Универзитет у Београду

Електротехнички факултет



Анализа социјалних мрежа

Пројектни задатак

|  |  |
| --- | --- |
| Студенти: | Школска година: |
| Ана Стакић 2020/3103 Страхиња Стефановић 2020/3104 | 2021/2022. |

Београд, јануар 2022.

Садржај

[Садржај 2](#_Toc95339222)

[1. Чишћење података 3](#_Toc95339223)

[2. Моделовање мрежа 5](#_Toc95339224)

[2.1. Subreddit network – SNet 5](#_Toc95339225)

[2.2. Filtered subreddit network – SNetF 6](#_Toc95339226)

[2.3. Targeted subreddit network – SNetT 8](#_Toc95339227)

[2.4. User network – UserNet 9](#_Toc95339228)

[3. Статистичка обрада података 10](#_Toc95339229)

[4. Основна анализа моделованих мрежа 13](#_Toc95339230)

[5. Анализа мера централности 27](#_Toc95339231)

[6. Детекција комуна 35](#_Toc95339232)

[7. Поређење SNet и SNetT мрежа 36](#_Toc95339233)

[Списак слика 37](#_Toc95339234)

[Списак табела 38](#_Toc95339235)

# 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***, ***permalink***, ***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) нужан 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.

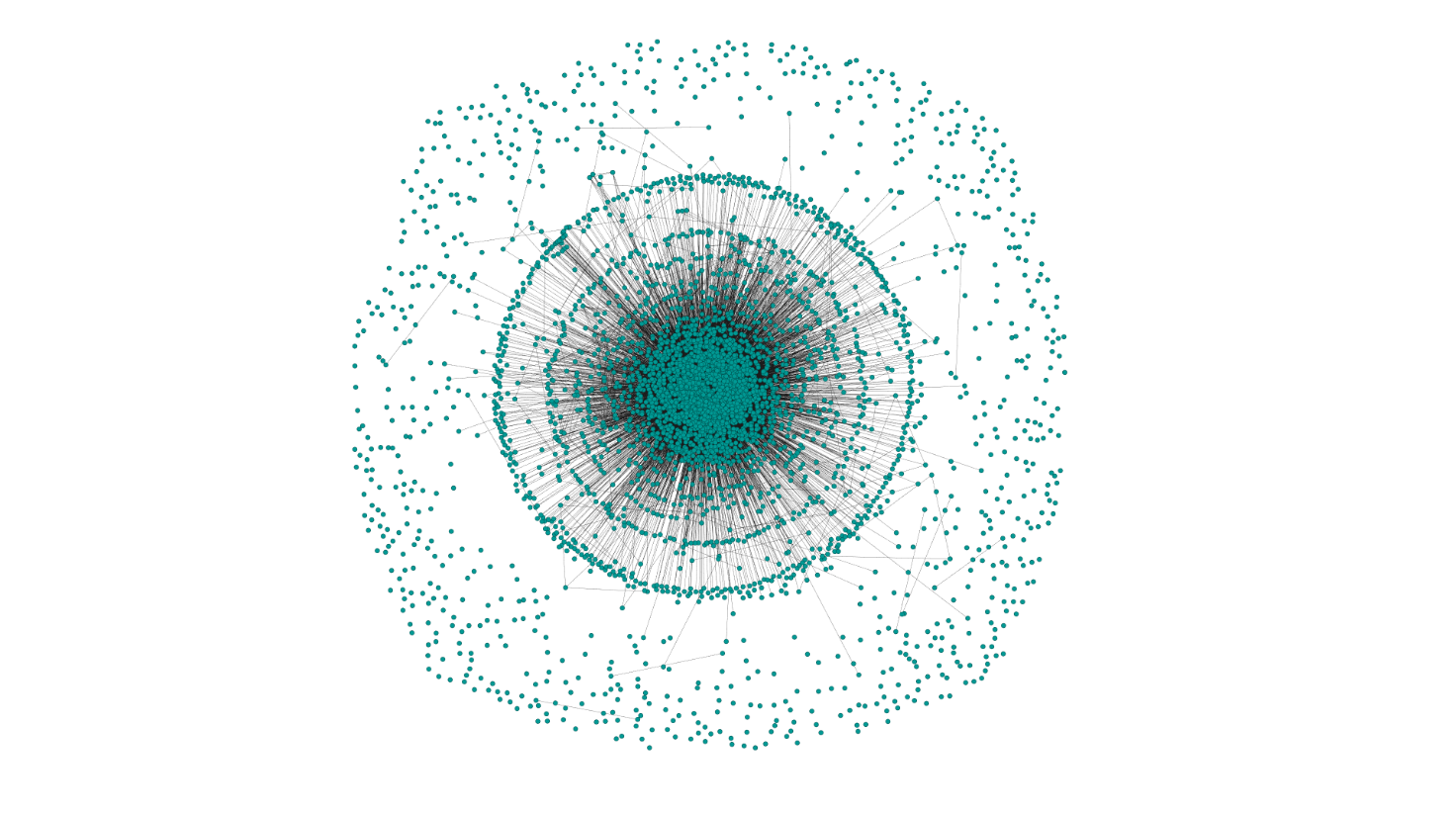
Табела 1.1 Утицај чишћења на величину скупа података

|  |  |  |
| --- | --- | --- |
|  | Примарни скуп података | Секундарни скуп података |
| Објаве | 2.519.853 × 16 | 2.044.810 × 12 |
| Коментари | 7.242.871 × 12 | 3.783.039 × 10 |

# Моделовање мрежа

## 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

## 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.

A picture containing chart

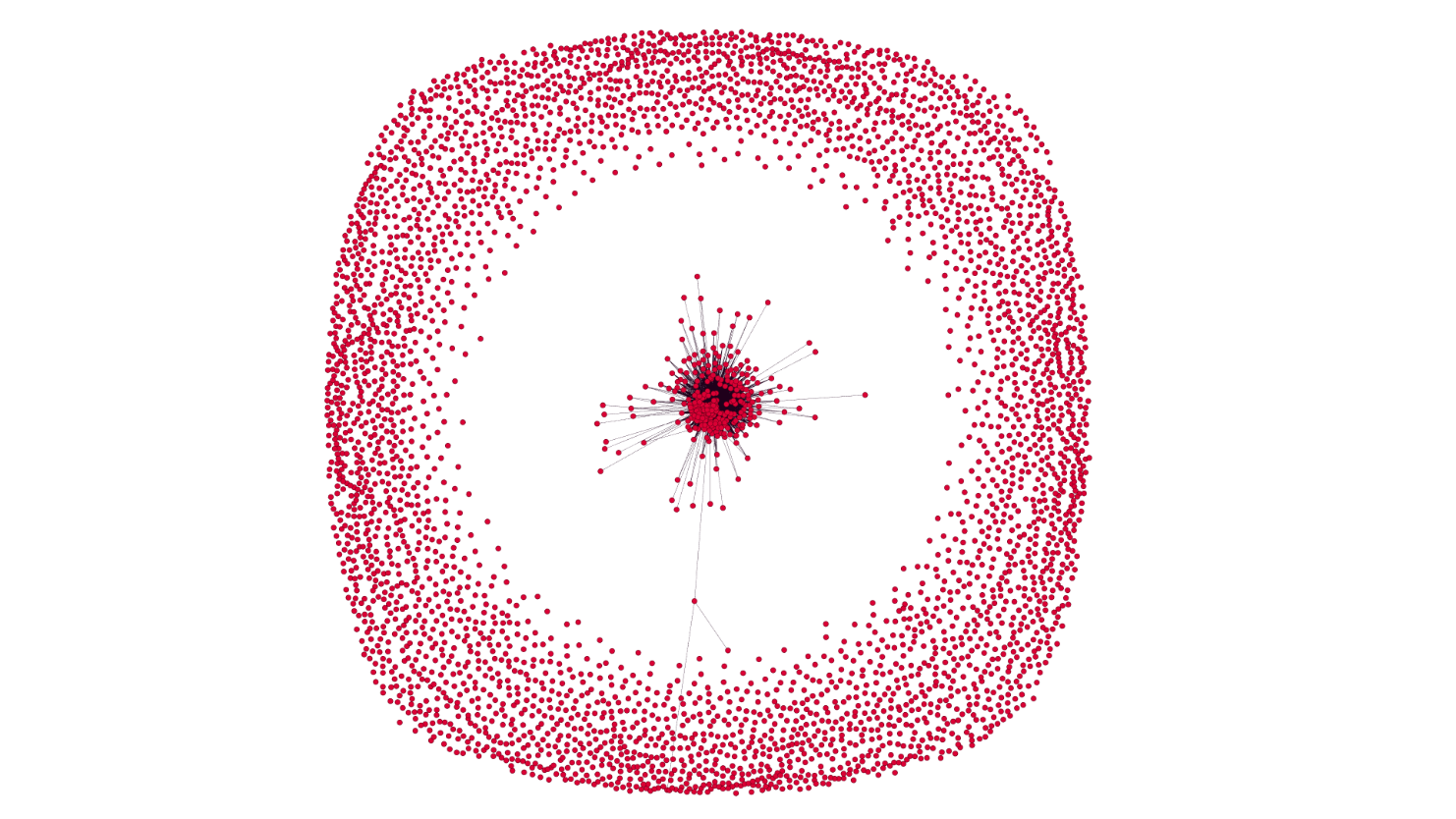
Description automatically generated

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

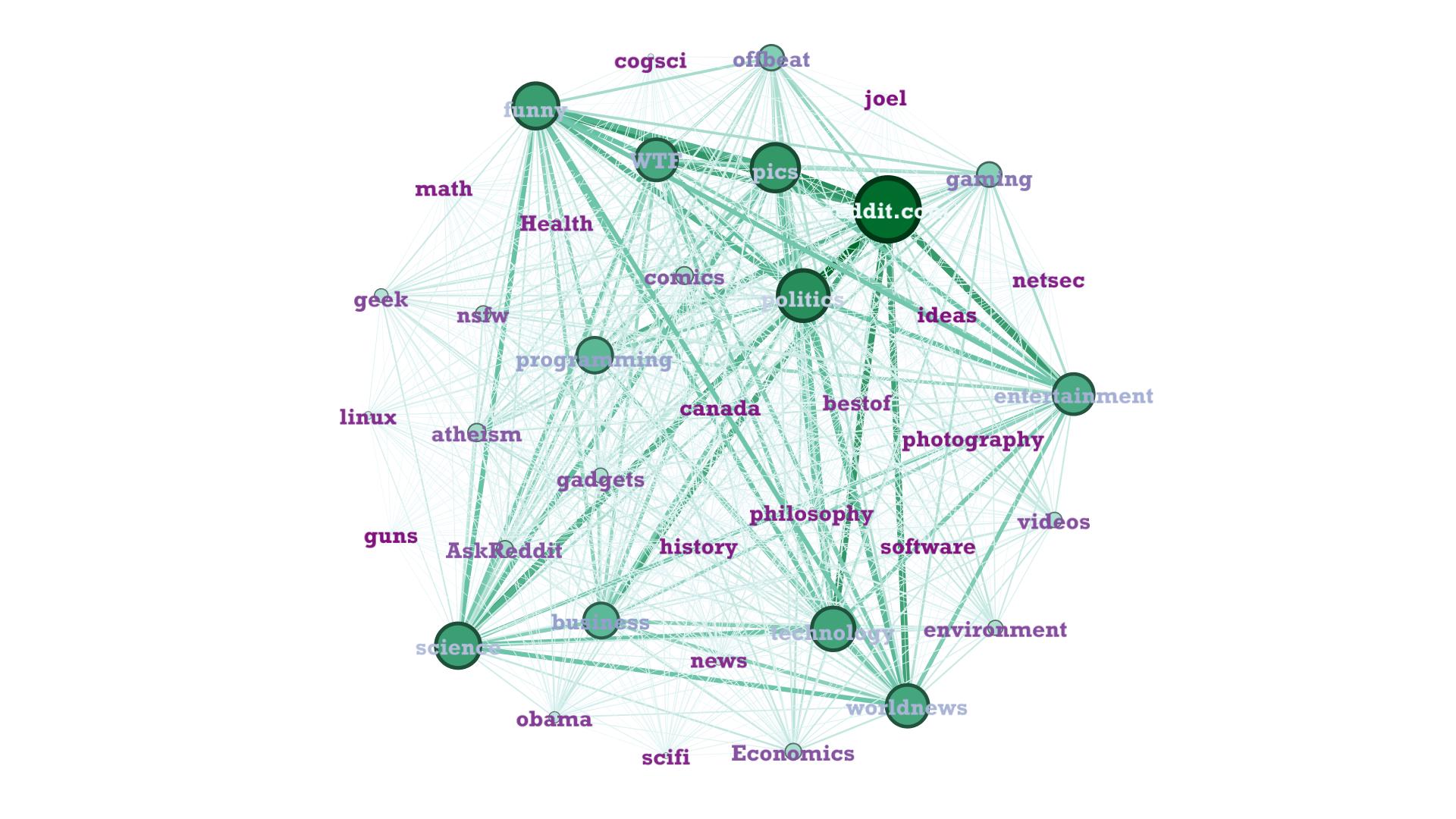
|  |  |  |
| --- | --- | --- |
|  | Пре филтрирања | После филтрирања |
| Број грана | 137.264 | 9.913 |



Слика 2.3 Мрежа SNetF

## Targeted subreddit network – SNetT

Још једна подмрежа првобитне SNet мреже с тим да се сада уклањају сви сабредити који не припадају задатом скупу сабредита чија тематика је блиска теми економске кризе. Укупан број сабредита је **4191**, док подскуп који се посматра броји **39** сабредита.

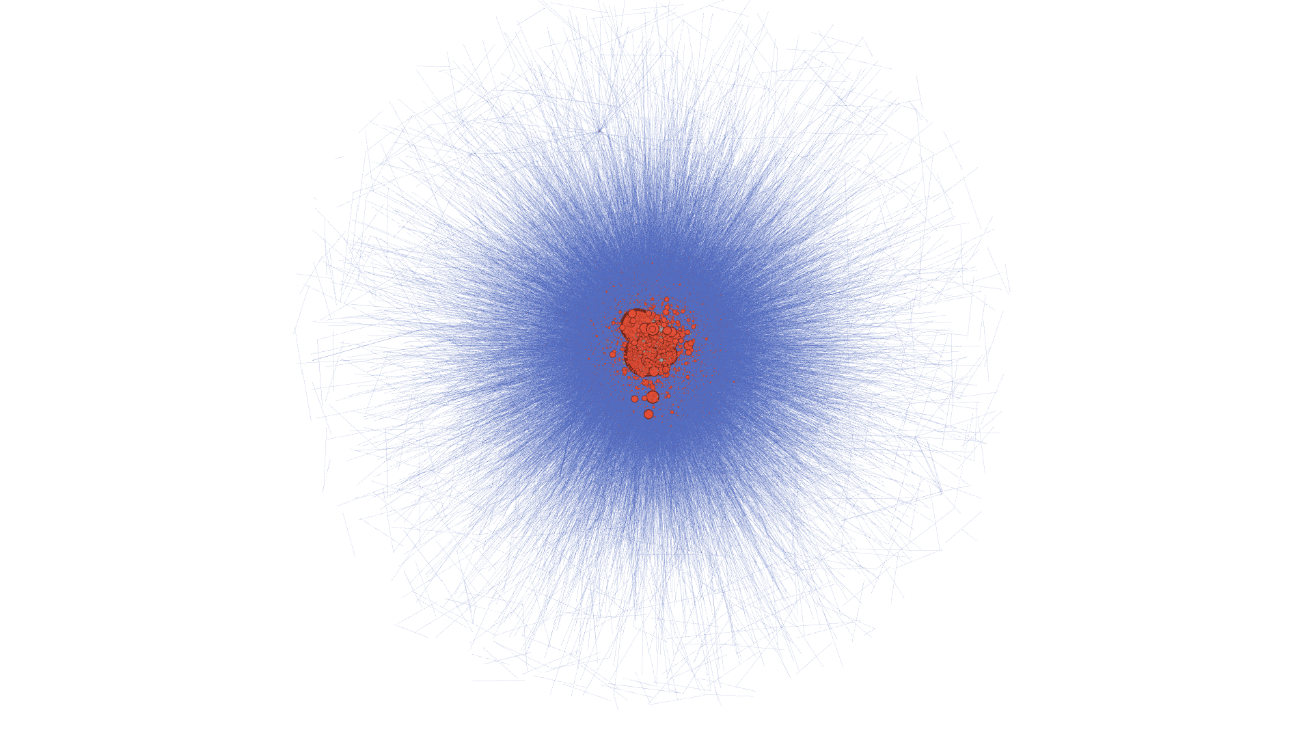


Слика 2.4 Мрежа SNetT

## 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

# 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

1. 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**

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

Табела 3.2 Анализа корисника

1. Који корисници су активни на највећем броју сабредита? На колико су сабредита активни?

Табела 3.3 Најактивнији корисници



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

Chart, scatter chart

Description automatically generated

Слика 3.1 Пирсонов коефицијент корелације

1. 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



\*num\_comments – број коментара на објави из истоимене колоне поченог сета података

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

# Основна анализа моделованих мрежа

1. Колика је густина мреже?
2. Колике су просечне дистанце у оквиру мреже и дијаметар мреже?

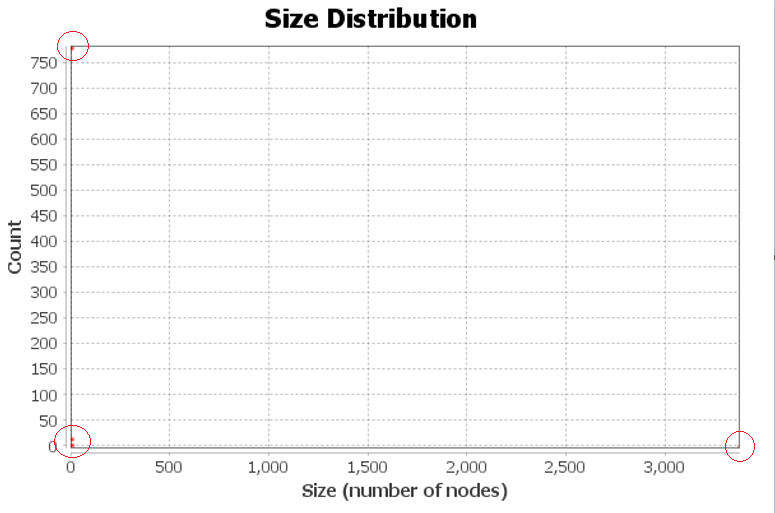
Табела 4.1 Густина и дијаметар мреже

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 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 |

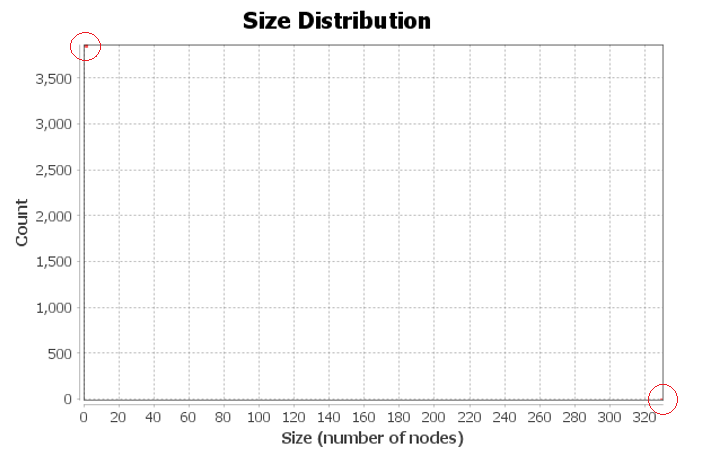
1. У којој мери је мрежа повезана и централизована? Навести број и величине повезаних компонената и проценити да ли постоји гигантска компонента.

Табела 4.2 Број повезаних компонената мреже

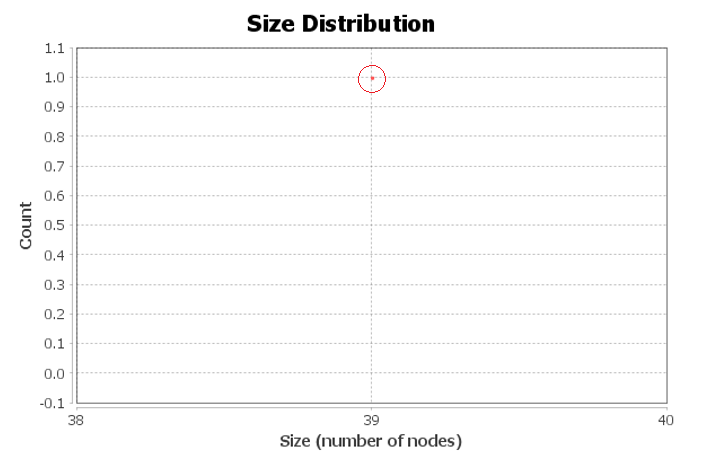
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | SNet | SNetF | SNetT | UserNet |
| Број повезаних компонената | 799 | 3863 | 1 | 201 |



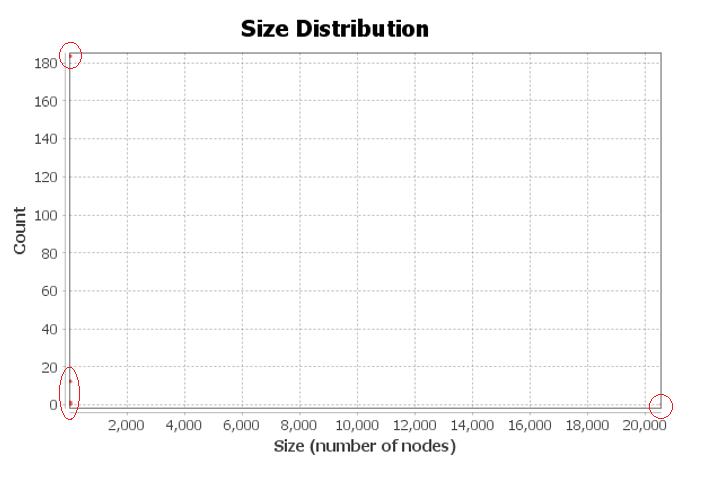
Слика 4.1 Величина компонената мреже SNet



Слика 4.2 Величина компонената 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.

1. 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.

Chart

Description automatically generatedChart, histogram

Description automatically generated

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 саСлика 2.1, па самим тим кластерисање ове мреже јесте изражено.

Chart

Description automatically generatedChart

Description automatically generated

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.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

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.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

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 слабој повезаности чворова, него што је то случај у осталим мрежама. Са друге стране, уочава се велика сличност расподеле оригиналне и случајно генерисане мреже. Да ли је кластерисање изражено?

1. 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 | НE (низак степен кластеризације) |
| SNetT | 1.0 | 1.0 | ДА |
| UserNet | 4.196 | 0.028932 | НE (низак степен кластеризације) |

1. 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 Aсортативност 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 | - | | 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).

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

Chart

Description automatically generated

Слика 4.9 Kоефицијент клуба богатих мреже SNet

Chart, line chart

Description automatically generated

Слика 4.10 Коефицијент клуба богатих мреже SNetF

Chart, box and whisker chart

Description automatically generated

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 притупе платформи ће највероватније постати чланови тих сабредита (*The rich get richer*).

1. Каква је дистрибуција чворова по степену и да ли прати *power law* расподелу?

Table

Description automatically generated

Слика 4.12 Дистрибуција чворова по степену мреже SNet

Chart, table

Description automatically generated

Слика 4.13 Дистрибуција чворова по степену мреже SNetF

Table

Description automatically generated

Слика 4.14 Дистрибуција чворова по степену мреже SNetT

Table

Description automatically generated  
  
Table

Description automatically generated  
  
Table

Description automatically generated

Слика 4.15 Дистрибуција чворова по степену мреже UserNet

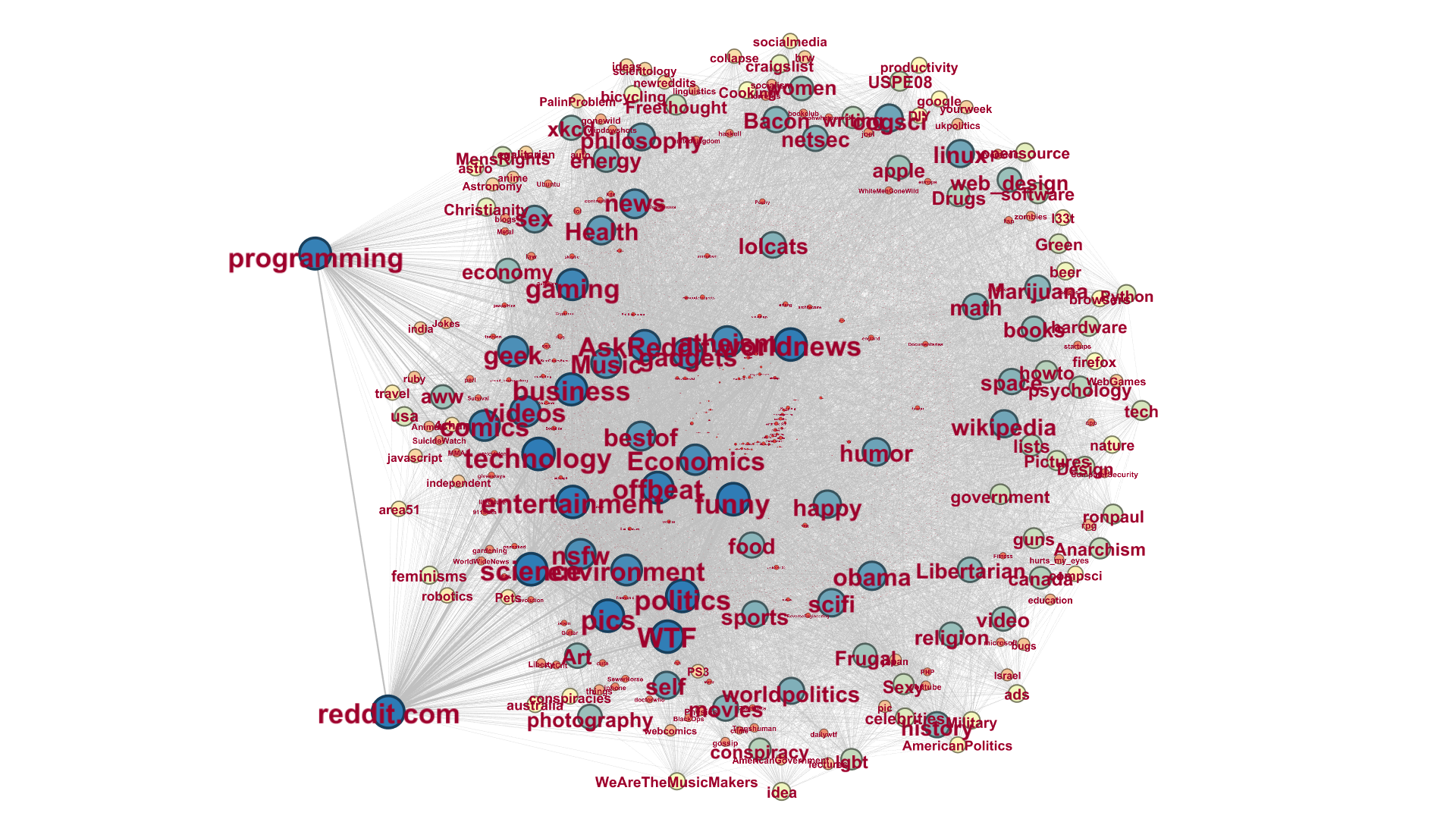
Табела 4.5 Преглед scale-free особине мрежа

|  |  |
| --- | --- |
|  | *Scale-free* мрежа |
| SNet | ДА (*truncated power law*) |
| SNetF | НЕ |
| SNetT | НЕ |
| UserNet | НЕ (вредност параметра *p* није довољно мала) |

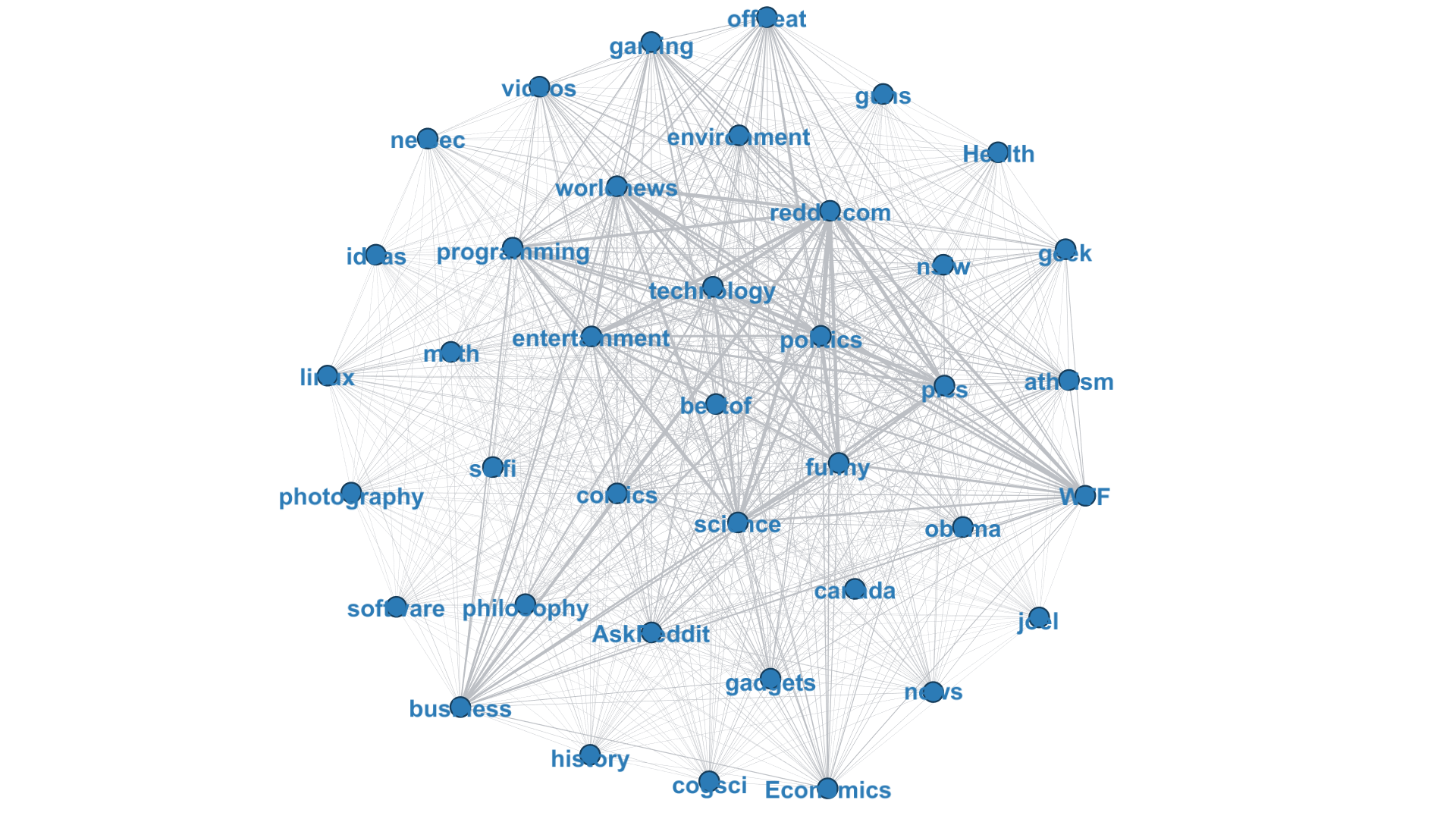
1. Одредити најважније хабове и ауторитете у мрежи. Како су они распоређени и уграђени у мрежу, да ли су на периферији или у језгру мреже?



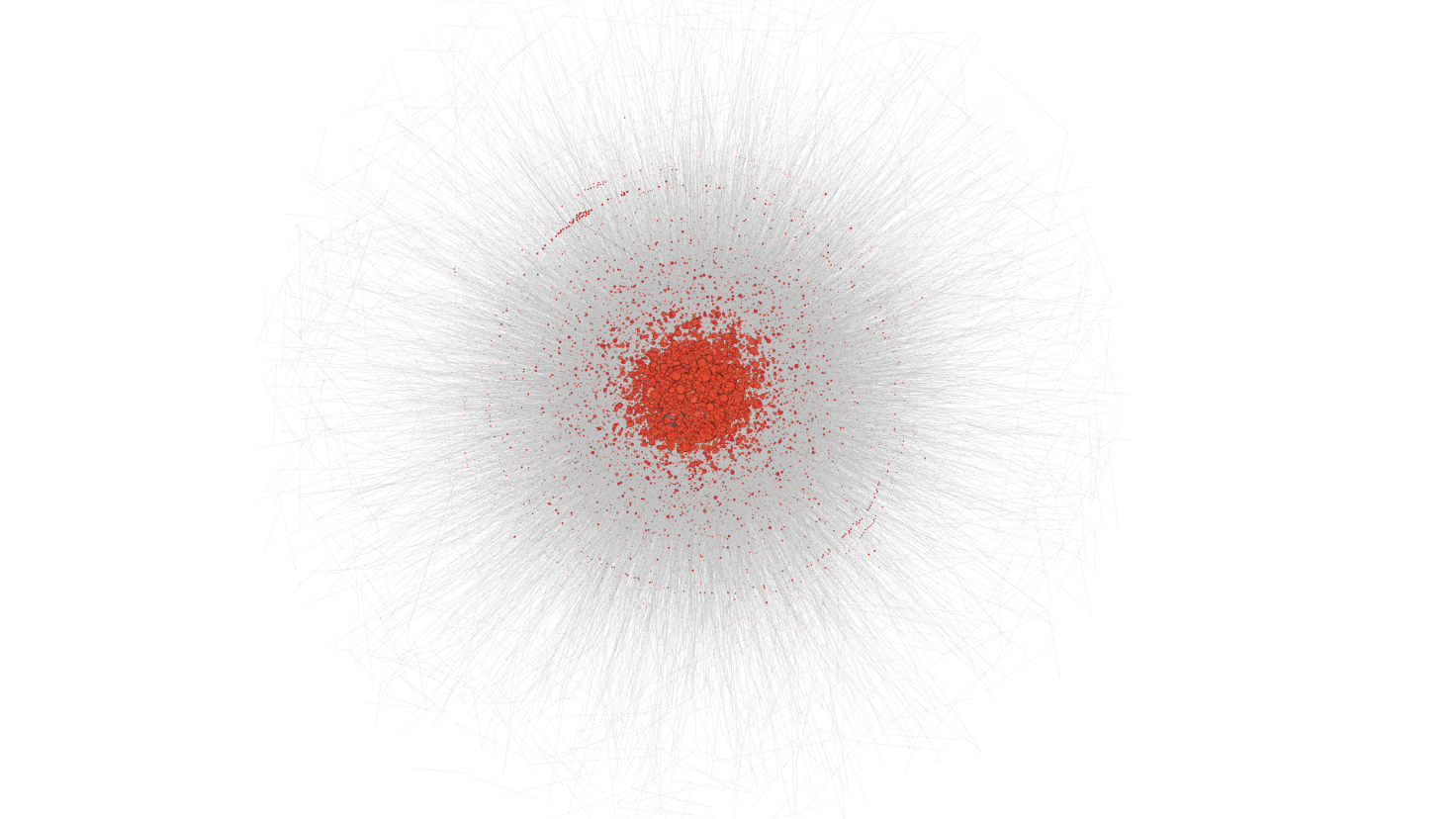
Слика 4.16 Хабови мреже SNet



Слика 4.17 Хабови мреже SNetF



Слика 4.18 Хабови мреже SNetT



Слика 4.19 Хабови мреже UserNet

# Анализа мера централности

1. 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 |
| alllie | 0.066303881 | **alllie** | 0.298119079 | **alllie** | 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 |

1. 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 | |

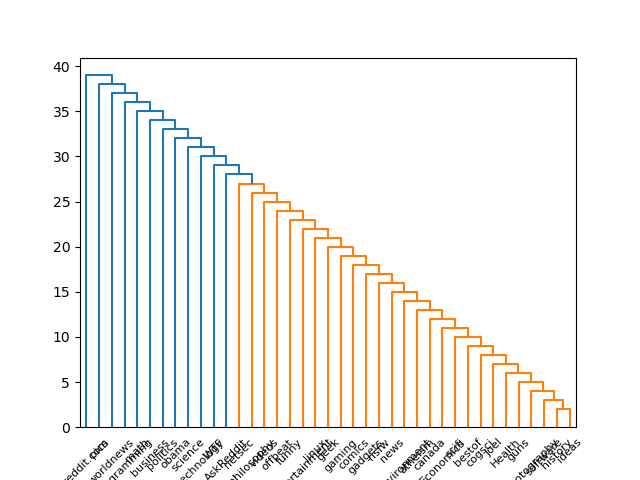
1. Рангирати чворове по Кацовој централности (енг. *Katz centrality*) са варијацијом параметара. При рачунању Кацове централности, експериментисати са додељивањем другачије вредности параметра β за сабредит који се у приложеним CSV фајловима идентификује врeдношћу колоне ***subreddit*** једнаком *reddit.com*. 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 SNetТ   |  |  | | --- | --- | |  | 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 | | alllie | 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 | | |

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

# Детекција комуна

1. Ако величина мреже дозвољава, спектралном анализом или анализом дендрограма проценити потенцијалне кандидате за број комуна у мрежи.
   * *SNet* мрежа
   * *SNetT* мрежа



Слика 6.1 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.

* + *SNetF* мрежа
  + *UserNet* мрежа

1. 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**

Diagram

Description automatically generatedChart, table

Description automatically generatedA picture containing shape

Description automatically generated

Слика 6.2 Кластерисање мреже SNet Лувенском методом

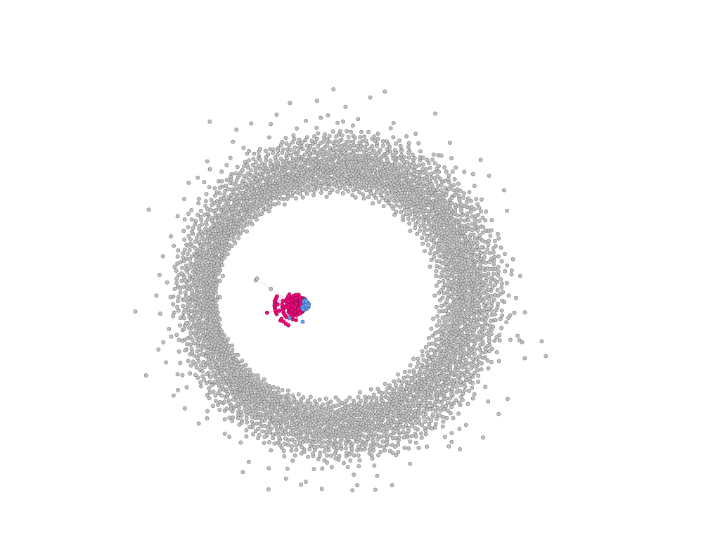
Кластерисањем мреже *SNetF* Лувенском методом издвојиле су се четири комуне веће од осталих. Чворови исте комуне обојеи су истом бојом као на слици изнад. Највећа детектована комуна броји 30,35% чворова, наредна 27,42%, а трећа највећа 7,59%.

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 6.3 Кластерисање мреже SNetF Лувенском методом

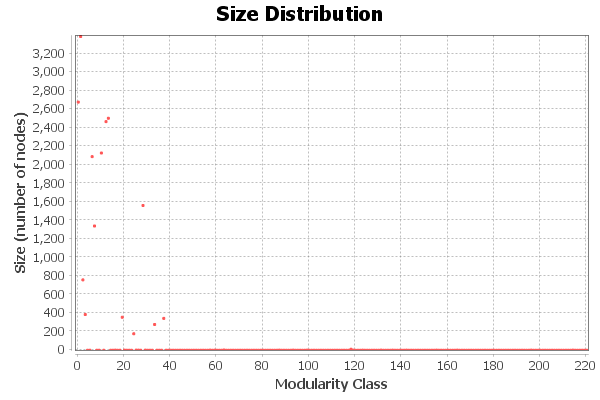
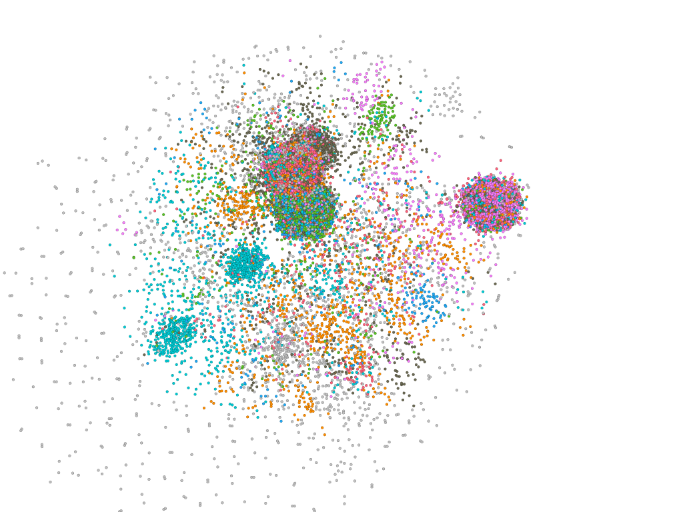
Издвојиле су се две веће комуне кластерисањем мреже *SNetF* Лувенском методом. Чворови једне комуне обојени су розе бојом и обухватају 4,22%, односно 176 чворова. Друга комуна, представљена је чворовима плаве боје, којих има 3,56%, односно 146.

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 садржи мање од 0,05% укупног броја чворова.

Резултат:

* Modularity: 0.080
* Modularity with resolution: 0.080
* Number of Communities: 3865
  + ***UserNet* мрежа**

A picture containing shape

Description automatically generated

Слика 6.4 Кластерисање мреже UserNet Лувенском методом

Кластерисањем мреже *UserNet* Лувенском методом, издвојено је шест комуна са већим бројем чворова у односу на остале. Комуне 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 садржи мање од 0,05% укупног броја чворова.

Резултат:

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

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

* *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 сабредити 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.

1. Ко су актери који се могу окарактерисати као кључни брокери (мостови) у мрежи? Шта их чини брокерима?
   * *SNet* мрежа
   * *SNetT* мрежа – сваки чвор може бити окарактерисан као брокер (мост), сви су једнаког значаја у мрежи.
   * *SNetF* мрежа
   * *UserNet* мрежа

# Поређење SNet и SNetT мрежа

1. Упоредити карактеристике SNet и SNetT мрежа. Коментарисати потенцијалне разлике и проценити да ли су сабредити из SNetT активнији и боље повезани од остатка мреже.
2. Како су распоређени чворови из SNetT у оквиру SNet мреже? Да ли припадају језгру или периферији или су мешовито распоређени?

Списак слика

[Слика 2.1 Мрежа SNet 5](#_Toc95338928)

[Слика 2.2 Расподела тежина грана 6](#_Toc95338929)

[Слика 2.3 Мрежа SNetF 7](#_Toc95338930)

[Слика 2.4 Мрежа SNetT 8](#_Toc95338931)

[Слика 2.5 Мрежа UserNet 9](#_Toc95338932)

[Слика 3.1 Пирсонов коефицијент корелације 12](#_Toc95338933)

[Слика 4.1 Величина компонената мреже SNet 14](#_Toc95338934)

[Слика 4.2 Величина компонената мреже SNetF 14](#_Toc95338935)

[Слика 4.3 Величина компонената мреже SNetT 15](#_Toc95338936)

[Слика 4.4 Величина компонената мреже UserNet 15](#_Toc95338937)

[Слика 4.5 Расподела коефицијента кластеризације мреже SNet и случајне мреже истих димензија 16](#_Toc95338938)

[Слика 4.6 Расподела коефицијента кластеризације мреже SNetT и случајне мреже истих димензија 17](#_Toc95338939)

[Слика 4.7 Расподела коефицијента кластеризације мреже SNetF и случајне мреже истих димензија 17](#_Toc95338940)

[Слика 4.8 Расподела коефицијента кластеризације мреже UserNet и случајне мреже истих димензија 18](#_Toc95338941)

[Слика 4.9 Kоефицијент клуба богатих мреже SNet 19](#_Toc95338942)

[Слика 4.10 Коефицијент клуба богатих мреже SNetF 20](#_Toc95338943)

[Слика 4.11 Коефицијент клуба богатих мреже SNetT 20](#_Toc95338944)

[Слика 4.12 Дистрибуција чворова по степену мреже SNet 21](#_Toc95338945)

[Слика 4.13 Дистрибуција чворова по степену мреже SNetF 21](#_Toc95338946)

[Слика 4.14 Дистрибуција чворова по степену мреже SNetT 21](#_Toc95338947)

[Слика 4.15 Дистрибуција чворова по степену мреже UserNet 22](#_Toc95338948)

[Слика 4.16 Хабови мреже SNet 23](#_Toc95338949)

[Слика 4.17 Хабови мреже SNetF 24](#_Toc95338950)

[Слика 4.18 Хабови мреже SNetT 25](#_Toc95338951)

[Слика 4.19 Хабови мреже UserNet 26](#_Toc95338952)

Списак табела

[Табела 1.1 Утицај чишћења на величину скупа података 4](#_Toc95338953)

[Табела 2.1 Промена броја грана приликом филтрације 6](#_Toc95338954)

[Табела 3.1 Анализа сабредита 10](#_Toc95338955)

[Табела 3.2 Анализа корисника 11](#_Toc95338956)

[Табела 3.3 Најактивнији корисници 11](#_Toc95338957)

[Табела 3.4 Објаве са највећим бројем коментара 12](#_Toc95338958)

[Табела 4.1 Густина и дијаметар мреже 13](#_Toc95338959)

[Табела 4.2 Број повезаних компонената мреже 13](#_Toc95338960)

[Табела 4.3 Коефицијент кластеризације мреже 16](#_Toc95338961)

[Табела 4.4 Aсортативност по степену чвора 19](#_Toc95338962)

[Табела 4.5 Преглед scale-free особине мрежа 23](#_Toc95338963)

[Табела 5.1 Преглед најважнијих актера мреже SNet према централности 27](#_Toc95338964)

[Табела 5.2 Преглед најважнијих актера мреже SNetF према централности 28](#_Toc95338965)

[Табела 5.3 Преглед најважнијих актера мреже SNetT према централности 28](#_Toc95338966)

[Табела 5.4 Преглед најважнијих актера мреже UserNet према централности 29](#_Toc95338967)

[Табела 5.5 Централност по сопственом вектору мреже SNet 30](#_Toc95338968)

[Табела 5.6 Централност по сопственом вектору мреже SNetF 30](#_Toc95338969)

[Табела 5.7 Централност по сопственом вектору мреже SNetT 30](#_Toc95338970)

[Табела 5.8 Централност по сопственом вектору мреже UserNet 30](#_Toc95338971)

[Табела 5.9 Katz централност 32](#_Toc95338972)

[Табела 5.10 Модификована Katz централност мреже SNet 32](#_Toc95338973)

[Табела 5.11 Katz централност 32](#_Toc95338974)

[Табела 5.12 Модификована Katz централност мреже SNetF 32](#_Toc95338975)

[Табела 5.13 Katz централност 33](#_Toc95338976)

[Табела 5.14 Модификована Katz централност мреже SNetТ 33](#_Toc95338977)

[Табела 5.15 Katz централност мреже UserNet 33](#_Toc95338978)