УНИВЕРЗИТЕТ У БЕОГРАДУ

ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

АНАЛИЗА СОЦИЈАЛНИХ МРЕЖА

Пројектни задатак

Предметни професори:

др Марко Мишић, доцент

др Јелица Протић, редовни професор

Предраг Обрадовић, асистент

Студент:

Марко Милићевић 3136/20

Београд, Јул 2022.

Садржај

1. ЧИШЋЕЊЕ ПОДАТАКА

Пре анализе, потребно је очистити и средити дате податке. Дати су нам подаци о *submission* и *comment*, односно *csv* табеле које садрже информације о објавама и коментарима на друштвеној мрежи „Reddit" у току 2008. године. Прва група садржи податке као што су: id, url, permalink, author, назив *subreddit-a*, док друга садржи: идентификацију коментара, аутора, идентификација родитеља који може бити или други коментар или објава, као и идентификација објаве на којој је коментар и назив *subreddit-a*.

Најпре су проверене основне информације, постоје ли празна поља у идентификационим колонама. Табела објава је била „чиста", док је међу коментарима постојао само један ред са празном идентификацијом. Како је то само један ред од седам милиона, одлучено је да је могуће једноставно уклонити тај ред. Поновљена је провера јединствености, овог пута за URL, који се показао уникатним у обе табеле. Посматрајући табеле, уочено је да постоје аутори означени са *„[deleted]"*, што значи да ови аутори више нису корисници мреже. Овакве ауторе је неопходно уклонити из обе табеле, након чега је одлучено понављање провере јединствености и степена попуњености. Уклоњених коментара било је око 2.5 милиона. Следећи корак је био преименовање идентификација *id* у нешто мени дескриптивније – *submissions_id* и *comments_id.*

У неким даљим фазама пројекта је запажено понављање префикса „t3_, *t1_"* у вредностима колона „link_id" и *„parent_id"*, који су након доказа хипотезе да су искључиво ово префикси, одстрањени из својих колона. Такође, пронађени су коментари који нису остављани на постојећим објавама, као и коментари који су остављани на непостојећим коментарима. Ово је, вероватно, последица коментарисања објава и коментара из претходних година, на пример у 2008. години коментарисати објаву из 2007. године, или коментарисања објава и коментара који су у ранијим итерацијама чишћења уклоњени. Како ови подаци не фигурирају, уклоњени су и они.

Брисањем неискоришћених колона, табеле су постале доста уредније и прегледније. Ова филтрација настала је након одређеног броја итерација других тачака пројекта.

```
comments = comments.drop(["distinguished", "gilded", "controversiality"], axis=1)
submissions = submissions.drop(["distinguished", "domain", "stickied", "locked", "hide_score"], axis=1)
```

Слика 2.1.1 Некоришћене колоне

Резултати чишћења података су следећи:

Табела 1.1.1.1 Приказ филтрирања података

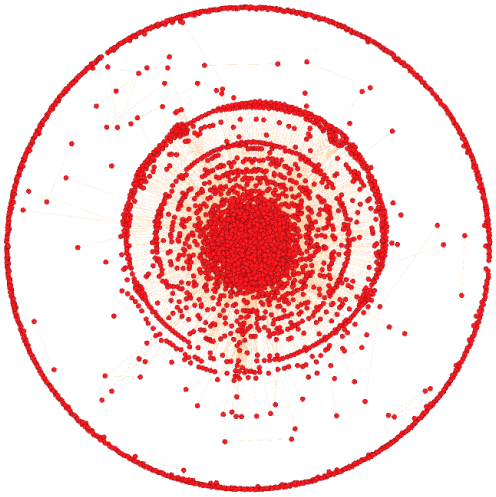
	Пре филтрације	После филтрације
Submissions	2519853	2044810
Comments	7242871	3210167

2. МОДЕЛОВАЊЕ МРЕЖА

У овом поглављу ће бити приказани графови, након филтрирања података. За ову тачку, коришћен је алата „Gephi”. Потребно је да се направе четири мреже: SNet, SNetF, SNetT, UserNet.

2.1. SNet – subreddit network

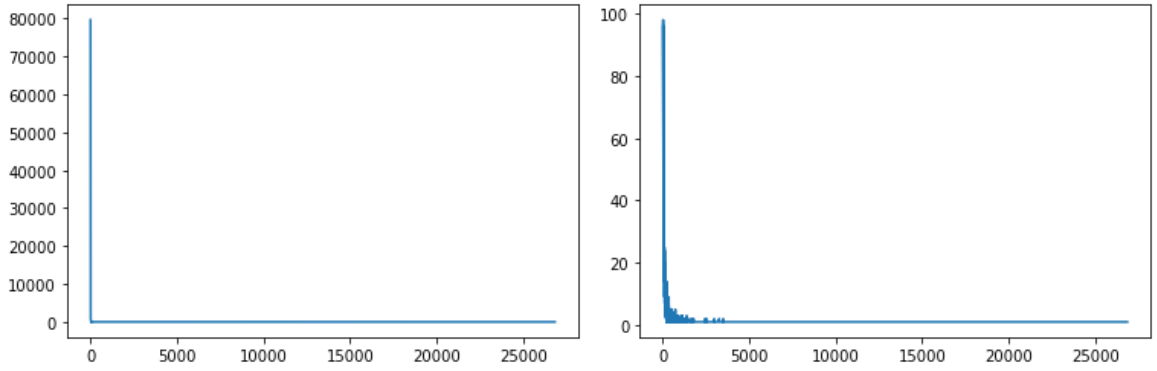
Мрежа је представља неусмерени тежински граф. Чворова, сабредита, има 4191, а ивица, корисника који су интераговали са два subreddit-a, 135974. Тежине представљају збир појављивања ивица.



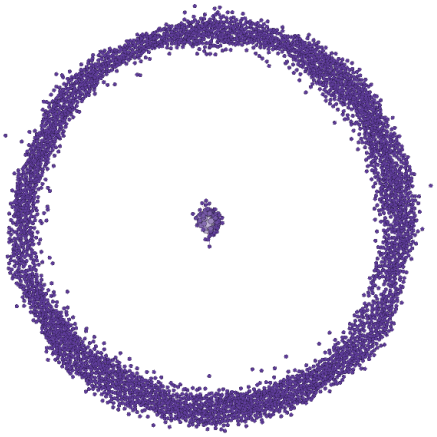
Слика 2.1.1 модел SNet

2.2. SNetF – subreddit network filtered

Слично претходној, мрежа је представља неусмерени тежински граф. Чворова, subreddita, има 4191, а ивица, корисника који су интераговали са два subreddit-a, 6113. Број грана је филтриран такав да су присутне само гране са тежином већом од *w*. Детаљан процес избора се налази у „Network creation.ipynb”. Већина грана су имале тежину 1 или 2. Тежине до 41 чине чак 95.66% свих грана. Из тог разлога је *w=41*, након тог броја нема троцифреног појављивања истих тежина. Тежине представљају збир појављивања ивица, као и код SNet.



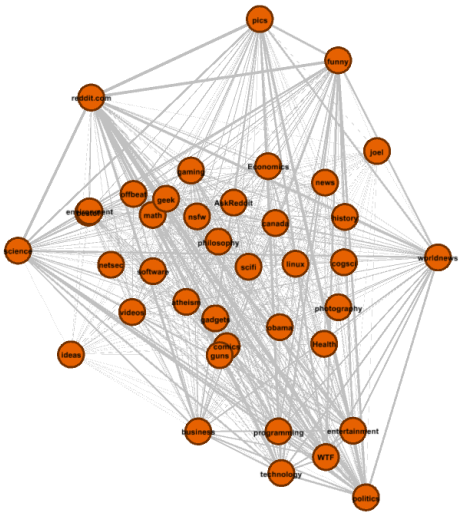
Слика 2.2.1 Дистрибуција тежина



Слика 2.2.2 Модел SNetF

2.3. SNetT – subreddit network targeted

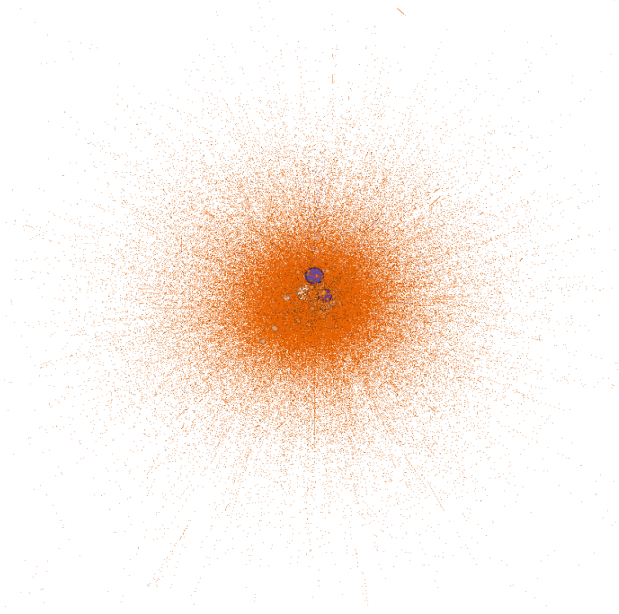
Слично претходним мрежама, ова мрежа је представља неусмерени тежински граф. Чворова, subreddita, има 39, а ивица, корисника који су интераговали са два subreddit-a, 741. Тежине представљају збир појављивања ивица, као и код Sent. Ова мрежа је потпуна, односно сваки чвор је повезаним са свим другим. Листа чворови је експлицитно дефинисина у поставци задатка.



Слика 2.3.1 Модел SNetT

2.4. UserNet – User network

Ова мрежа се разликује у потпуности од претходних, пре свега јер је усмерена. Чворова, корисника који су коментарисали објаву или коментар другим корисницима, има 18975, а грана 102043.



Слика 2.4.1 Модел UserNet

3. ИСТРАЖИВАЧКА ПИТАЊА

3.1. Статистичка анализа

3.1.1. Колико постоји различитих subreddit-а који се појављују у посматраном периоду? Који су најважнији по броју корисника, а који по броју коментара?

Укупан број subreddita је 4191 (број чворова SNet-а).

Табела 3.1.1.1 Преглед најважнијих сабредита

Subreddit	Author count
reddit.com	158641
politics	34902
pics	26759
technology	26070
funny	25615
entertainment	24717
programming	23629
business	23347
science	23332
worldnews	22472

Subreddit	Comment count
reddit.com	738291
politics	526425
programming	255419
pics	187328
science	161279
worldnews	150029
WTF	150029
funny	114429
technology	99609
entertainment	96118

3.1.2. Какав је просечан број забележених корисника активних у посматраном поериоду по субредиту? Корисник се сматра активним ако је забележио барем један коментар или објава тог корисника на субредиту.

Просечан број корисника по субредиту је: **141.32283.**

3.1.3. Ко су корисници са највећим бројем објава, а ко корисници са највећим бројем коментара?

Табела 3.1.3.1 Преглед најактивнијих корисника по типу активности

Users with the most submissions:

	author	counts
84823	gst	18870
141813	qgyh2	12238
147359	rmuser	9822
173691	twolf1	8597
13172	IAmperfectlyCalm	8308
141766	qazamisan	6927
54960	charlatan	5998
90683	igeldard	5373
130852	noname99	5334
64933	democracy101	5332

Users with the most comments:

	author	counts
11308	NoMoreNicksLeft	8465
51139	malcontent	7423
52040	matts2	6975
567	7oby	5640
52943	mexicodoug	5499
55356	mutatron	4980
62351	randomb0y	4896
18866	aletoledo	4880
58533	otakucode	4744
19141	alllie	4734

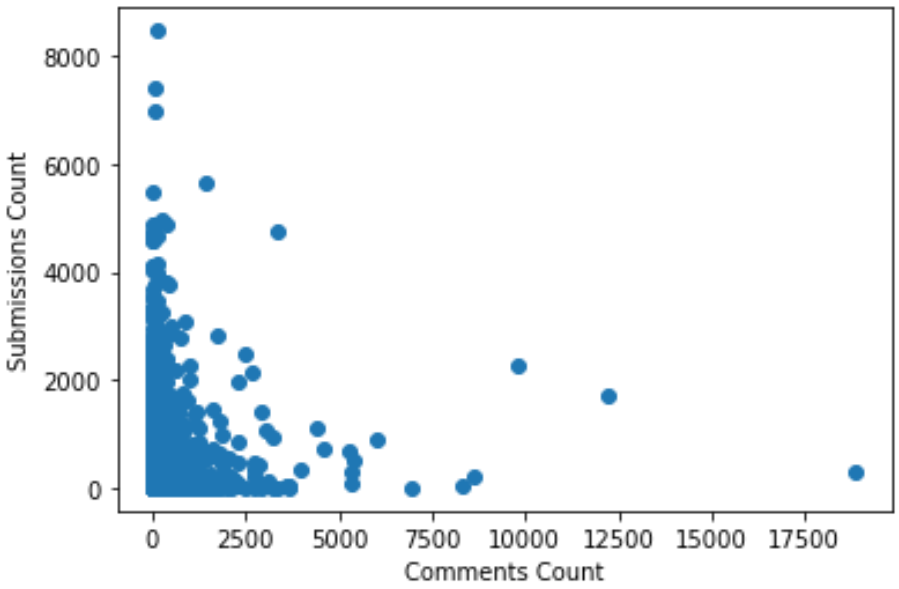
3.1.4. Који корисници су активни на највећем броју субредита? На колико субредита су активни?

Табела 3.1.4.1 Преглед најактивнијих корисника и број субредита

author	subreddit
Escafane	152
MrKlaatu	147
omfgninja	111
scientologist2	106
codepoet	103
krugerlive	102
turkourjurbs	101
tuoder	101
bobcat	98
b34nz	97

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

Пирсонов коефицијент је: **0.155797.**



Слика 3.1.1 Пирсонова расподела

3.1.6. Које објаве поседују највећи број коментара и на којим су субредитима постављене? Приказати податке о тим објавама, укључујући то на којем су сабредиту постављене и шта им је садржај (ако је поље објаве „over18“ постављено на false).

Табела 3.1.6.1 Објаве са највише коемнтара

	author	subreddit	url	permalink	score	counts
154071	zekel	science	http://hundredpushups.com	/r/science/comments/6nz1k/got_six_weeks_try_th...	1621	1313
27108	matiasklein	reddit.com	https://www.reddit.com/r/reddit.com/comments/6...	/r/reddit.com/comments/675oj/post_the_funniest...	1098	1225
347032	willjohnston	politics	https://www.reddit.com/r/politics/comments/7be...	/r/politics/comments/7beo2/obama_wins_the_pres...	8538	1074
27334	rpi22	reddit.com	http://www.washingtonpost.com/wp-dyn/content/c...	/r/reddit.com/comments/676ja/new_study_confirm...	669	961
244944	georgeb	programming	http://www.google.com/chrome	/r/programming/comments/6z9op/chrome_is_here/	1904	913
428848	Schlichten	worldnews	http://www.dailykos.com/story/2008/12/28/11443...	/r/worldnews/comments/7m6m4/today_i_end_my_sup...	1589	904
200875	TheRealStick	politics	https://www.reddit.com/r/politics/comments/6tv...	/r/politics/comments/6tvaz/im_a_bleedingheart_...	788	840
162829	jtmarlin	reddit.com	http://www.nytimes.com/aponline/business/AP-Sc...	/r/reddit.com/comments/6p30u/supreme_court_rul...	1071	787
285037	IM_A_REPTILIAN	politics	http://www.msnbc.msn.com/id/26884523/?	/r/politics/comments/7488a/bailout_does_not_pa...	3361	765
133321	valeriepieris	reddit.com	http://www.txstudy.org/please.html	/r/reddit.com/comments/6lbr7/is_reddit_really_...	492	762

3.2. Основна карактеризација моделованих мрежа

3.2.1. Колика је густина мреже?

Табела 3.2.1.1 Густине мрежа

Мрежа:	SNet	SNetF	SNetT	UserNet
Густина:	0.015	0.001	1.0	0.0

3.2.2. Колике су просечне дистанце у оквиру мреже и дијаметар мреже?

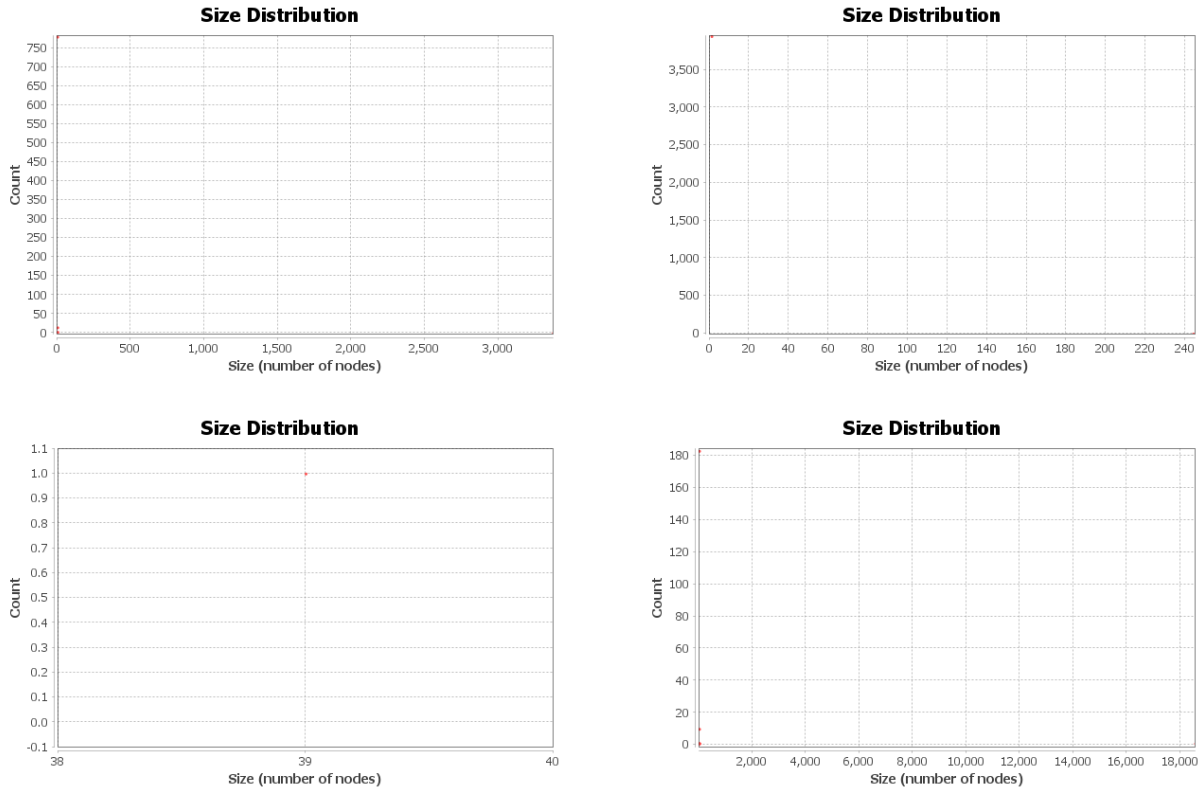
Табела 3.2.2.1 Просечне дистанце и дијаметри

Мрежа	SNet	SNetF	SNetT	UserNet
Просечна дистанце	2.107	1.814	1	4.190
Дијаметар	5	4	1	12

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

Табела 3.2.3.1 Повезаност мрежа

Мрежа	SNet	SNetF	SNetT	UserNet
Број слабо повезаних компоненти:	799	3948	1	197 (13761 јако повезаних)



Слика 3.2.1 Преглед дистрибуције величина SNet, SNetF, SNetT, UserNet

Све мреже поседују гигантске мреже, осим SNetT која је комплетан граф. Гигантска мрежа се може тестирати коришћењем адекватног филтера у алату *Gephi*.

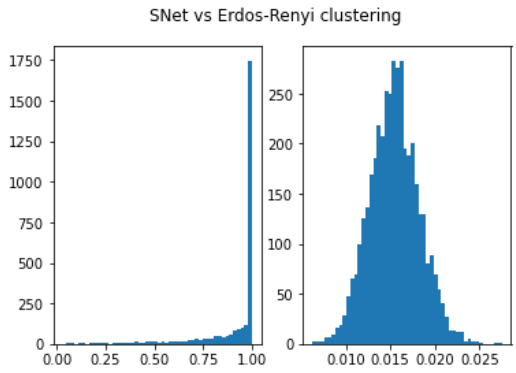
3.2.4. *Колики је просечни, а колики глобални коефицијент кластеризације мреже? Каква је расподела локалног коефицијента кластеризације њених чворова? Да ли је кластерисање изражено или не? Одговор дати упоређивањем са случајно генерисаном Ердош-Рењи мрежом истих димензија.*

Табела 3.2.4.1 Преглед коефицијената кластеризације

Мрежа	SNet	SNetF	SNetT	UserNet
Просечни са тежинама	0.000247	0.000523	0.060745	0.000379
Глобални са тежинама	0.001219	0.014698	0.139317	0.085552

Просечни без тежина	0.618428	0.046684	1.0	0.028293
Глобални без тежина	1.0	1.0	1.0	1.0
Просечни EP	0.015472	0.000409	1.0	0.000556
Глобални EP	0.027484	1.0	1.0	0.013889

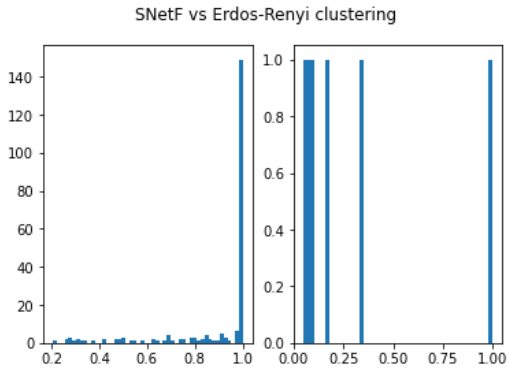
Интересантно је да за SNet и SNetF, Gephi и NetworkX не дају исте резултате.



Слика 3.2.2 Висуелизација SNet EP

Визуелизација SNet:

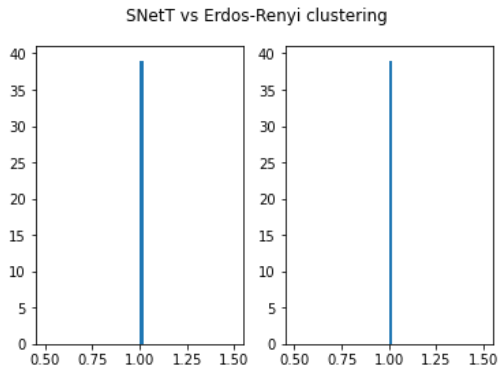
На слици се примећује да већина чворова има степен кластеризације вредности 1. Ова мрежа је добро кластерисана, и то у један велики кластер.



Слика 3.2.3 Висуелизација SNetF EP

Визуелизација SNetF:

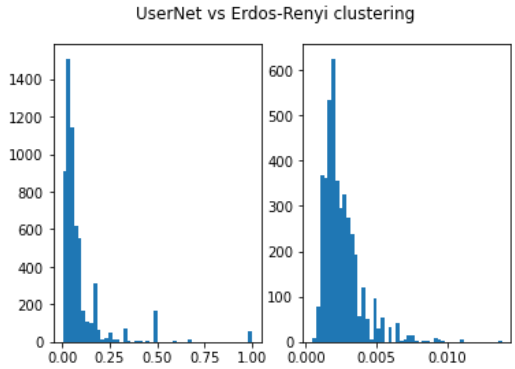
Слика јако подсећа на слику расподеле код SNet мреже. Кластеризација је додатно изражена, што је и очекивано с обзиром да је велики број грана графа уклоњен. И ова мрежа је добро кластерисана.



Слика 3.2.4 Висуелизација SNetT EP

Визуелизација SNetF:

Како се креирање случајне мреже врши на основу броја ивица и чворова, као и оригинална SNetT мрежа, Ердош-Рењи је такође комплетан граф. Отуд су вредности на слици идентичне. Ова мрежа је један једини кластер.



Слика 3.2.5 Висуелизација UserNet EP

Визиелизација UserNet:

На први поглед, две расподеле јако личе. Из тога се може закључити да не постоји нека природан начин повезивања чворова, већ је, слично Ердош-Рењи мрежи, повезивање насумично.
Такође, са слике се примећује да су вредности расподеле коефицијента кластеризације јако мали, па се долази до закључка да мрежа није изражено кластерисана.

3.2.5. На основу питања 8 и 10, закључити да ли мреже показују особине малог-света.

Особину малог-цвета мрежа показује уколико је мала вредност просечне удаљености и уколико је висок степен кластеризације.

Како се резултати из Gephi и из NetworkX не подударају, а за неке графове чак није могуће ни покренути због неповезаности потпуно повезаног графа, није било могуће срачунати одговоре за све мреже.

Sigma = (C/Cr) / (L/Lr) > 1 == small-world *(r = Ердош-Рењи мреже)*

Резултати који су успели да се изврше:

SNet: Јесте, по формули за вредност сигме.

UserNet: Јесте, по формули за вредност сигме.

Слободна процена:

SNet: Да, јер је просечни пут само 2 и има највиши степен кластеризације

SNetF: Не, иако је још мањи просечни пут, ова мрежа има јако низак просечни степен кластеризације због слабо повезаних компоненти

SNetT: Да, јер је свако са сваким повезан.

UserNet: Не, јер је виши просечни пут, а постоји шанса да и не постоји пут с обзиром да је граф усмерен.

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

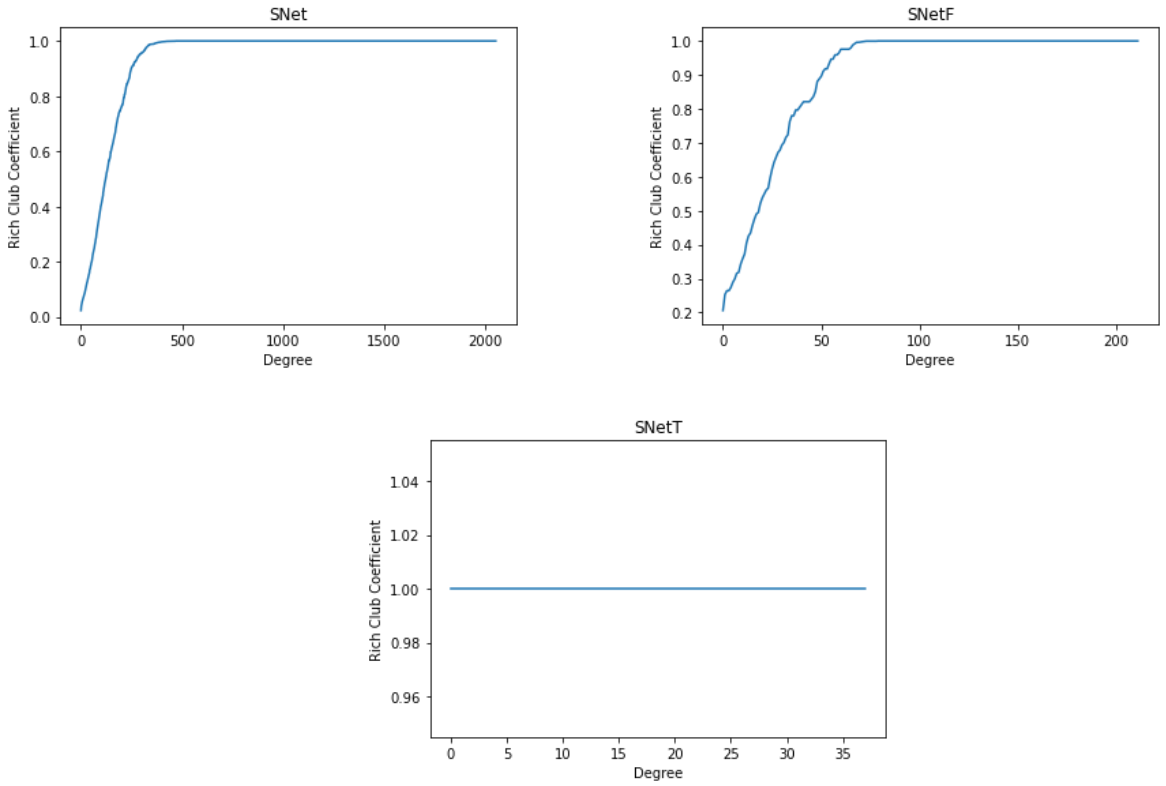
Табела 3.2.6.1 Асортативна анализа

Мрежа	SNet	SNetF	SNetT	UserNet
Коефицијент асортативности	-0.42823	-0.61035	Неодређено	In: -0.06096 Out: -0.00102

Асортативно мешање је присутно у мрежама SNet и SNetF јер су коефицијенти ближи -1.
Асортативно мешање није могуће за комплетан граф SNetT.
Асортатибно мешање није присутно за мрежу UserNet јер су коефицијенти ближи 0.

3.2.7. Да ли мрежа испољава феномен клуба богатих (rich-club-phenomenon)?.

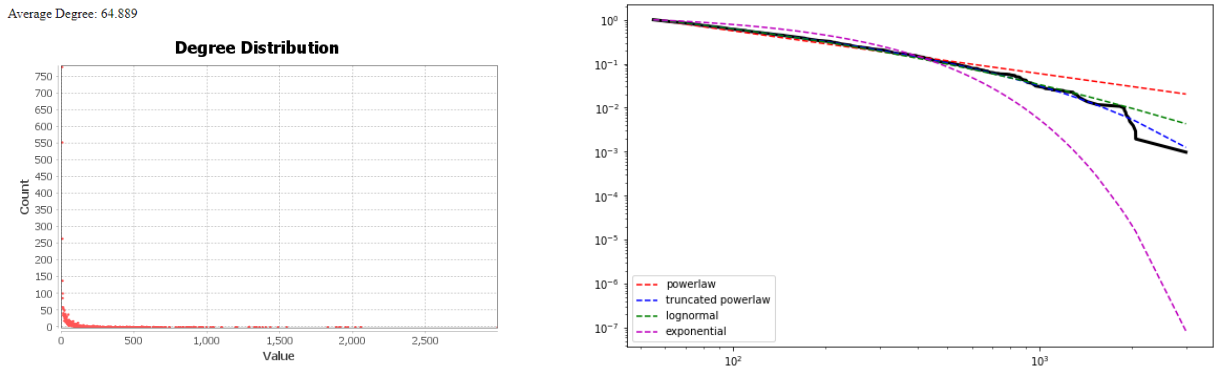
Слично пословици „богати постају богатији, а сиромашни сиромашнији“, мрежа мора да има мали број хабова, а велики број слабо повезаних компоненти. Иако је можда већ доказано, исцртавањем решења долазимо до закључка.



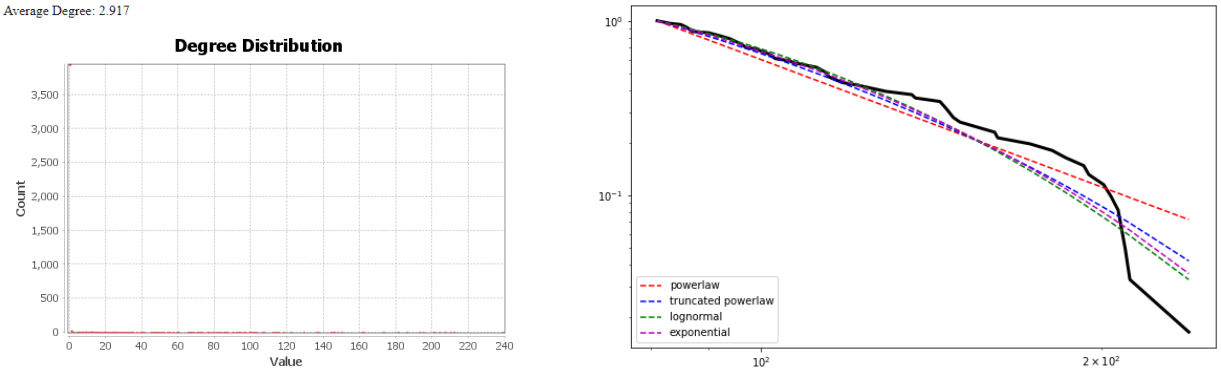
Слика 3.2.6 Феномен Клуба-богатих *Rich-club*

Као што смо и могли да претпоставимо, прве две мреже испољавају феномен, док трећа то не чини. Како је трећа мрежа комплетна, коефицијент се понаша као константа. UserNet као усмерен граф не може да испољава овај феномен. Поред усмерених, ни мултиплекс графови и графови са чворовима који су повезани сами са собом (*self-looping)* не испољавају овај феномен.

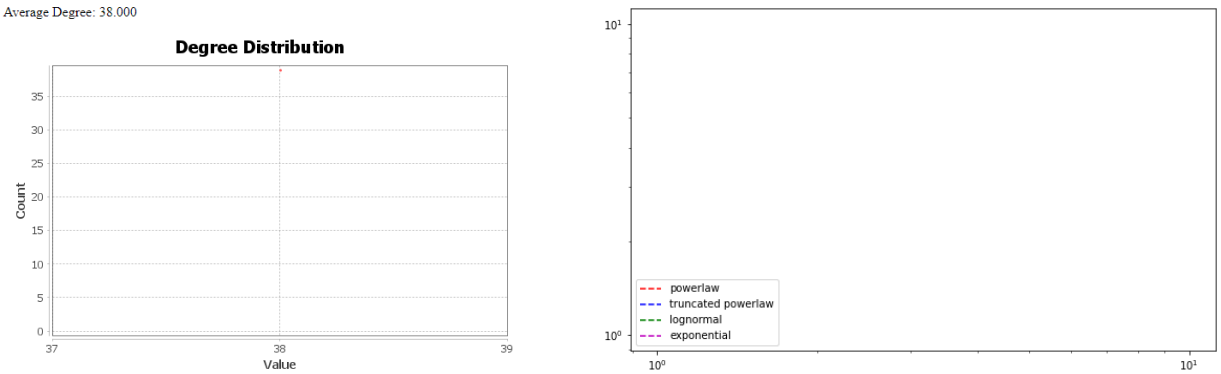
3.2.8. Каква је дистрибуција по степену и да ли прати power-law дистрибуцију?



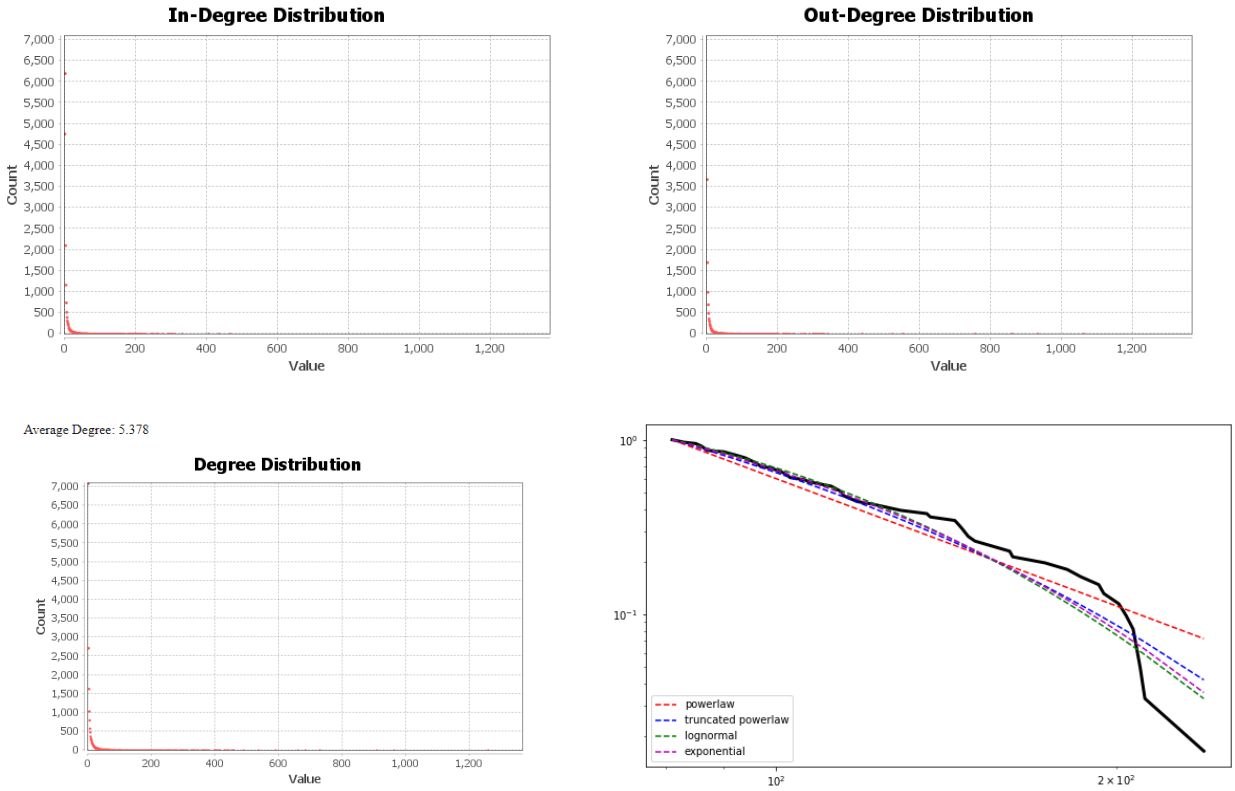
Слика 3.2.7 Дистрибуција степена SNet мреже



Слика 3.2.8 Дистрибуција степена SNetF мреже



Слика 3.2.9 Дистрибуција степена SNetT мреже

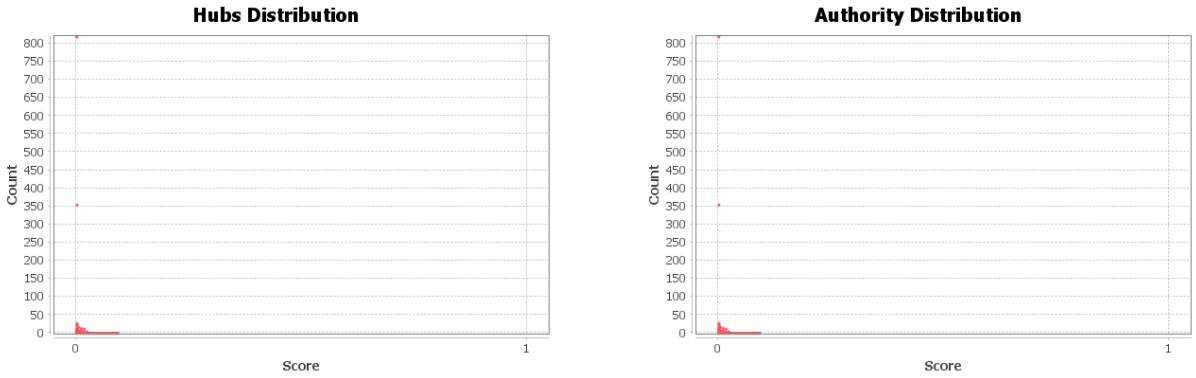


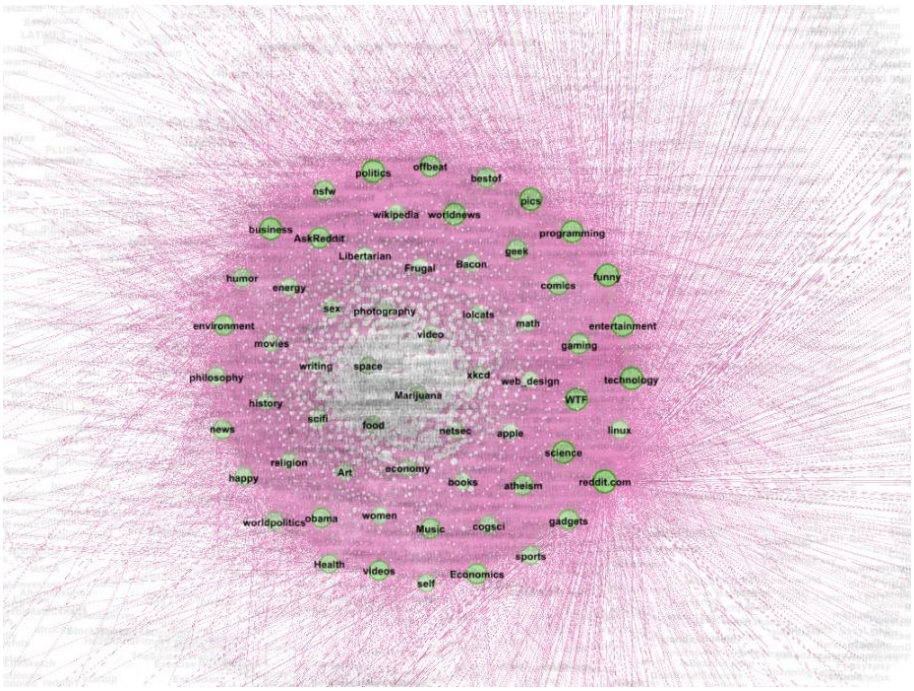
Слика 3.2.10 Дистрибуција степена UserNet мреже

Тумачење графова:

Мрежа	Разлог	Одлука
SNet	Truncated power-law	Да
SNetF	Fluctuates between power-law and lognormal	Не
SNetT	NaN	Не
UserNet	Fluctuates too much	Не

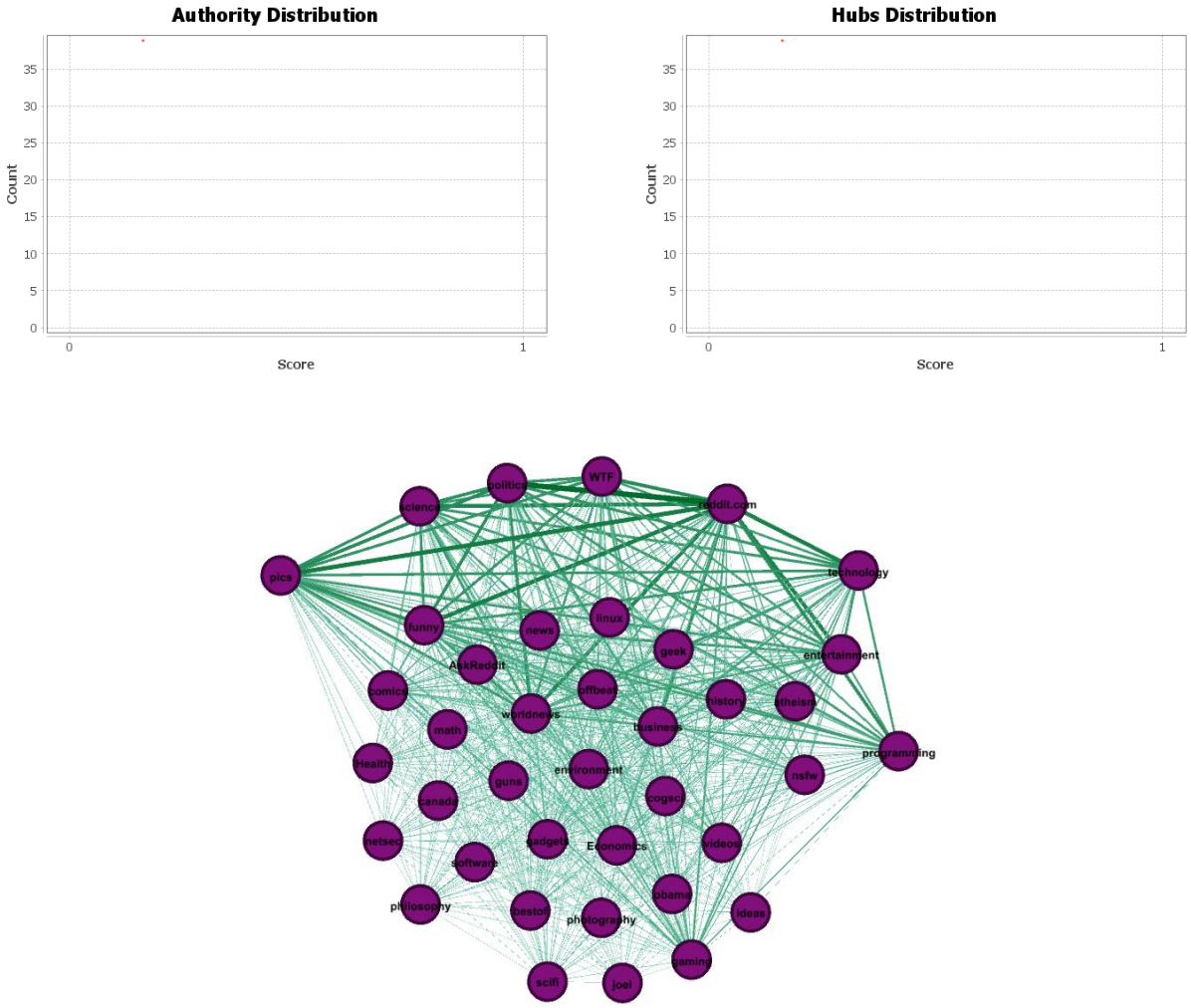
3.2.9. Одредити најважније хабове и ауторитете у мрежи. Како су они распоређени и уграђени у мрежу? Да ли су на периферији или у језгру?



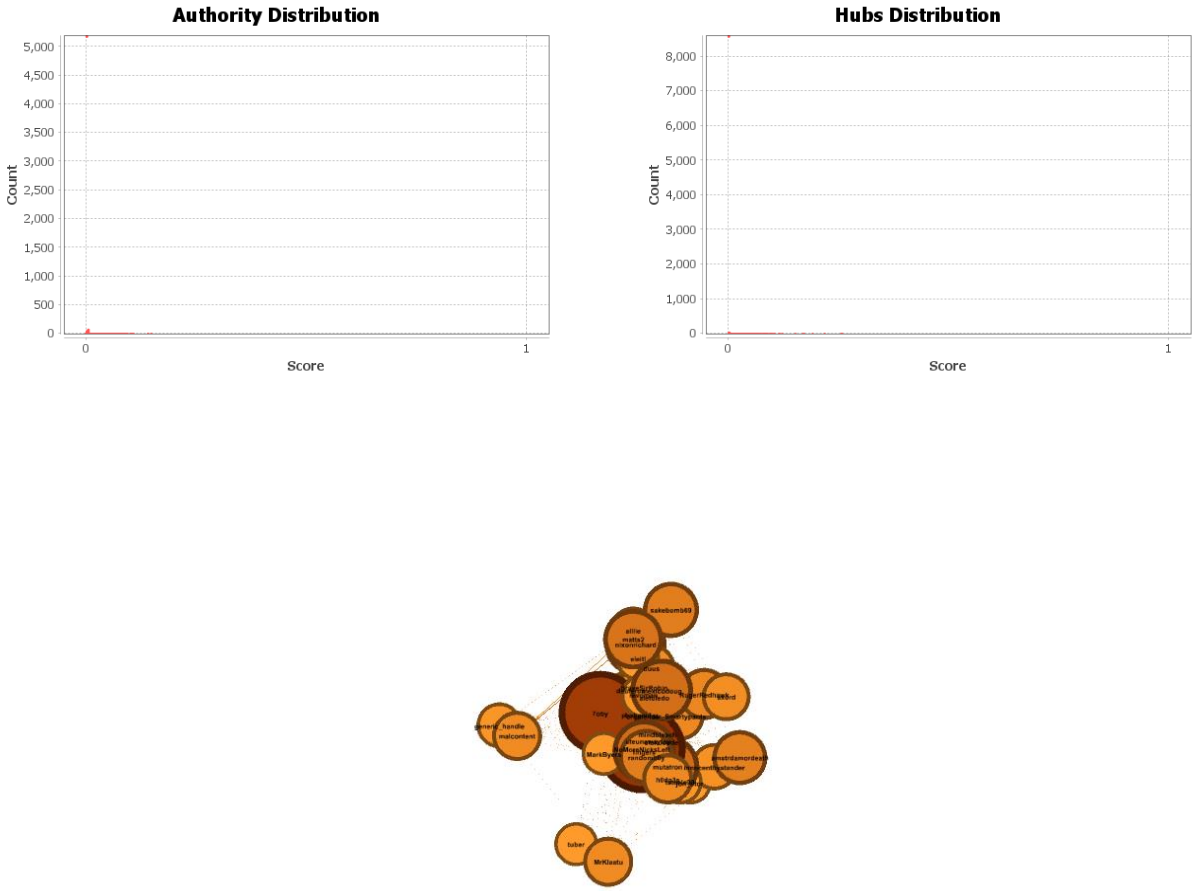


Слика 3.2.11 SNet хабови и ауторитети

Слика 3.2.12 SNetF хабови и ауторитети



Слика 3.2.13 SNetT хабови и ауторитети



Слика 3.2.14 UserNet хабови и ауторитети

3.3. Анализа мера централности

3.3.1. Спровести анализе мера централности по степену, блискости и релационој централности. Дати преглед најважнијих актера по свакој од њих.

Табела 3.3.1.1 Мере централности SNet

subreddit	dc	subreddit	cc	subreddit	bc
reddit.com	0.71432	reddit.com	0.72284	reddit.com	0.24826
politics	0.48998	technology	0.57667	technology	0.03186
technology	0.48950	politics	0.5763	programming	0.02954
pics	0.48091	pics	0.57157	politics	0.02828
funny	0.46850	funny	0.56514	business	0.02677
science	0.46468	science	0.56326	pics	0.02523
entertainment	0.45632	entertainment	0.55942	entertainment	0.02474
worldnews	0.45465	worldnews	0.55873	worldnews	0.02168
programming	0.45012	programming	0.55576	funny	0.02113
WTF	0.44940	WTF	0.55576	science	0.02055

Табела 3.3.1.2 Мере централности SNetF

subreddit	dc	subreddit	cc	subreddit	bc
reddit.com	0.05704	reddit.com	0.05706	reddit.com	0.00086
pics	0.0506	pics	0.05106	politics	0.00018
politics	0.05012	politics	0.05069	pics	0.00018
science	0.0494	science	0.05015	science	0.00015
technology	0.0494	technology	0.05015	technology	0.00015
funny	0.04869	funny	0.04962	programming	0.00014
worldnews	0.04797	worldnews	0.0491	funny	0.00014
WTF	0.04654	WTF	0.0481	worldnews	0.00013
entertainment	0.04606	entertainment	0.04777	WTF	0.00011
programming	0.04439	programming	0.04667	entertainment	0.0001

Табела 3.3.1.3 Мере централности SNetT

subreddit	dc	subreddit	cc	subreddit	bc
reddit.com	1.0	reddit.com	1.0	reddit.com	0.0
funny	1.0	funny	1.0	funny	0.0
atheism	1.0	atheism	1.0	atheism	0.0
bestof	1.0	bestof	1.0	bestof	0.0
technology	1.0	technology	1.0	technology	0.0
WTF	1.0	WTF	1.0	WTF	0.0
canada	1.0	canada	1.0	canada	0.0
geek	1.0	geek	1.0	geek	0.0
photography	1.0	photography	1.0	photography	0.0
history	1.0	history	1.0	history	0.0

Табела 3.3.1.4 Мере централности UserNet

subreddit	dc	subreddit	cc	subreddit	bc
7oby	0.07199	7oby	0.19143	7oby	0.03876
alllie	0.06625	NoMoreNicksLeft	0.18919	alllie	0.02146
rmuser	0.0507	mutatron	0.18328	qgyh2	0.01541
qgyh2	0.04785	fingers	0.18261	rmuser	0.01456
deuteros	0.03842	Poromenos	0.18236	NoMoreNicksLeft	0.01248
tsteele93	0.036	amstrdamordeath	0.18161	mutatron	0.01097
NoMoreNicksLeft	0.03478	MrKlaatu	0.18122	tsteele93	0.01073
bobcat	0.03104	sakebomb69	0.18045	deuteros	0.00984
mutatron	0.02825	uteunawaytay	0.17882	bobcat	0.00883
jordanlund	0.02577	nixonrichard	0.17757	glengyron	0.00851

subreddit	in_dc	subreddit	out_dc
NoMoreNicksLeft	0.02451	alllie	0.05587
7oby	0.02287	7oby	0.04912
mutatron	0.02129	rmuser	0.04527
Poromenos	0.01739	qgyh2	0.03979
fingers	0.01634	deuteros	0.02909
matts2	0.01597	tsteele93	0.02751
amstrdamordeath	0.01581	bobcat	0.02303
MrKlaatu	0.01581	grauenwolf	0.01792
nixonrichard	0.01549	Aerik	0.01734
malcontent	0.01528	AMerrickanGirl	0.01713

3.3.2. Ко су најважнији актери по централности по сопственом вектору? Шта нам то говори о њима?

Ова мера централности је варијанта централности по степену. Говори нам о утицају, односно моћи, сваког чвора. Моћнији чвор је чвор који је окружен чворовима са нижим степенима централности. Утицајнији чвор је чвор који има висок степен централности, баш као и његови суседи. Табеле у наставку представљају утицај.

Табела 3.3.2.1 Својствени вектор SNet

subreddit	evc
reddit.com	0.38428
politics	0.3157
pics	0.29206
funny	0.28054
science	0.27384
technology	0.26536
worldnews	0.25857
WTF	0.25355
entertainment	0.25304
programming	0.22023

Табела 3.3.2.2 Својствени вектор SNetF

subreddit	evc
reddit.com	0.38434
politics	0.31576
pics	0.29211
funny	0.28058
science	0.27388
technology	0.2654
worldnews	0.25861
WTF	0.25358
entertainment	0.25308
programming	0.22025

Табела 3.3.2.3 Својствени вектор SNetT

subreddit	evc
reddit.com	0.38898
politics	0.31938
pics	0.29518
funny	0.28349
science	0.27677
technology	0.26828
worldnews	0.26141
WTF	0.25601
entertainment	0.25554
programming	0.22282

Табела 3.3.2.4 Својствени вектор UserNet

subreddit	evc
malcontent	0.61088
glengyron	0.48943
matts2	0.19386
eaturbrainz	0.15458
NoMoreNicksLeft	0.15251
7oby	0.136
43P04T34	0.10537
mexicodoug	0.09835
otakucode	0.09579
sakebomb69	0.09521

Табела 3.3.2.5 Преглед централности по својственом вектор по мрежама

3.3.3. Рангирати чворове по Кацовој централности са варијацијом параметара. При рачунању, експериментисати са додељивањем другачије вредности параметра бета за сабредит који се у приложеним CSV фајловима идентификује вредношћу колоне subreddit reddit.com. Дати преглед најважнијих актера у случају да је бета исто за све сабредите и у случају да је бета наведеног сабредита значајно веће.

Пре свега, у овом задатку нема смисла испитивати мрежу UserNet, јер она не поседује сабредите као чворове.Даље, Кацова централност рачуна централност посматраног чвора у односу на централност његових суседа. Ово је генерализација централности по својственом вектору. Параметар бета служи за пондерисање, док параметар алфа контролише рад самог

алгоритма. Алфа је потребно да буде мање од реципрочне вредности максималног ламбда. За детаљнији опис, погледати сам рад. Бета је тестирано за 1, 2, 4, 64 и 4096. Алфа је одабрано да буде $6e^{-8}$ јер је ова вредност мања од сваког максималног алфа. Повећавањем параметра бета долази до повећавања јаза између вредности централности и сатуризације најцентралнијег. Приказане су вредности за бета 1, 2 и 4096, респективно. Остале вредности су доступне у свесци.

Табела 3.3.3.1 Кац централност SNet

Бета:	1		2		4096
subreddit	katz	subreddit	katz	subreddit	katz
reddit.com	0.01577	reddit.com	0.0312	reddit.com	0.99986
politics	0.0157	politics	0.01572	politics	0.00187
pics	0.01569	pics	0.0157	pics	0.00155
funny	0.01568	funny	0.01569	funny	0.00147
science	0.01568	science	0.01569	technology	0.00147
technology	0.01567	technology	0.01568	entertainment	0.00143
worldnews	0.01566	worldnews	0.01567	science	0.00143
WTF	0.01566	WTF	0.01567	worldnews	0.00135
entertainment	0.01566	entertainment	0.01567	business	0.00131
programming	0.01563	programming	0.01564	WTF	0.00129

Табела 3.3.3.2 Кац централност SNetF

Бета:	1		2		4096
subreddit	katz	subreddit	katz	subreddit	katz
reddit.com	0.01576	reddit.com	0.0312	reddit.com	0.99986
politics	0.0157	politics	0.01572	politics	0.00187
pics	0.01568	pics	0.0157	pics	0.00155
funny	0.01567	funny	0.01569	funny	0.00147
science	0.01567	science	0.01568	technology	0.00147
technology	0.01566	technology	0.01567	entertainment	0.00143
worldnews	0.01565	worldnews	0.01567	science	0.00143
WTF	0.01565	WTF	0.01566	worldnews	0.00135
entertainment	0.01565	entertainment	0.01566	business	0.00131
programming	0.01562	programming	0.01563	WTF	0.00129

Табела 3.3.3.3 Кац централност SNetT

Бета:	1		2		4096
subreddit	katz	subreddit	katz	subreddit	katz
reddit.com	0.16202	reddit.com	0.3095	reddit.com	0.99999
politics	0.16146	politics	0.15577	politics	0.00187
pics	0.16132	pics	0.15558	pics	0.00154
funny	0.16123	funny	0.15549	funny	0.00147
science	0.1612	science	0.15545	technology	0.00147
technology	0.16113	technology	0.15538	entertainment	0.00143
worldnews	0.16107	worldnews	0.15531	science	0.00143
WTF	0.16105	WTF	0.15528	worldnews	0.00135
entertainment	0.16103	entertainment	0.15528	business	0.00131
programming	0.1608	programming	0.15504	WTF	0.00129

Табела 3.3.3.4 Преглед Кац централности по мрежама

3.3.4. *На основу претходна три питања, предложити и конструисати композитну меру централности за проналажење најважнијих актера. Обратити пажњу на тип усмерености мреже и сходно томе прилагодити колико различите мрежне метрике утичу на хеуристику.*

Провером корелације и доказом корелисаности, закључено је:

- Не постоји корелација у SNetT, из разлога што је то комплетан граф.
- Постоји позитивна корелација мера централности у мрежама SNet I SNetF.
- За UserNet је потребно смислити другачију композитну меру из разлога што је out-degree негативно корелисан.

Предлози:

Composite Centrality = DC x CC x BC x EC

Composite Centrality Directed = (InDC / OutDC) x CC x BC

Табела 3.3.4.1 Композитна централност SNet

subreddit	composite_rank
reddit.com	1.0
politics	48.0
technology	72.0
pics	288.0
funny	900.0
science	1800.0
programming	2565.0
entertainment	3087.0
worldnews	3584.0
business	6655.0

Табела 3.3.4.2 Композитна централност SNetF

subreddit	composite_rank
reddit.com	1.0
politics	36.0
pics	36.0
technology	486.0
science	506.25
funny	1008.0
worldnews	2744.0
WTF	4608.0
programming	6000.0
entertainment	7290.0

Табела 3.3.4.3 Композитна централност SNetT

subreddit	composite_rank
reddit.com	8000.0
politics	16000.0
pics	24000.0
funny	32000.0
science	40000.0
technology	48000.0
worldnews	56000.0
WTF	64000.0
entertainment	72000.0
programming	80000.0

Табела 3.3.4.4Композитна централност UserNet

subreddit	composite_rank
NoMoreNicksLeft	1.25
7oby	6.0
mutatron	9.10843
malcontent	17.3444
fingers	40.82111
amstrdamordeath	41.29412
matts2	43.67418
sakebomb69	72.7504
Poromenos	80.82474
MrKlaatu	184.48427

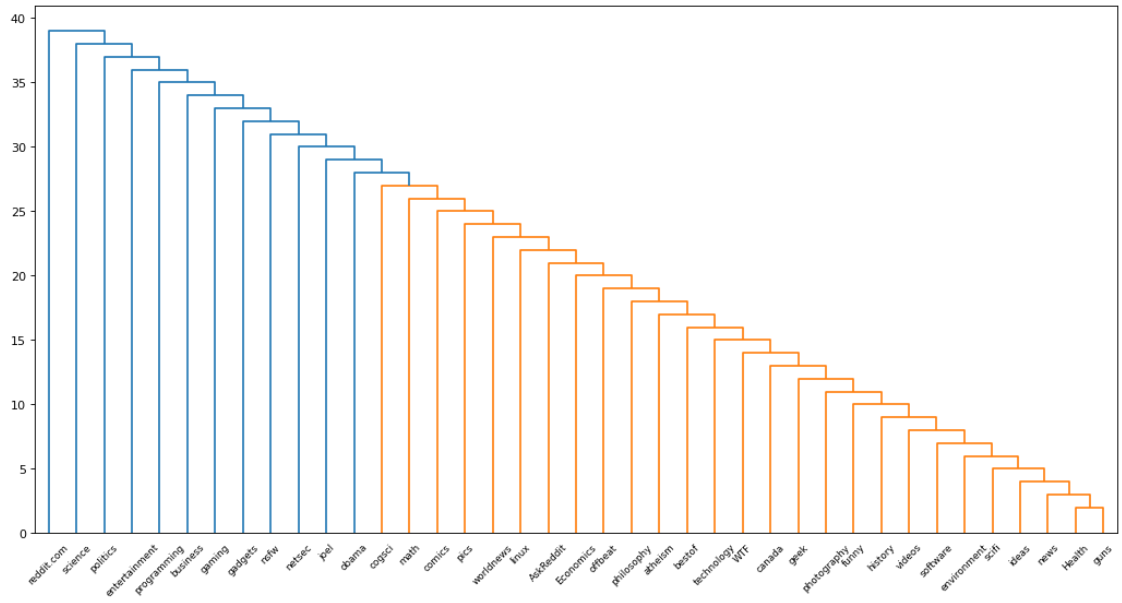
Вредност која се добија за усмерени граф је „лева“ централност по свој. век и представља вредност добијену у односу улазећи степен централности. Да би се добио „десни“, потребно је окренути граф, односно позвати методу UserNet.*reverse()* па поновити.

3.4. Детекција комуна

3.4.1. Ако величина мреже дозвољава, спкетралном анализом или анализом дендрограма проценити потенцијалне кандидате за број комуна у мрежи.

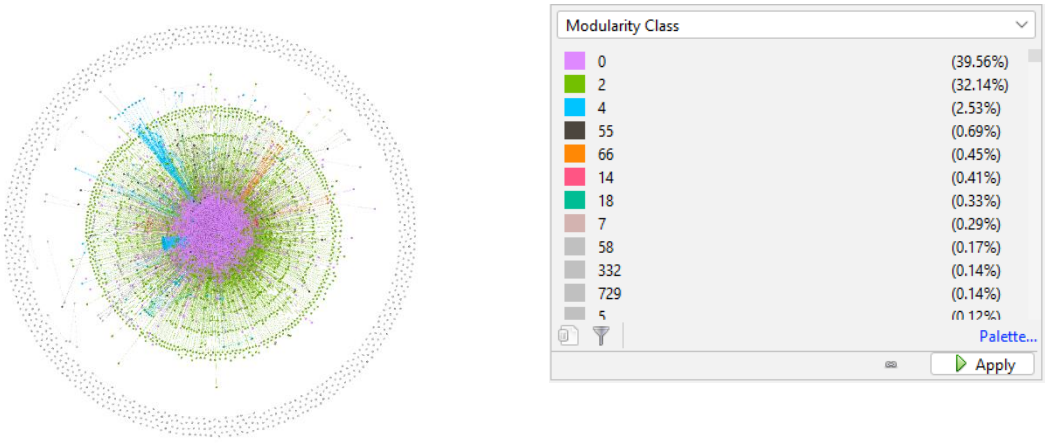
UserNet је био исувише велики како би се показао било на који начин.

Како су мреже исувише велике за сликање дендрограма, овим методом је осликан а једино мрежа **SNetT**. Очекивано би било да постоји само једна комуна, што се и може закључити аналитом дендрограма .



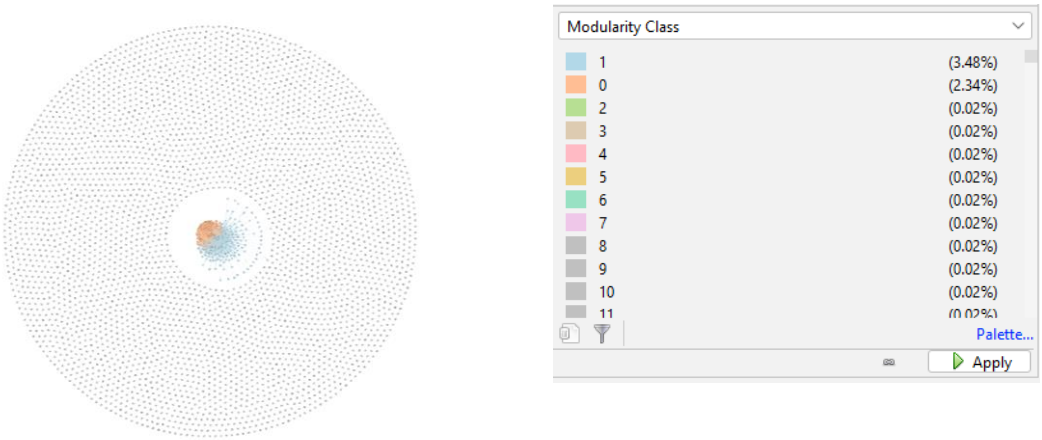
Слика 3.4.1 SNet дендрограм

За **SNet** је јасно се издвајају 3 комуне. Трећа, иако доста мања, ипак чини комуну у односу на број чворова и број јако малих комуна.



Слика 3.4.2 Спектрограм SNet

За **SNetF** мрежу се јасно примећују 2 комуне и велики број неповезаних компоненти.



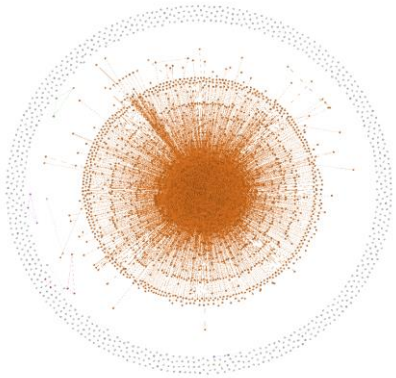
Слика 3.4.3 Спектрограм SNetF

3.4.2. 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).

По објашњењу које стоји у алату, вредност параметра резолуције мања од 1 ће резултирати мањим бројем комуна, док ће вредност преко 1 резултирати већим.

SNet мрежа има три комуне као што је претходно приказано. Уколико повећамо вредност резолуције преко прага 1, доћи ће до спајања комуна у једну велику и велики број неповезаних. Уколико се смањи, враћа се на 3, од којих су 2 велике са по 30-40%, а трећа до 2%. Покушаване су вредности 0.8, 0.5, 0.25, 0.05 али нису примећене значајне промене. Вероватна последица великог броја слабо повезаних компоненти.

Табела 3.4.2.1 Преглед кластерисања Лувенским методом SNet

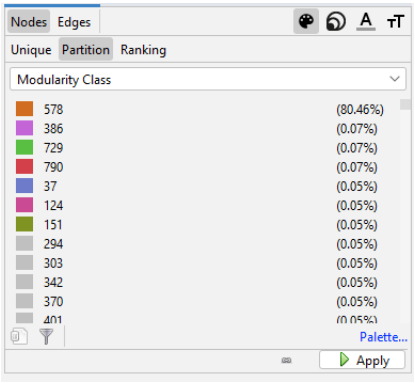


Резолуција: 5.0

Модуларност: 0.000

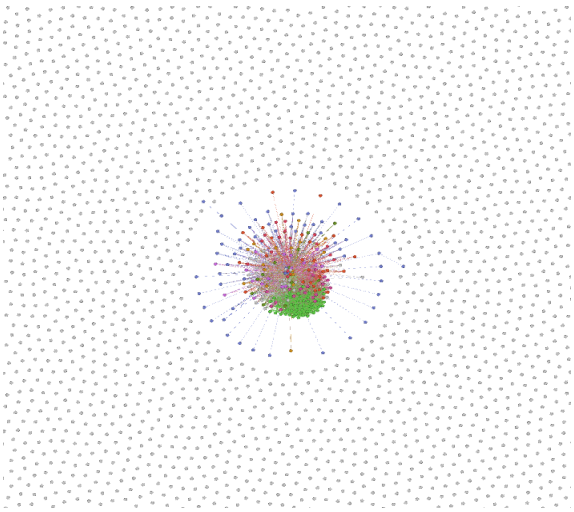
Модуларност са резолуцијом: 4.000

Број комуна: 799



SNetF мрежу карактерише још израженији број неповезаних компоненти с обзиром да су гране испод прага брисане. Понашање је исто као и у претходном примеру. Када је вредност резолуције већа од 1, долази до сједињавања унутрашње две комуне (претходно добијене). Уколико је мања од 1, долази до поделе на већи број комуна, које су занемарљивих величина.

Табела 3.4.2.2 Преглед кластерисања Лувенским методом SNetF



Резолуција: 0.5

Модуларност: 0.004

Модуларност са резолуцијом: -0.046

Број комуна: 3960

SNetT мрежа је потпуно повезан граф. Број комуна варира између 1 и 2. Са резолуцијом 1 и већом, број комуна је два, док са било чим мањим, број комуна је један. Овај граф би требало да има само једну комуну, тако показује да вредност резолуције не мора да буде прецизна и тачна.

UserNet мрежа је огромна. Иако је могуће поделити на више комуна, модуларност за резолуцију вредности један враћа 0.320 модуларност и 0.320 модуларност са резолуцијом, као и 221 комуну, због саме величине није могуће у неком реалном времену приказати изглед овог графа.

3.4.3. Које заједнице се могу уочити приликом анализе мреже? Да ли постоји неко објашњење за детектоване комуне.

За ово питање има смисла посматрати само SNetF мрежу. Остале мреже су или превелике или немају комуне. У поменутој мрежи није уочено правило поделе. Ако посматрамо чворове са највећим степенима у обе комуне, видећемо „WTF”, „entertainment”, „ funny”, али и „politics”, „programming”, „business” у првој комуни, док се у другој издвајају „Obama”, „economics”, „askreddit”, „linux” итд. Подела не делује да је направљена према теми, али можда постоји нека друга, скривенија, информација. Те поделе могу да буду локације интераговања па тиме и другачија количина грана (асортативност и дељење међу пријатељима у САД и Европе), или подела на то да ли неке сабредите људи више читају, дајући већи значај објавама него коментарима, односно активни и пасивни облик интеракција. Са датог графа није могуће направити закључак.

3.4.4. Ко су брокери (мостови) и мрежи? Да ли припадају језгру или периферији или су мешовито распоређени?

Библиотека *networkx* поседује методе за проверу мостова (грана) у графу. Уколико је грана мост, вероватно су и чворови који су повезани њоме брокери. Уз помоћ ове методе, доказано је да у мрежи SNetT нема мостова, док за UserNet нису дефинисани. Брокери се карактеришу као чворови са високом релационом централношћу, али и ниским мрежним ограничењима, па се са тим у виду вредност брокера може дефинисати као:

Brokerage = BC/NC

Како су се неке операције предуго извршавале, а знамо да су већина тежина грана једнаке 1, за вредност NC је у великим графовима наметнута вредност 1.

SNet

Табела 3.4.4.1 SNet брокери

subreddit	brokerage	BC	NC
reddit.com	0.21779	0.21779	1.0
technology	0.02635	0.02635	1.0
programming	0.0258	0.0258	1.0
business	0.02528	0.02528	1.0
entertainment	0.02136	0.02136	1.0
politics	0.02051	0.02051	1.0
pics	0.0201	0.0201	1.0
worldnews	0.01729	0.01729	1.0
funny	0.01586	0.01586	1.0
science	0.01484	0.01484	1.0

Табела 3.4.4.2 SNetF брокери

subreddit	brokerage	BC	NC
reddit.com	0.00509	0.00063	0.12449
linux	0.00179	0.00019	0.10482
programming	0.00141	0.00018	0.12763
offbeat	0.001	0.00011	0.11174
cogsci	0.00095	9e-05	0.09096
worldnews	0.00094	0.00012	0.1268
scifi	0.00093	9e-05	0.09233
entertainment	0.00091	0.00011	0.12689
business	0.00085	0.00011	0.12912
nature	0.0008	8e-05	0.09991

Табела 3.4.4.3 SNetT брокери

subreddit	brokerage	BC	NC
ideas	4.63276	0.70199	0.15153
joel	3.86897	0.73044	0.18879
funny	0.0	0.0	0.18962
atheism	0.0	0.0	0.18024
bestof	0.0	0.0	0.16313
technology	0.0	0.0	0.18989
WTF	0.0	0.0	0.18877
canada	0.0	0.0	0.16721
geek	0.0	0.0	0.16901
photography	0.0	0.0	0.16069

Табела 3.4.4.4 UserNet брокери

subreddit	brokerage	BC	NC
7oby	0.01991	0.01991	1.0
alllie	0.00999	0.00999	1.0
jordanlund	0.00974	0.00974	1.0
deuteros	0.00943	0.00943	1.0
Prysorra	0.00927	0.00927	1.0
mutatron	0.00871	0.00871	1.0
tsteele93	0.00857	0.00857	1.0
linkedlist	0.00759	0.00759	1.0
bobcat	0.00732	0.00732	1.0
moogle516	0.00695	0.00695	1.0

3.5. Поређење SNet и SNetT мрежа

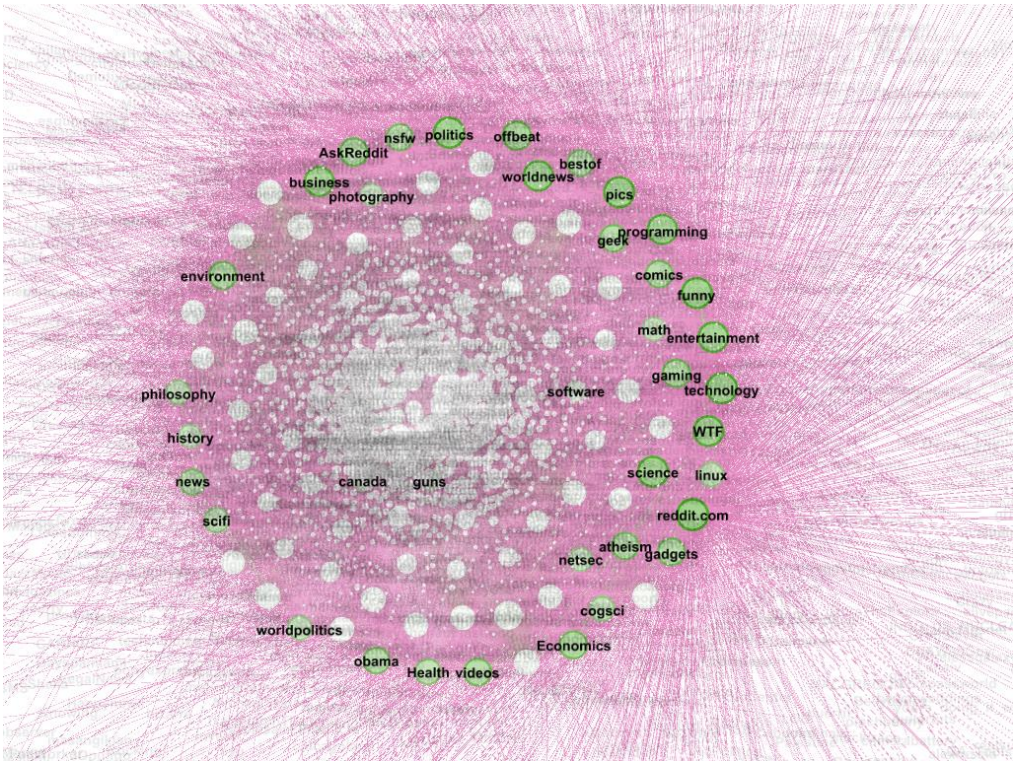
3.5.1. Упоредити карактеристике две мреже. Коментарисати потенцијалне разлике и проценити да ли су сабредити из SNetT активнији и боље повезани од остатка мреже.

Уколико се активност посматра кроз степен централности, најбољих 39 SNet-а се разликује за само 8 елемената у односу на SNet. Те разлике заузимају последња места, тако да се може сматрати да чворови SNetT-а представљају најактивније језгро, барем првих 20. Првих 30 се разликују само за субредите „music“ и „sports“. Ако се сетимо SNetF мреже, у њој је доказано да је скоро 96% грана између привих 41.

Мрежа SNetT се може сматрати кликом мреже SNet. Што се осталих карактеристика тиче, тешко је поредити огромну мрежу са комплетним графом.

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

Као што је у ранијим поглављима објашњено, чворови са вишим степеном централности се налазе у језгру. Како се првих 20 налази заиста у језгру и чине исту комуну, сигурно су и других 19 блиско повезани са првим, с обзиром да је SNetT само узорак SNet-а, и да је SNetT комплетан граф. Потребно је извршити филтрацију по regex-у у алату Gephi.



Слика 3.5.1 Приказ локације SNetT чворова у SNet мрежи

Списак слика

Списак табела