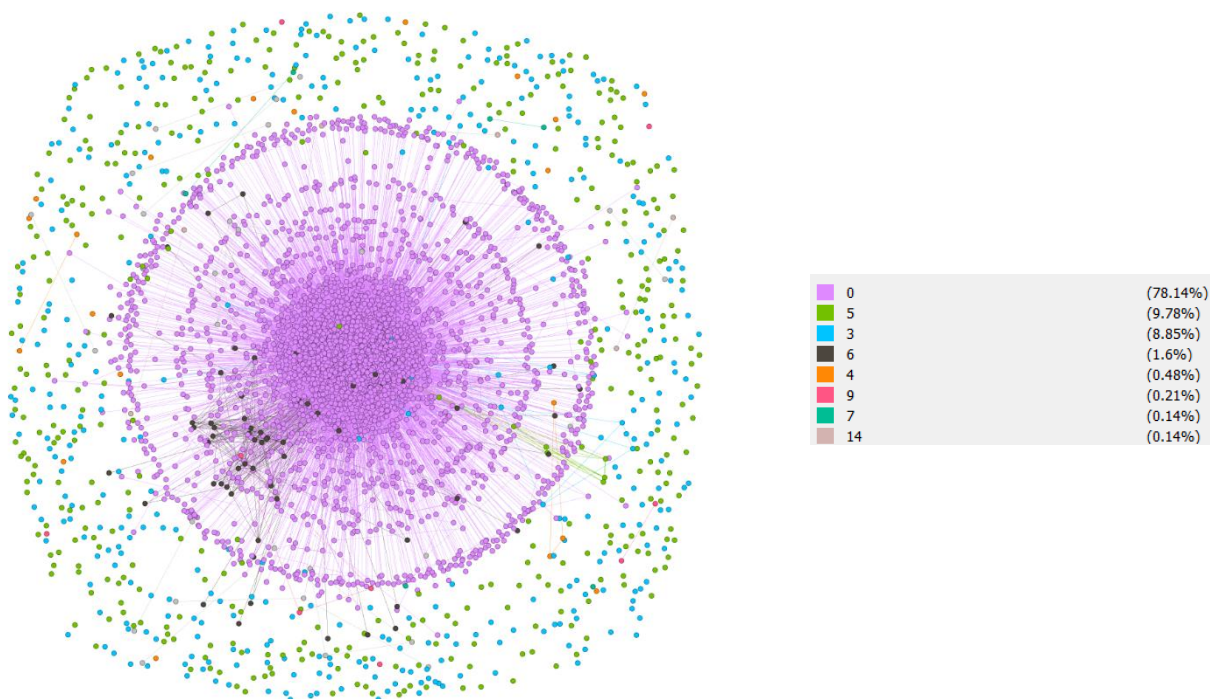
	Composite Rank
7oby	6
allie	24
rmuser	36
qgyh2	480
tsteele93	756
deuteros	4760
bobcat	4860
Aerik	7680
glengyron	25080
AMerrickanGirl	35420

6. ДЕТЕКЦИЈА КОМУНА

20) Ако величина мреже дозвољава, спектралном анализом или анализом дендрограма проценити потенцијалне кандидате за број комуна у мрежи.

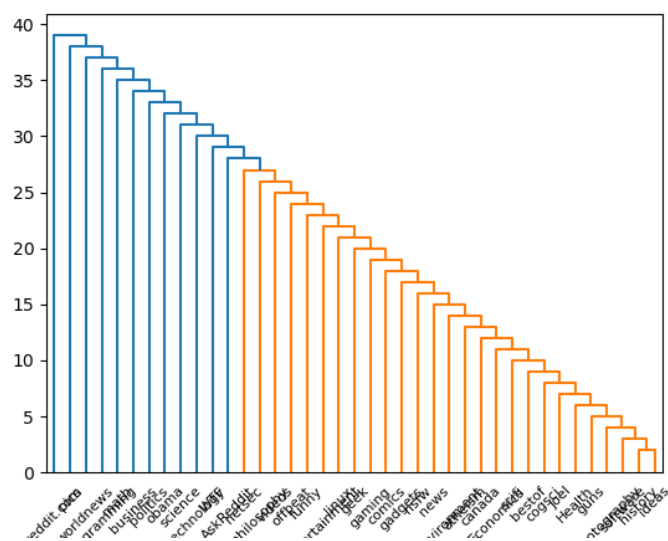
- **SNet мрежа**

Спектралном анализом мреже за број кластера у опсегу од 2 до 30 примећено је да се јасно издвајају 3 комуне.



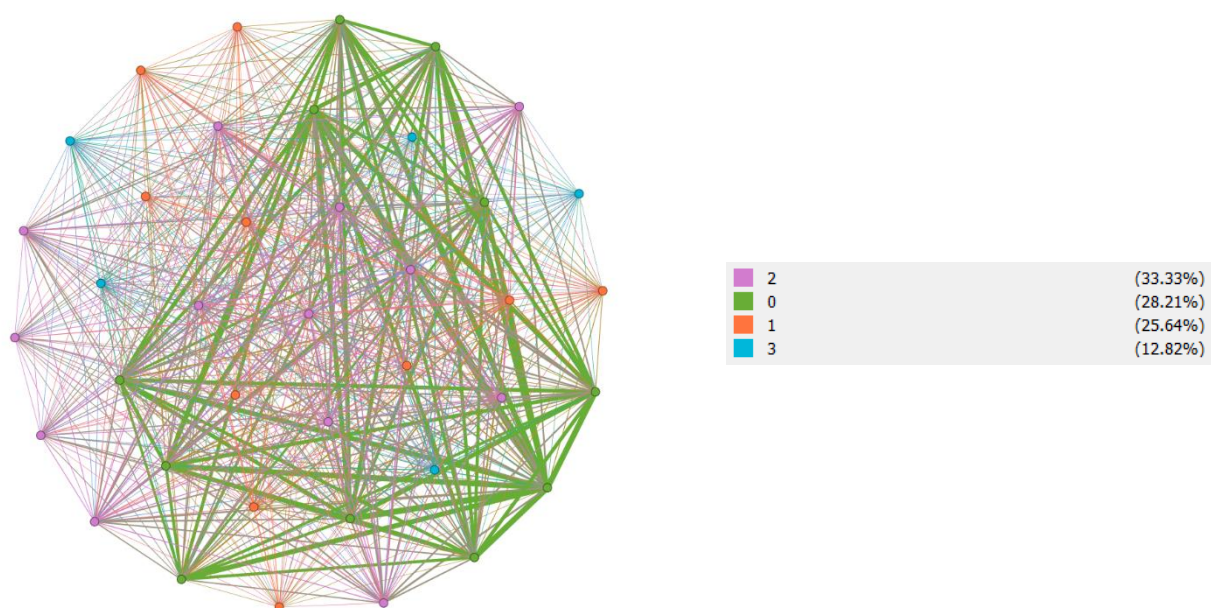
Слика 6.1 Спектрална анализа мреже SNet

- *SNetT* мрежа



Слика 6.2 Дендрограм *SNetT* мреже

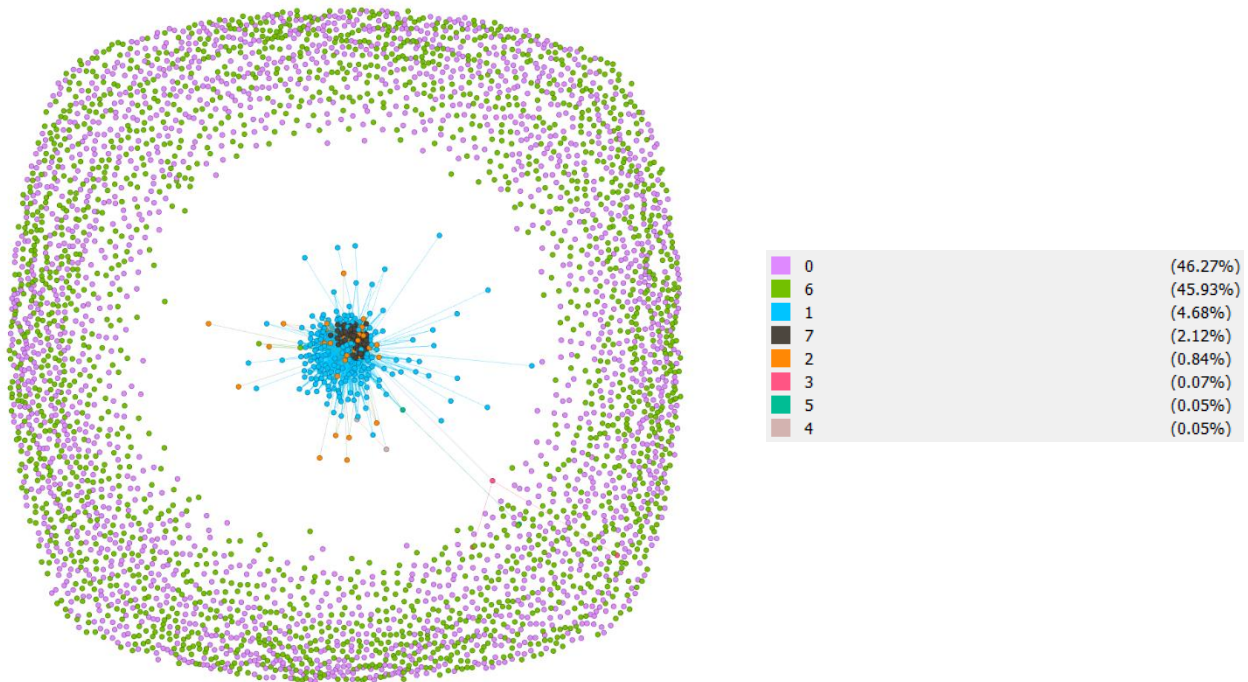
Како мрежа *SNetT* представља потпун граф, и сам дендрограм је симетричног облика. Подела на комуне код ове мреже и нема много смисла, цела мрежа је тачно једна комуна. Међутим, спектралном анализом мреже за број кластера у опсегу од 2 до 30 примећено је да се јасно издвајају 4 комуне.



Слика 6.3 Спектрална анализа мреже *SNetT*

- **SNetF мрежа**

Спектралном анализом мреже за број кластера у опсегу од 2 до 30 примећено је да се јасно издвајају 2 комуне, с тим да се за неко k појаве и највише две комуне величине око 100 актера које настају цепањем поменутих доминантинх комуна (око 1900 чланова).



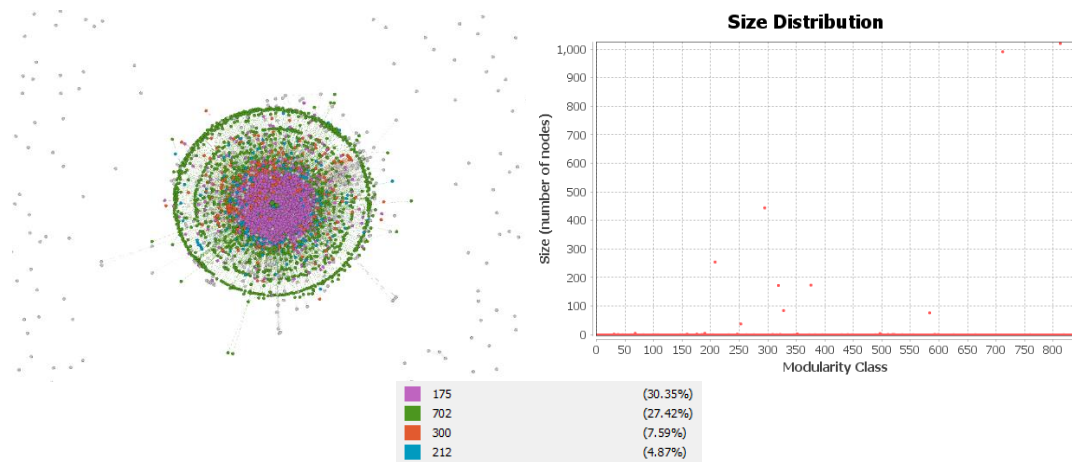
Слика 6.4 Спектрална анализа мреже SNetF

- **UserNet мрежа**

Мрежа корисника је превеликог обима да би спектрална анализа или анализа дендрограма могла да се изврши у разумном времену.

- 21) Спровести кластерисање Лувенском методом (максимизацијом модуларности) у алату *Gephi* за три различите вредности параметра резолуције. Конструисати визуелизације и дискутовати избор параметра резолуције на добијено кластерисање (број и величина кластера).

- ***SNet* мрежа**



Слика 6.5 Кластерисање мреже *SNet* Лувенском методом

Кластерисањем мреже *SNetF* Лувенском методом издвојиле су се четири комуне веће од осталих. Чворови исте комуне обојени су истом бојом као на слици изнад. Највећа детектована комуна броји 30,35% чворова, наредна 27,42%, а трећа највећа 7,59%.

Коришћени параметри:

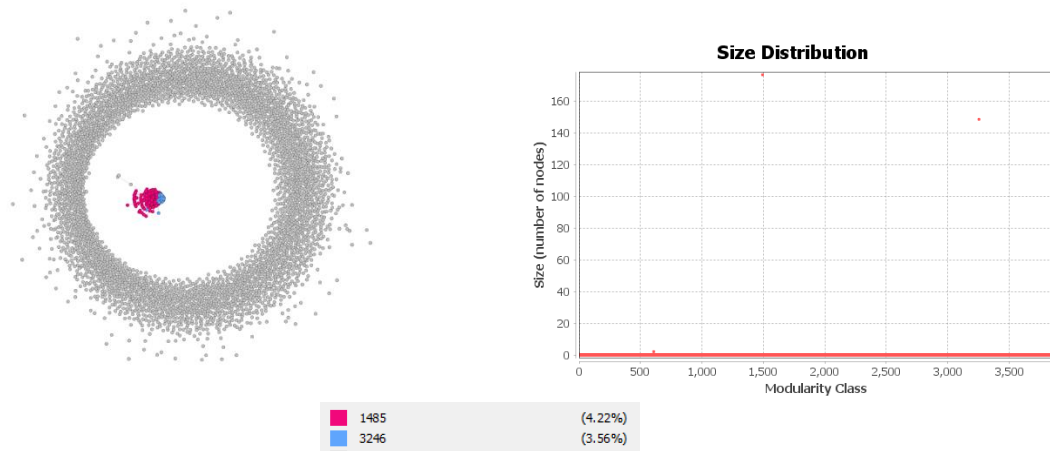
- Randomize: On
- Use edge weights: On
- Resolution: 0.9 – одабрана је вредност како би се детектовао већи број мањих комуна. За све вредности овог параметра веће од један, издваја се само једна комуна са приметно већим бројем елемената, док све остале садрже не више од 0,05% укупног броја чворова.

Резултат:

- Modularity: 0.095
- Modularity with resolution: 0.051
- Number of Communities: 836

- ***SNetT* мрежа** – мрежа је потпун граф, дакле и једна комплетна комуна/кластер

- *SNetF* мрежа



Слика 6.6 Кластерисање мреже *SNetF* Лувенском методом

Издвојиле су се две веће комуне кластерисањем мреже *SNetF* Лувенском методом. Чворови једне комуне обојени су розе бојом и обухватају 4,22%, односно 176 чворова. Друга комуна, представљена је чворовима плаве боје, којих има 3,56%, односно 146.

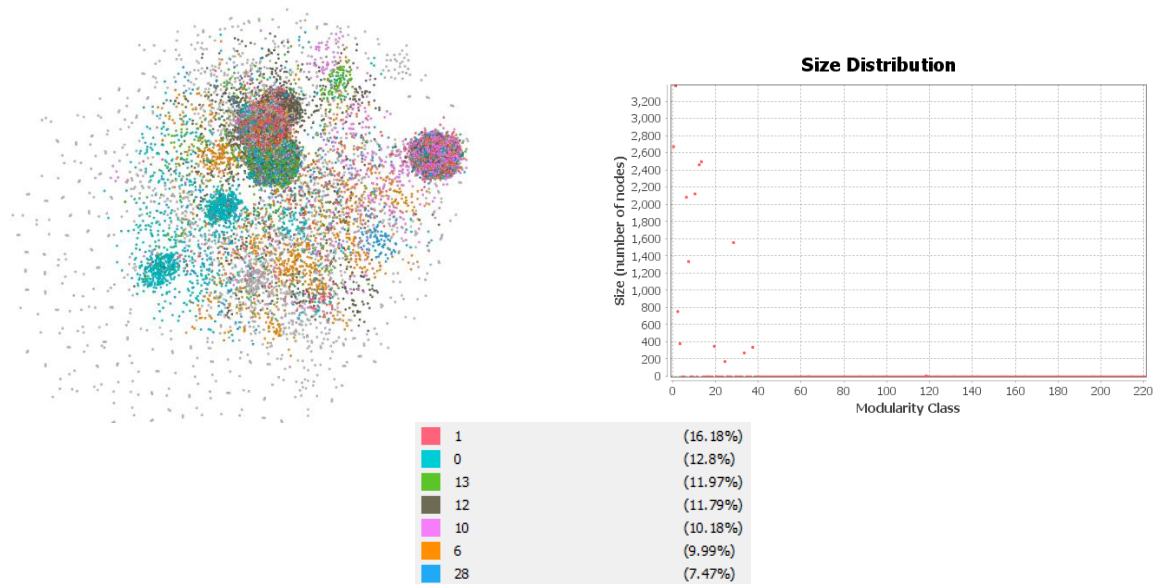
Коришћени параметри:

- Randomize: On
- Use edge weights: On
- Resolution: 1.0 – За вредности овог параметра мање од одабране детектује се нешто више комуна, али је њихова величина мања и приближно иста, док се за параметре веће од један, издваја само једна већа комуна, а све остале садрже мање од 0,05% укупног броја чворова.

Резултат:

- Modularity: 0.080
- Modularity with resolution: 0.080
- Number of Communities: 3865

- **UserNet мрежа**



Слика 6.7 Кластерисање мреже UserNet Лувенском методом

Кластерисањем мреже *UserNet* Лувенском методом, издвојено је шест комуна са већим бројем чворова у односу на остале. Комуна су обојене различитим бојама, а највећа обухвата 16,18% свих чворова мреже.

Коришћени параметри:

- Randomize: On
- Use edge weights: On
- Resolution: 1.08 – За вредности овог параметра мање од одабране детектује се нешто више комуна, али је њихова величина мања и приближно иста, док се за параметре веће од одабраног, издваја само једна већа комуна, а све остале садрже мање од 0,05% укупног броја чворова.

Резултат:

- Modularity: 0.313
- Modularity with resolution: 0.349
- Number of Communities: 221

22) Које заједнице (комуна) се могу уочити приликом анализе мреже? Да ли постоји неко објашњење за детектоване комуна?

- **SNet мрежа** – Није уочена правилност

- **SNetT мрежа** – Нема детектованих комуна. Мрежа је потпун граф, те су сви чворови једнаким, максималним бројем грана повезани са осталим чворовима.
- **SNetF мрежа** – Не уочава се нека изражена правилност. Доста (описно) сличних сабредита припада различитим комунама. На пример, сабредит *Linux* припада једној комуни, док сабредит *Ubuntu* припада другој. Слична ситуација је и са паровима *Django – Python*, *Celebrity – Celebrities*, *Bookclub – Books*. Са друге стране може се приметити да у мањој (плавој) комуни има нешто израженији број сабредита политичко-економских тема (*AmericanGovernment*, *AmericanPolitics*, *Anarchism*, *conspiracies*, *conspiracy*, *Economics*, *economy*, *government*, *news*, *obama*, *ukpolitics*, *usa*, *worldpolitics*, *WorldWideNews*).
- **UserNet мрежа** – корисници који су имали највише интеракција групицани су у комуне. Вероватно би даља анализа могла да покаже да корисници исте комуне припадају и великом броју истих сабредита.

23) Ко су актери који се могу окарактерисати као кључни брокери (мостови) у мрежи? Шта их чини брокерима?

Брокере у мрежи идентификујемо на основу метрика **релационе централности** и **мрежних ограничења**. Мрежна ограничења нам говоре да што су суседи актера боље повезани, то су његове могућности за деловање више редуковане. Самим тим брокери у мрежи имају високу релациону централност и ниска мрежна ограничења.

Увешћемо појам **брокер коефицијента** који ћемо рачунати на следећи начин:

$$\text{релациона_централност} + (1.0 - \text{мрежно_ограничење})$$

- ***SNet* мрежа**

Израчунавање мрежног ограничења је занемарено у овом случају услед величине посматране мреже (комплексност израчунавања је превелика). Из тог разлога је у анализи мостова разматрана само релациона централност.

Табела 6-1 Брокери мреже SNet

	Betweenness Centrality	Constraint	Broker Coefficient
reddit.com	0.216913	1	0.216913
technology	0.026364	1	0.026364
programming	0.025611	1	0.025611
business	0.025554	1	0.025554
entertainment	0.021284	1	0.021284
politics	0.020419	1	0.020419
pics	0.019917	1	0.019917
worldnews	0.017213	1	0.017213
funny	0.015591	1	0.015591
science	0.014692	1	0.014692

- ***SNetT* мрежа**

Табела 6-2 Брокери мреже SNetT

Subreddit	Betweenness Centrality	Constraint	Broker Coefficient
joel	0.756757	0.186713	1.570044
ideas	0.687055	0.150257	1.536799
software	0.000000	0.155381	0.844619
history	0.000000	0.155587	0.844413
cogsci	0.000000	0.158193	0.841807

philosophy	0.000000	0.158330	0.841670
guns	0.000000	0.159231	0.840769
photography	0.000000	0.159650	0.840350
scifi	0.000000	0.160525	0.839475
Health	0.000000	0.162094	0.837906

- ***SNetF*** мрежа

Табела 6-3 Брокери мреже SNetF

Subreddit	Betweenness Centrality	Constraint	Broker Coefficient
government	0.000022	0.075754	0.924268
usa	0.000003	0.076281	0.923721
Pictures	0.000009	0.077097	0.922911
psychology	0.000006	0.077425	0.922582
lists	0.000007	0.077832	0.922175
howto	0.000005	0.077919	0.922085
history	0.000041	0.078669	0.921372
Art	0.000038	0.079210	0.920828
USPE08	0.000001	0.079206	0.920795
Frugal	0.000024	0.079351	0.920673

- ***UserNet*** мрежа

Израчунавање мрежног ограничења је занемарено у овом случају услед величине посматране мреже (комплексност израчунавања је превелика). Из тог разлога је у анализи мостова разматрана само релациона централност.

Табела 6-4 Брокери мреже UserNet

User	Betweenness Centrality	Constraint	Broker Coefficient
7oby	0.018202	1	0.018202
allie	0.01027	1	0.01027
jordanlund	0.010068	1	0.010068
mutatron	0.009445	1	0.009445
Pryorra	0.008865	1	0.008865
tsteele93	0.008328	1	0.008328
bobcat	0.007335	1	0.007335
deuteros	0.007257	1	0.007257
stesch	0.007105	1	0.007105
Aerik	0.006658	1	0.006658

7. ПОРЕЂЕЊЕ SNet и SNetT МРЕЖА

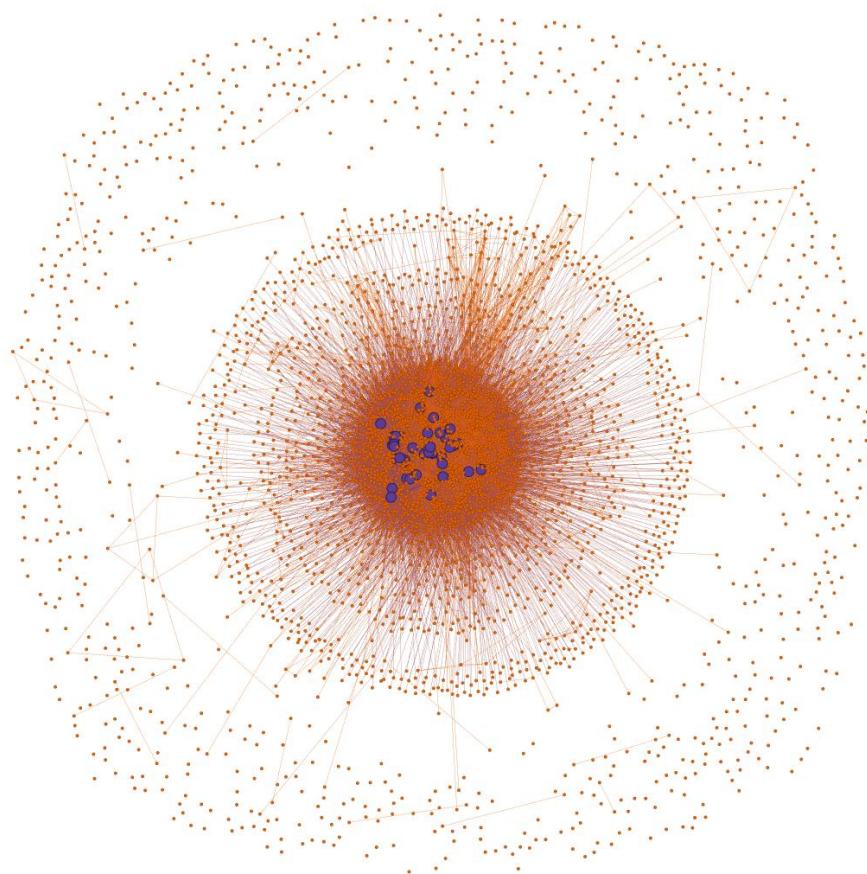
24) Упоредити карактеристике SNet и SNetT мрежа. Коментарисати потенцијалне разлике и проценити да ли су сабредити из SNetT активнији и боље повезани од остатка мреже.

Посматрајући карактеристике мрежа SNet и SNetT из поглавља **4. Основна анализа моделованих мрежа** можемо уочити велике разлике између њих. Самим тим што је мрежа SNetT потпун повезан граф и метрике унутар ње биће очекивано довољно другачије од мреже SNet. Ако упоредимо сабредите по степену чвора, можемо приметити да су они заједнички за обе мреже јако добро повезани и са осталим сабредитима. Табела у наставку, приказује 25 најбоље повезаних сабредита у SNet мрежи. Једини сабредит из ове групе, који не припада и SNetT мрежи је сабредит *Music* обележен жутом бојом.

Табела 7-1 Најбоље повезани сабредити у SNet мрежи

Label	Degree
reddit.com	2994
technology	2067
politics	2058
pics	2022
funny	1970
science	1957
entertainment	1921
worldnews	1917
programming	1900
WTF	1890
business	1848
offbeat	1554
gaming	1500
AskReddit	1449
Economics	1411
environment	1405
videos	1369
gadgets	1352
comics	1342
atheism	1331
geek	1296
Music	1292
nsfw	1291
obama	1213

25) Како су распоређени чворови из SNetT у оквиру SNet мреже? Да ли припадају језгру или периферији или су мешовито распоређени?



Слика 7.1 Чворови мреже SNetT унутар мреже SNet

Чворови из циљане групе сабредита се налазе у самом језгру мреже свих сабредита где су груписани актери са већим степеном чвора. Ти сабредити дефинитвно припадају групи боље повезаних чворова и стиче се утисак да су то једни од најактивнијих сабредита на платформи у 2008. години, што и није толико изненађење с обзиром на важност и утицај теме као што је светска економска криза.

Списак слика

Слика 2.1 Мрежа SNet.....	5
Слика 2.2 Расподела тежина грана.....	6
Слика 2.3 Мрежа SNetF.....	7
Слика 2.4 Мрежа SNetT	8
Слика 2.5 Мрежа UserNet.....	9
Слика 3.1 Пирсонов коефицијент корелације.....	12
Слика 4.1 Величина компонената мреже SNet	14
Слика 4.2 Величина компонената мреже SNetF	14
Слика 4.3 Величина компонената мреже SNetT	15
Слика 4.4 Величина компонената мреже UserNet	15
Слика 4.5 Расподела коефицијента кластеризације мреже SNet и случајне мреже истих димензија	16
Слика 4.6 Расподела коефицијента кластеризације мреже SNetT и случајне мреже истих димензија	17
Слика 4.7 Расподела коефицијента кластеризације мреже SNetF и случајне мреже истих димензија	17
Слика 4.8 Расподела коефицијента кластеризације мреже UserNet и случајне мреже истих димензија	18
Слика 4.9 Коефицијент клуба богатих мреже SNet.....	20
Слика 4.10 Коефицијент клуба богатих мреже SNetF.....	20
Слика 4.11 Коефицијент клуба богатих мреже SNetT.....	21
Слика 4.12 Дистрибуција чворова по степену мреже SNet	21
Слика 4.13 Дистрибуција чворова по степену мреже SNetF	22
Слика 4.14 Дистрибуција чворова по степену мреже SNetT	22
Слика 4.15 Дистрибуција чворова по степену мреже UserNet	23
Слика 4.16 Хабови мреже SNet.....	24
Слика 4.17 Хабови мреже SNetF.....	25
Слика 4.18 Хабови мреже SNetT.....	26
Слика 4.19 Хабови мреже UserNet.....	27
Слика 6.1 Спектрална анализа мреже SNet	37
Слика 6.2 Дендрограм SNetT мреже.....	38
Слика 6.3 Спектрална анализа мреже SNetT	38
Слика 6.4 Спектрална анализа мреже SNetF	39
Слика 6.5 Кластерисање мреже SNet Лувенском методом	40
Слика 6.6 Кластерисање мреже SNetF Лувенском методом	41
Слика 6.7 Кластерисање мреже UserNet Лувенском методом	42
Слика 7.1 Чворови мреже SNetT унутар мреже SNet.....	48

Списак табела

Табела 1-1 Утицај чишћења на величину скупа података.....	4
Табела 2-1 Промена броја грана приликом филтрације	6
Табела 3-1 Анализа сабредита	10
Табела 3-2 Анализа корисника.....	11
Табела 3-3 Најактивнији корисници	11
Табела 3-4 Објаве са највећим бројем коментара.....	12
Табела 4-1 Густина и дијаметар мреже.....	13
Табела 4-2 Број повезаних компонената мреже.....	13
Табела 4-3 Коефицијент кластеризације мреже	16
Табела 4-4 Асортативност по степену чвора	19
Табела 4-5 Преглед scale-free особине мрежа.....	23
Табела 5-1 Преглед најважнијих актера мреже SNet према централности.....	28
Табела 5-2 Преглед најважнијих актера мреже SNetF према централности.....	29
Табела 5-3 Преглед најважнијих актера мреже SNetT према централности.....	29
Табела 5-4 Преглед најважнијих актера мреже UserNet према централности.....	30
Табела 5-5 Централност по сопственом вектору мреже SNet	31
Табела 5-6 Централност по сопственом вектору мреже SNetF	31
Табела 5-7 Централност по сопственом вектору мреже SNetT	31
Табела 5-8 Централност по сопственом вектору мреже UserNet	31
Табела 5-9 Katz централност мреже SNet	33
Табела 5-10 Модификована Katz централност мреже SNet.....	33
Табела 5-11 Katz централност мреже SNetF	33
Табела 5-12 Модификована Katz централност мреже SNetF	33
Табела 5-13 Katz централност мреже SNetT	34
Табела 5-14 Модификована Katz централност мреже SNetT	34
Табела 5-15 Katz централност мреже UserNet	34
Табела 5-16 Композитна централност мреже SNet.....	35
Табела 5-17 Композитна централност мреже SNetF.....	35
Табела 5-18 Композитна централност мреже SNetT	36
Табела 5-19 Композитна централност мреже UserNet.....	36
Табела 6-1 Брокери мреже SNet.....	44
Табела 6-2 Брокери мреже SNetT.....	44
Табела 6-3 Брокери мреже SNetF.....	45
Табела 6-4 Брокери мреже UserNet.....	46
Табела 7-1 Најбоље повезани сабредити у SNet мрежи	47