Универзитет у Београду

Електротехнички факултет



Анализа социјалних мрежа

Извештај за други пројектни задатак

|  |
| --- |
| Студенти: |
| Марија Андрејевић 2020/3226 Невена Цветковић 2020/3233 |

Београд, Септембар 2021.

Садржај

[Садржај 2](#_Toc81442140)

[1. Увод 3](#_Toc81442141)

[2. Поступак израде 4](#_Toc81442142)

[2.1. Генерисање секундарног скупа података и графова мреже 4](#_Toc81442143)

[3. Одговори на истраживачка питања 6](#_Toc81442144)

[Списак скраћеница 29](#_Toc81442145)

[Списак слика 30](#_Toc81442146)

[Списак табела 31](#_Toc81442147)

1. Увод

Приликом израде пројектног задатка на предмету анализа социјалних мрежа, вршена је анализа стања професионалног мушког тениса у периоду од 2018. до 2020. године. У овом извештају наведени су резултати анализе и одговори на постављена истраживачка питања.

За испитивање података коришћен је програмски језик *Python* са својим библиотекама *NetworkX* и *pandas* за анализу и визуелизацију података и мрежа. Окружење у којем је вршена анализа је *Jupyter Notebook* *.* За финију визуелизацију коришћен је алат *Gephi*.

У наредним поглављљима ће бити дат поступак израде пројектног задатка, биће наведена постављена истраживачка питања и дати донети закључци.

1. Поступак израде

У оквиру израде пројектног задатка вршена је квантитативна и квалитативна аналза стања професионалног мушког тениса у сингл конкуренцији применом техника за анализу социјалних и комплексних мрежа и статистичких метода.

Примарни скуп података је преузет са репозиторијума који одржава Џеф Сакман. Подаци су организовани у виду неколико CSV датотека које садрже податке о мечевима у периоду од 2018., до 2020., податке о свим икада рангираним играчима, податке о рангирању до 2019. године и ранг активних тенисера на ATP листи током 2020. године.

На основу примарног скупа података је помоћу библиотека програмског језика *Python* креиран секундарни скуп података. Користећи секундарни скуп података је потом генерисан граф који представља мрежу тенисера и одговорено је на постављена истраживачка питања.

* 1. Генерисање секундарног скупа података и графова мреже

Подаци из датих CSV датотека су учитани помоћу *Pandas* библиотеке и сачуване у променљиве типа pandas.DataFrame Потом је за сваки од скупова података анализирана његова структура, проверавано да ли су подаци валидни и извршена је селекција колона које су потребне за даљи рад. Такође, одрађена је и анализа кандидата поља која би могла да буду примарни кључ и ако су таква недостајала, додата је аутоматски генерисана идентификациона колона.

Тако пречишћен скуп података представља секундарни скуп података. Овај скуп је потом искоришћен за генерисање неусмереног графа мреже.

Чворови графа су играчи, а гране које их повезују међусобно одиграни мечеви. Креирана су три графа, за мечеве који су одиграни по годинама, за 2018., 2019. и 2020. годину. Генерисан је и агрегиран граф који садржи унију података за све три године. Гране графа представљају међусобне сусрете тенисера. За гране у графовима по годинама, као тежина гране узет је број међусобних сусрета тенисера. У агрегираном графу, тежине грана су формиране као број мечева тенисера, где мечеви из 2018. и 2019. године имају тежину један, а мечеви из 2020. године у коначан збир тежина улазе помножени фактором 0.5. Сматра се да су мечеви из 2020. године имали мању важност јер је стање у тенису те године било под великим утицајем епидемије корона вируса и велик број турнира није ни одржан.

На основу секундарног скупа података и генерисаних графова приступљено је анализи социјалне мреже.

1. Одговори на истраживачка питања

У наредном поглављу ће бити приказан списак истраживачких питања и одговора. Део анализе и сам поступак израде налазе се у приложеним *Jupyter Notebook* фајловима - *2018.ipynb*, *2019.ipynb*, *2020.ipynb* и *agregated.ipynb*.

1. Колики је просечан број тенисера (саиграча) по сваком тенисеру?

Просечан број саиграча по сваком тенисеру је заправо просечан степен чвора графа. Овај податак се у алату *Gephi* након што се учита неки од појединачних графова може пронаћи у менију *Statistics -> Network Overview -> Average Degree*. По годинама, резултати су 2018 – 11.881, 2019 – 13.066, 2020 – 7.681 и агрегирано – 18.348.

1. Који тенисери су се сусретали са највећим бројем других тенисера?

У наредној табели је дат преглед по пет тенисера који су се сусретали са највећим бројем тенисера. У колони бр тен се налази број тенисера са којима су се сусрели.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **2018.** | | **2019.** | | **2020.** | | **Агрегирани подаци** | |
| **Име** | **Бр тен** | **Име** | **Бр тен** | **Име** | **Бр тен** | **Име** | **Бр тен** |
| F Fognini | 61 | D Medvedev | 60 | A Rublev | 43 | K Khachanov | 103 |
| D Thiem | 55 | S Tsitsipas | 51 | F A Aliassime | 37 | S Tsitsipas | 99 |
| S Tsitsipas | 53 | A Zverev | 51 | S Tsitsipas | 34 | Daniil Medvedev | 97 |
| A Zverev | 52 | B Paire | 51 | N Djokovic | 32 | F Fognini | 96 |
| R Gasquet | 51 | D Schwartzman | 50 | K Khachanov | 31 | A Zverev | 95 |

Табела 1 Тенисери који имају највише сусрета са другим тенисерима

1. Који тенисери су учествовали на највећем броју (различитих) турнира?

У наредној табели је дат преглед тенисера који су учествовали на највећем броју различитих турнира (колона Бр тур).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **2018.** | | **2019.** | | **2020.** | | **Агрегирани подаци** | |
| **Име** | **Бр тур** | **Име** | **Бр тур** | **Име** | **Бр тур** | **Име** | **Бр тур** |
| D Dzumhur | 32 | D Shapovalov | 32 | A Mannarino | 17 | D Shapovalov | 75 |
| Albert Ramos | 30 | K Khachanov | 31 | F A Aliassime | 17 | A Mannarino | 74 |
| Robin Haase | 30 | Taylor Fritz | 31 | A Bublik | 17 | B Paire | 73 |
| M Zverev | 30 | B Paire | 31 | John Millma | 32 | J L Struff | 71 |
| A Mannarino | 30 | J L Struff | 16 | J L Struff | 16 | S Tsitsipas | 70 |

Табела 2 Тенисери који су учесвовали на највећем броју турнира

1. Који тенисери су добри кандидати за представнике професионалних тенисера? Да ли и шта се мења уколико се уместо једног бира скуп од неколико представника? Добар представник или представници би требало и да буду у контакту са што већим скупом тенисера.

Као што је у тексту задатка наведено, добар представник тенисера треба да буде у контакту са што већим скупом тенисера. Користећи Degree centrality меру централности анализирано је који чворови имају везе са највећим бројем других чворова. Листа тенисера је дата у табели 3, са њиховим ранговима.

1. Како су рангирани на ATP листи тенисери који су се сусретали са највећим бројем других тенисера?

Као и за питање број четири, тенисери који су се сусретали са највећим бројем других тенисера су одређени користећи Degree centrality меру централности. У следећој табели су дати резултати по годинама, имена тенисера и број играча са којима су се сусрели.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **2018.** | | **2019.** | | **2020.** | | **Агрегирани подаци** | |
| **Име** | **Бр тен** | **Име** | **Бр тен** | **Име** | **Бр тен** | **Име** | **Бр тен** |
| F Fognini | 13 | D Medvedev | 5 | A Rublev | 8 | K Khachanov | 20 |
| D Thiem | 8 | B Paire | 24 | F A Aliassime | 21 | S Tsitsipas | 6 |
| S Tsitsipas | 6 | S Tsitsipas | 6 | S Tsitsipas | 6 | D Medvedev | 4 |
| A Zverev | 4 | A Zverev | 7 | N Djokovic | 1 | F Fognini | 17 |
| R Gasquet | 26 | D Schwartzman | 14 | K Khachanov | 20 | A Zverev | 7 |

Табела 3 Тенисери који су се сусретали са највећим бројем тенисера

1. Из којих земаља долази највећи број активних играча у протеклом периоду?

Из скупа свих играча су селектовани они који су играли бар један меч у 2018., 2019. или 2020. години, као и у агрегираном скупу мечева. Добијени резултат је затим груписан по држави порекла и сортиран опадајуће по броју тенисера из државе. Добијени резултати су дати у следећој табели. Наведена су имена држава и број активних играча из тих држава.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2018.** | | **2019.** | | **2020.** | | | **Агрегирани подаци** | |
| **Држава** | **Бр тен** | **Држава** | **Бр тен** | | **Држава** | **Бр тен** | **Држава** | **Бр тен** |
| USA | 34 | USA | 31 | | USA | 28 | USA | 44 |
| FRA | 25 | GER | 22 | | FRA | 22 | FRA | 33 |
| ESP | 24 | FRA | 22 | | ITA | 18 | ITA | 27 |
| GER | 18 | ITA | 21 | | GER | 15 | ESP | 27 |
| ITA | 17 | ESP | 18 | | AUS | 14 | GER | 26 |
| ARG | 15 | ARG | 13 | | ESP | 14 | ARG | 19 |
| AUS | 14 | AUS | 12 | | ARG | 13 | AUS | 17 |
| AUT | 8 | SRB | 9 | | RUS | 9 | RUS | 15 |

Табела 4 Државе из којих долази највећи број играча

1. Из којих земаља долазе најуспешнији играчи у смслу освојених поена на ATP листи?

Издвојен је скуп првопласираних тридесет играча, са подацима о играчима, рангу и поенима на ATP листи. Потом су подаци груписани по држави и тако је добијена табела 6 у којој се налазе подаци о имену државе и броју тенисера из те државе који су у првих 30. Изостављени су из приказа агрегирани подаци, јер су исти као и за 2020. годину.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2018.** | | | **2019.** | | | **2020.** | | |
| **Држ** | **Име** | **Поени** | **Држ** | **Име** | **Поени** | **Држ** | **Име** | **Поени** |
| SRB | Djokovic | 9045.0 | ESP | Nadal | 9985.0 | SRB | Djokovic | 12030.0 |
| ESP | Nadal | 7480.0 | SRB | Djokovic | 9145.0 | ESP | Nadal | 9850.0 |
| SUI | Federer | 6420.0 | SUI | Federer | 6590.0 | AUT | Thiem | 9125.0 |
| GER | Zverev | 6385.0 | AUT | Thiem | 5825.0 | RUS | Medvedev | 8470.0 |
| ARG | Del Potro | 5300.0 | RUS | Medvedev | 5705.0 | SUI | Federer | 6630.0 |
| RSA | Anderson | 4710.0 | GRE | Tsitsipas | 5300.0 | GRE | Tsitsipas | 5925.0 |
| CRO | Cilic | 4250.0 | GER | Zverev | 3345.0 | GER | Zverev | 5525.0 |
| AUT | Thiem | 4095.0 | ITA | Berrettini | 2870.0 | RUS | Rublev | 4119.0 |

Табела 5 Првих осам тенисера на ранг листи по годинама

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2018.** | | **2019.** | | **2020.** | |
| **Држава** | **Број тенисера** | **Држава** | **Број тенисера** | **Држава** | **Број тенисера** |
| ESP | 4 | FRA | 4 | ESP | 3 |
| FRA | 3 | RUS | 3 | FRA | 3 |
| ARG | 2 | ESP | 3 | CAN | 3 |
| RUS | 2 | SUI | 2 | RUS | 3 |
| CAN | 2 | AUS | 2 | SUI | 2 |
| ITA | 2 | CAN | 2 | USA | 2 |
| CRO | 2 | ARG | 2 | ITA | 2 |
| GBR | 1 | ITA | 2 | SRB | 2 |

Табела 6 Првих осам држава по броју тенисера у првих 30 по годинама

1. Како су у оквиру скупа података окарактерисани играчи из Србије?

Када се посматра који тенисери су се сусретали са највећим бројем других тенисера, за 2018. и 2019. годину, у првих 35 најбољих играча у овој категорији се налазе два српска тенисера, Новак Ђоковић и Душан Лајовић. У 2020. години у првих 30 играча се налазе три тенисера, Лајовић, Кецмановић и Ђоковић који заузима четврто место. У агрегираним подацима се истиче Душан Лајовић који заузима петнаесто место са 89 противника.

У категорији највећег броја одиграних турнира, српски тенисер Душан Лајовић се у сваком скупу података налази међу првих тридесет. Такође, Миомир Кецмановић се у 2020. години нашао међу најбољих тридесет тенисера по броју одиграних турнира.

Што се тиче најбољих представника тенисера, Новак Ђоковић се у 2020. години истакао као четврти најбољи представник са вредношћу 0.09302325581395349 параметра Degree centrality. У агрегираном скупу података Душан Лајовић је заузео шеснаесто место.

Србија је у 2018. години имала 7 активних играча, у 2019. години 9, у 2020. години 8, а у агрегираном скупу података за све три године је имала 10 играча.

По броју најуспешнијих играча по поенима на ATP листи, у првих 30 играча, Србија се у 2018. и 2019. години нашла на четвртом месту са једним тенисером, а у 2020. години и у агрегираном скупу података Србија је имала два играча у најбољих 30 и тако заузела друго место.

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

При анализи комуна/кластера у мрежи коришћена је метрика модуларности и анализа доступна и алату *Gephi*. Модуларност је мера квалитета партиционисања чворова мреже у одговарајуће кластере. Приказана су два начина поделе у кластере – користећи метрику модуларности и помоћу top-down Girvan-Newman метода.

Уочено је да су мреже са већом модуларношћу, подељене на кластере који су унутар себе густо повезани, а немају велики број веза са остатком мреже. Приложене уз овај документ се могу наћи слике мрежа у којима су означене комуне са највећом модуларншћу.

Користећи Girvan-Newman метод, по годинама, добијено је 2018. – 19 комуна, 2019. – 11 комуна. 2020. – 26 комуна и у агрегираним подацима 19 комуна. Свака од мрежа има једну гигантску комуну/кластер, док су остале знатно мање.

1. Спровести анализу кластерисања и асортативну анализу, па упоредити груписање на основу земље из које играчи долазе, на основу броја мечева које играју и на основу рејтинга.

У наредној табели су дате вредности параметара који се испитују у оквиру анализе кластерисања и асортативне анализе. Све вредности су израчунате користећи функције библиотеке *NetworkX*.

Вредности просечног коефицијента кластеризације за сваки од графова је реда величине 0,1, с тим да је вредност овог параметра знатно мања за 2020. годину, што је узроковано мањим бројем мечева у овој години. Максималан локални степен кластерисања има вредности реда величине 0.2 ~ 0.3. Коефицијенти асортативности на основу тежинског, нетежинског степена чвора и на основу државе порекла тенисера имају позитивне вредности, реда величине до ~0.2. Њихове позитивне вредности говоре да чворови сличних вредности атрибута (држава, нетежински и тежински степен чвора) теже да се повезују једни са другима. Негативне вредности коефицијента асортативности на основу ранга имају негативне вредности јер тенисери високог ранга чешће играју са тенисерима ниског ранга, него са тенисерима сличним себи. Вредност коефицијента асортативности на основу тежинског степена чвора има нема наведену вредност јер није било могуће израчунати овај параметар због тога што је атрибут тежине чвора у агрегираном графу реалан број.

|  |  |  |
| --- | --- | --- |
|  | **Име параметра** | **Вредност** |
| **2018** | Максималан локални степен кластерисања  Тенисер | 0.20000000000000004  Alexander Zverev |
| Просечан степен кластерисања | 0.12227548001603945 |
| Коефицијент асортативности – нетежински степен чвора | 0.21868975600564688 |
| Коефицијент асортативности – тежински степен чвора | 0.22945507317379313 |
| Коефицијент асортативности – држава | 0.017737013898119852 |
| Коефицијент асортативности – ранг | -0.005617682736737482 |
| **2019** | Максималан локални степен кластерисања  Тенисер | 0.2519842099789747  Daniil Medvedev |
| Просечан степен кластерисања | 0.13785479448611918 |
| Коефицијент асортативности – нетежински степен чвора | 0.1893631627055268 |
| Коефицијент асортативности – тежински степен чвора | 0.18750548289093133 |
| Коефицијент асортативности – држава | 0.00474047371358742 |
| Коефицијент асортативности – ранг | -0.0069502359615036206 |
| **2020** | Максималан локални степен кластерисања  Тенисер | 0.33333333333333337  Yen Hsun Lu |
| Просечан степен кластерисања | 0.09086792378213669 |
| Коефицијент асортативности – нетежински степен чвора | 0.23815555973416275 |
| Коефицијент асортативности – тежински степен чвора | 0.23637924173286778 |
| Коефицијент асортативности – држава | 0.0066677316949746955 |
| Коефицијент асортативности – ранг | -0.003351964280899542 |
| **Агрегирана вредност** | Максималан локални степен кластерисања  Тенисер | 0.16798947331931643  Julian Lenz |
| Просечан степен кластерисања | 0.1866939496703153 |
| Коефицијент асортативности – нетежински степен чвора | 0.16919387110544537 |
| Коефицијент асортативности – тежински степен чвора | - |
| Коефицијент асортативности – држава | 0.009843043206048577 |
| Коефицијент асортативности – ранг | -0.0038445015097198815 |

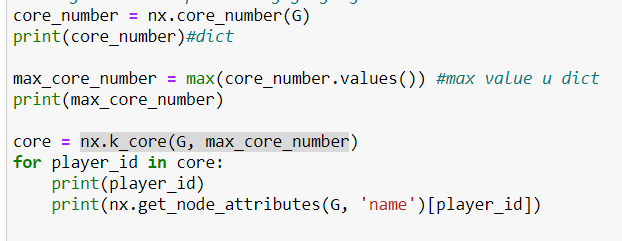
Табела 7 Вредности параметара за анализу кластерисања и асотативности

1. У којој мери тенисери имају тенденцију да се сусрећу са истим тенисерима? Да ли то утиче на њихово рангирање на ATP листи?

Како би се одредила тенденција тенисера да се сусрећу са истим тенисерима, одређиван је однос броја тежинских и нетежинских грана једног чвора. Просечан степен тенденције по годинама је 2018. - 1.1607071112896745, 2019. - 1.1694701429772918, 2020. - 1.0920754716981131, агрегирано - 1.1995309568480301. На основу резултата за појединачне играче може се закључити да тенисери већег ранга најчешће имају високу вредност степена тенденције тенисера да се сусрећу са истим тенисерима.

1. Који тенисери представљају језгро мреже?

Одређивање језгра мреже је рађено помоћу функција из библиотеке *NetworkX* користећи код приказан на слици.



Слика 1 Код за израчунавање језгра мреже

К-торка је максималан подграф која садржи чворове степена к или већег. Core number чвора у графу је величина највеће к-торке која садржи тај чвор. Функција nx.k\_core издваја подграф са максималном вредношћу core number параметра.

Језгра мреже су за све године величине 10 до 35 и већински садрже тенисере из првих 100, док свака од њих садржи тенисере првих десет тенисера.

1. Ко су тенисери који повезују различите групе у оквиру мреже?

Мостови (*bridge*) су гране у мрежи које повезују различите групе у оквиру мреже и њиховим уклањањем би настале две неповезане компоненте. Број бриџева у мрежи је 2018. – 122, 2019. – 115, 2020. – 115 и за агрегиране податке постоји 163 гране које представљају мост.

1. Колика је густина сваке од моделованих мрежа?

За израчунавање густине мрежа коришћена је функција библиотеке *NetworkX* и алат *Gephi*. Добијене су следеће вредности по годинама, прва бројка је добијена из *Gephi* алата, а друга помоћу *NetworkX*: 2018 – 0.028 / 0.028422651334345846, 2019 – 0.036 / 0.03599430872158145, 2020 – 0.022 / 0.022328951803168184 и агрегирано 0.032 / 0.03163392486200962.

1. У којој мери су мреже повезане и централизоване?

У наредној табели су дате вредности различитих врста централности по годинама и тенисер који је имао највећу вредност централности за дату годину. Вредност information\_centrality није могла да се израчуна јер ниједна од мрежа није повезана.

Како су максималне вредности Betweenness centrality јако ниске, а вредности Closeness centrality високе за сваки од наведених тенисера, а Betweenness centrality су за те чворове јако ниске што говори да се тај чвор налази на великом броју путања у мрежи, близак је другима, али су и они међусобно блиски.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Тип централности** | **Име и презиме** | **Вредност** |
| **2018** | Degree centrality: | Fabio Fognini | 0.14593301435406697 |
| Closeness centrality: | Fabio Fognini | 0.38868304352899985 |
| Betweenness centrality: | Pablo Cuevas | 0.06984202008882848 |
| **2019** | Degree centrality: | Daniil Medvedev | 0.1652892561983471 |
| Closeness centrality: | Daniil Medvedev | 0.4317160826594789 |
| Betweenness centrality: | Pablo Cuevas | 0.056098124415660075 |
| **2020** | Degree centrality: | Andrey Rublev | 0.10755813953488372 |
| Closeness centrality: | Andrey Rublev | 0.37152979651162793 |
| Betweenness centrality: | Novak Djokovic | 0.032802146597679875 |
| **Агрегирано** | Degree centrality: | Karen Khachanov | 0.17758620689655172 |
| Closeness centrality: | Karen Khachanov | 0.4315643185963693 |
| Betweenness centrality: | Stefanos Tsitsipas | 0.06815305924680275 |

Табела 8 Максималне вредности централности тенисера по годинама

1. Које су просечне дистанце, а колики дијаметар у оквиру моделованих мрежа?

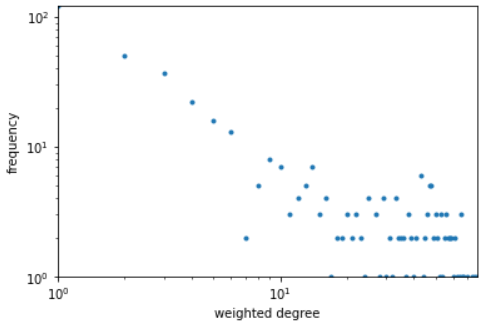
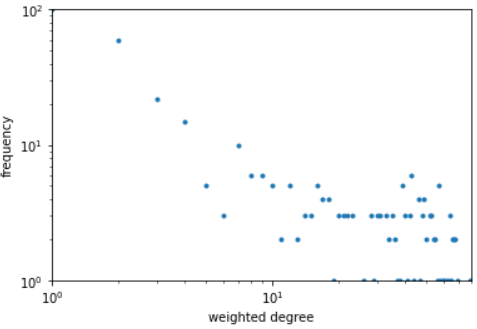
Просечне дистанце и дијаметар моделованих мрежа се могу пронаћи у оквиру алата *Gephi* и дате су у следећој табели.

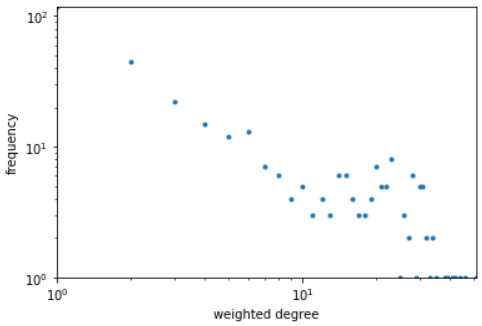
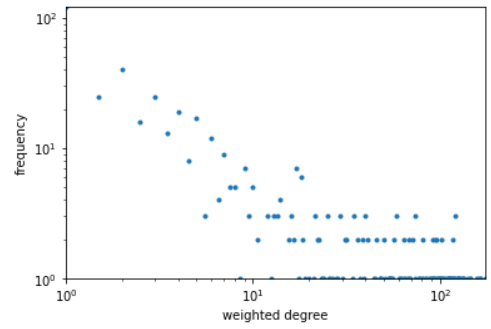
|  |  |  |
| --- | --- | --- |
|  | **Параметар** | **Вредности** |
| **2018** | Дистанца | 3.1371012285965936 |
| Дијаметар | 11 |
| **2019** | Дистанца | 3.045589369909987 |
| Дијаметар | 9 |
| **2020** | Дистанца | 3.1371012285965936 |
| Дијаметар | 11 |
| **Агрегирано** | Дистанца | 3.2336087527117874 |
| Дијаметар | 10 |

Табела 9 Вредности дистанце и дијаметра по годинама

1. Каква је дистрибуција чворова по степену и да ли прати неку закономерност? Како је степен чвора корелисан са рејтингом тенисера?

На следећим сликама су приказани графици дистрибуције чвора по тежинском степену чвора. Дистрибуција личи на *power-law (scale-free)* расподелу.

Слика 2 Дистрибуција чворова по степену за 2018., 2019., 2020. и агрегирано слева на десно

У наредној табели се налази група тенисера, са својим рангом и тежинским степеном чвора, по годинама. Примећује се да за већи степен чвора ранг има већу вредност, то јест тенисер је удаљенији на листи.

|  |  |  |  |
| --- | --- | --- | --- |
| **Година** | **Тенисер** | **Ранг** | **Степен чвора** |
| **2018** | Cheng Yu Yu | 883 | 1 |
| Mikhail Youzhny | 115 | 25 |
| Soon Woo Kwon | 235 | 2 |
| **2019** | Wishaya Trongcharoenchaikul | 625 | 1 |
| Gianluca Mager | 118 | 2 |
| Juan Martin del Potro | 122 | 13 |
| **2020** | Dusan Lajovic | 26 | 30 |
| Benjamin Bonzi | 165 | 4 |
| Gianluca Mager | 100 | 11 |
| **Агрегирано** | Kei Nishikori | 41 | 110.5 |
| Julius Tverijonas | 830 | 0.5 |
| Pablo Carreno Busta | 16 | 120.5 |

Табела 10 Корелација ранга и тежинског степена чвора

1. Да ли у мрежи постоје хабови и који су?

Хабови су чланови мреже са највећим утицајем у мрежи. Утицај се рачуна на основу улазне централности по степену. Најчешће се одређују за усмерене мреже, па тако и све функције за рачунање утицаја чвора као улаз очекују усмерен граф. Како су креирани графови мрежа неусмерени, ова функција би била непрецизна па зато није посматрана.

1. Да ли мрежа исказује особине малог света?

Како ниједна од мрежа није повезана, не исказују особине малог света. Уместо за целу мрежу, рађена је провера да ли највећа повезана компонента исказује особине малог света.

Како би мрежа имала особине малог света, неопходно је да има ниске вредности просечне удаљености чворова у односу на величину мреже, да има низак дијаметар и висок степен кластеризације.

Као што смо претходно видели, просечна дистанца за све мреже је реда величине 3, а дијаметар је реда величине 10, што су изразито мале вредности у односу на број тенисера у мрежи. Просечан коефицијент кластеризације по годинама редом има вредности 2018. - 0.13998203859759706, 2019. - 0.1467226467630041, 2020. - 0.10923147632347448 и агрегирано - 0.19829832679790357. Вредности коефицијента кластеризације су релативно ниске, па се тако може закључити да мреже имају неке карактеристике малог света.

1. Какве су карактеристике его мрежа чланова Велике тројке? У којој мери се те карактеристике разликују?

У наредној табели су приказане карактеристике его мрежа Велике тројке по годинама и агрегирано. У 2018. години, примећује се мања его мрежа Рафаела Надала у односу на остале. Надал има мању мрежу јер је 2018. године, због више различитих повреда играо само на девет турнира. У 2019. све три мреже имају јако сличне карактеристике. У 2020. години се примећује знатно мања мрежа Роџера Федерера. Узрок за ово је свакако епидемија корона вируса, али и чињеница да је Федерер играо само на једном турниру, а остатак сезоне пропустио, што због короне, што због последица повреде колена. У агрегираној мрежи се види да сва тројица тенисера имају јако сличне его мреже.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Име и презиме** | **Број чворова** | **Број грана** | **Просечан степен чвора** |
| **2018** | Новак Ђоковић | 43 | 324 | 15.0698 |
| Роџер Федерер | 45 | 355 | 15.7778 |
| Рафаел Надал | 36 | 229 | 12.7222 |
| **2019** | Новак Ђоковић | 48 | 378 | 15.7500 |
| Роџер Федерер | 47 | 381 | 16.2128 |
| Рафаел Надал | 49 | 377 | 15.3878 |
| **2020** | Новак Ђоковић | 33 | 117 | 34.0482 |
| Роџер Федерер | 7 | 11 | 3.1429 |
| Рафаел Надал | 30 | 104 | 6.9333 |
| **Агрегирано** | Новак Ђоковић | 81 | 1314 | 32.4444 |
| Роџер Федерер | 75 | 1389 | 37.0400 |
| Рафаел Надал | 83 | 1413 | 34.0482 |

Табела 11 Карактеристике его мрежа велике тројке, по годинама

1. Каква је позиција его чвора у свакој од его мрежа? Како су они структурно уграђени у мрежу?

На наредним сликама су приказане его мреже тенисера, по годинама, од 2018. до 2020.. Ове его мреже су генерисане помоћу алата *Gephi*. Агрегиране мреже су изостављена јер њихов приказ не би био јасан. Оне се могу наћи у приложеним фотографијама.

Text

Description automatically generated

Слика 3 Его мрежа Новак Ђоковић 2018.

Text

Description automatically generated with low confidence

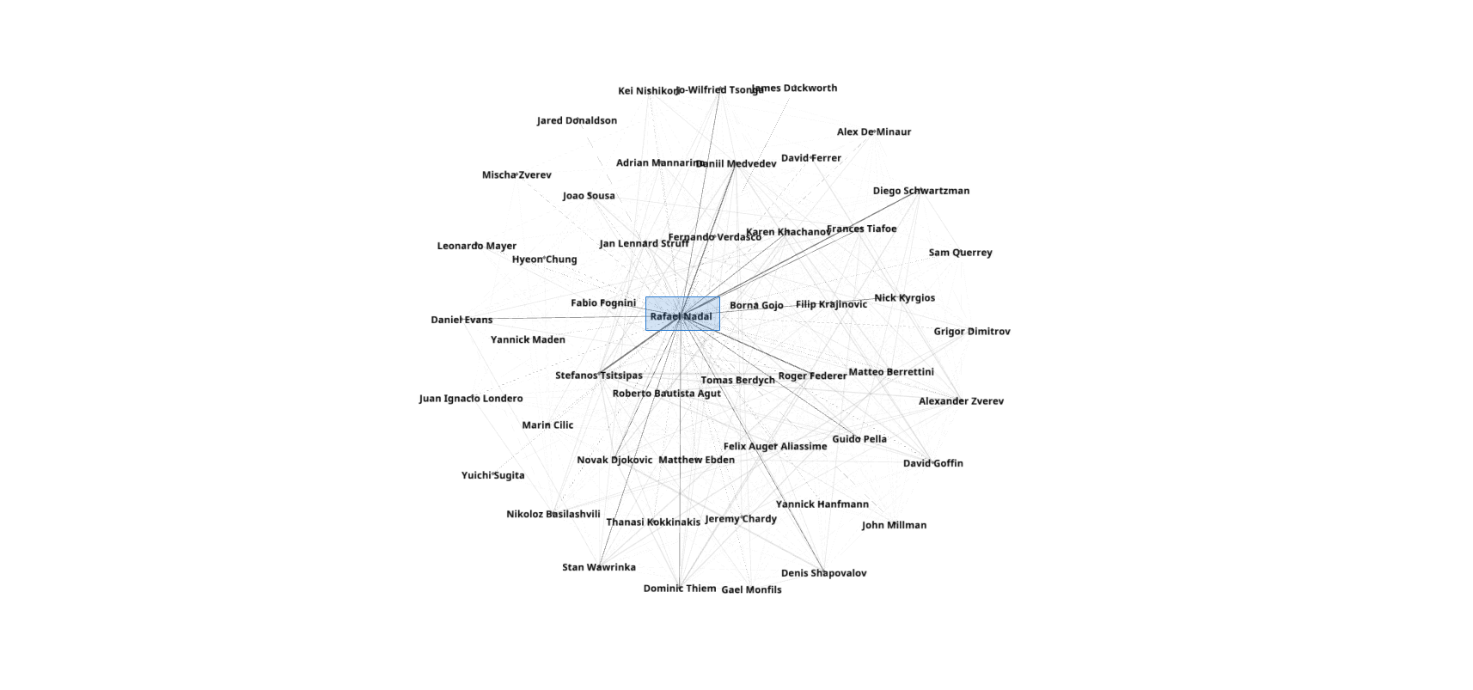
Слика 4 Его мрежа Новак Ђоковић 2019.



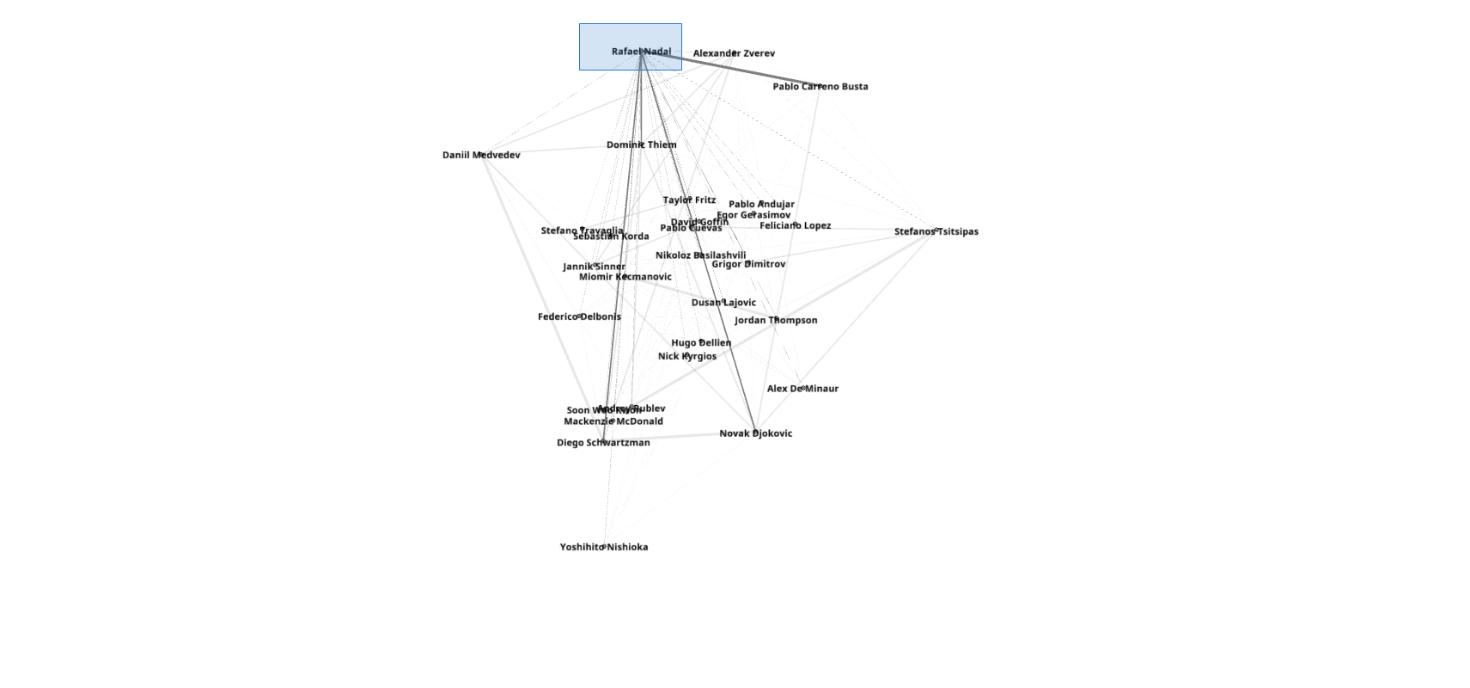
Слика 5 Его мрежа Новак Ђоковић 2020.



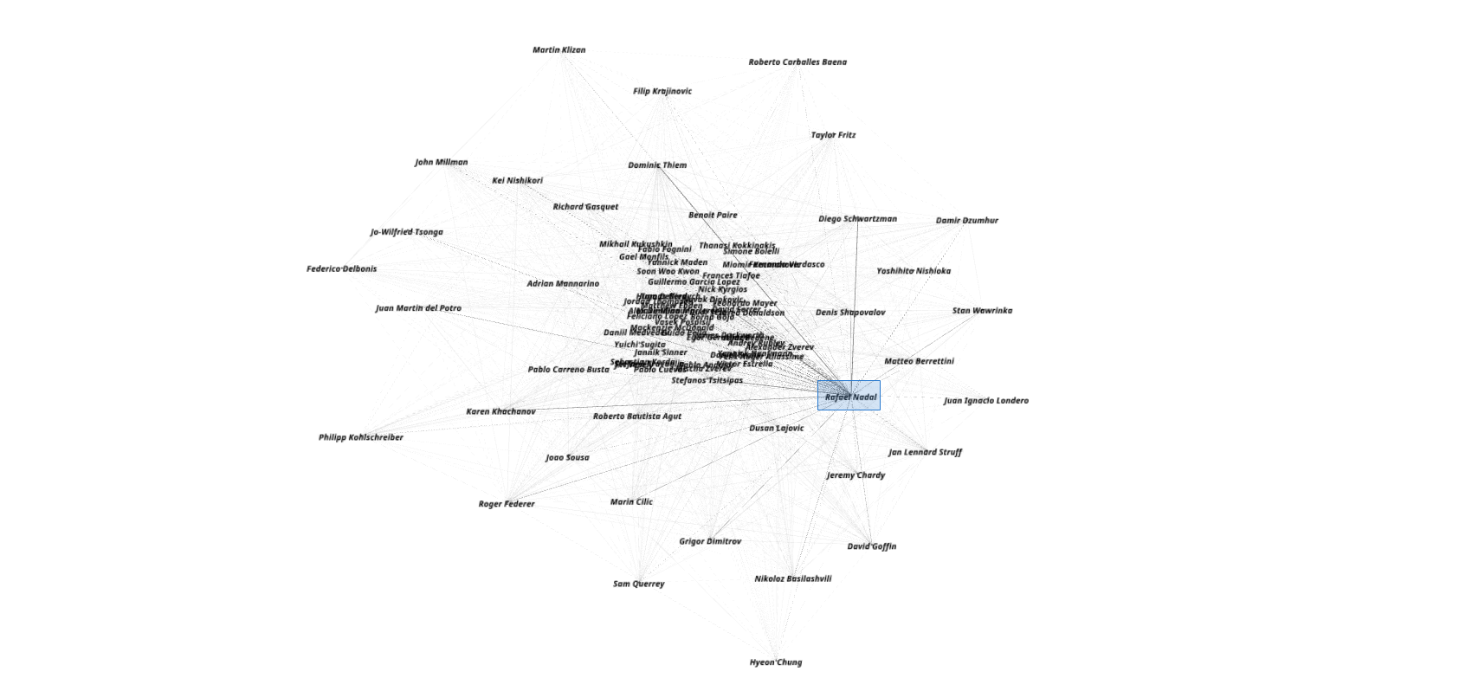
Слика 6 Его мрежа Рафаел Надал 2018.



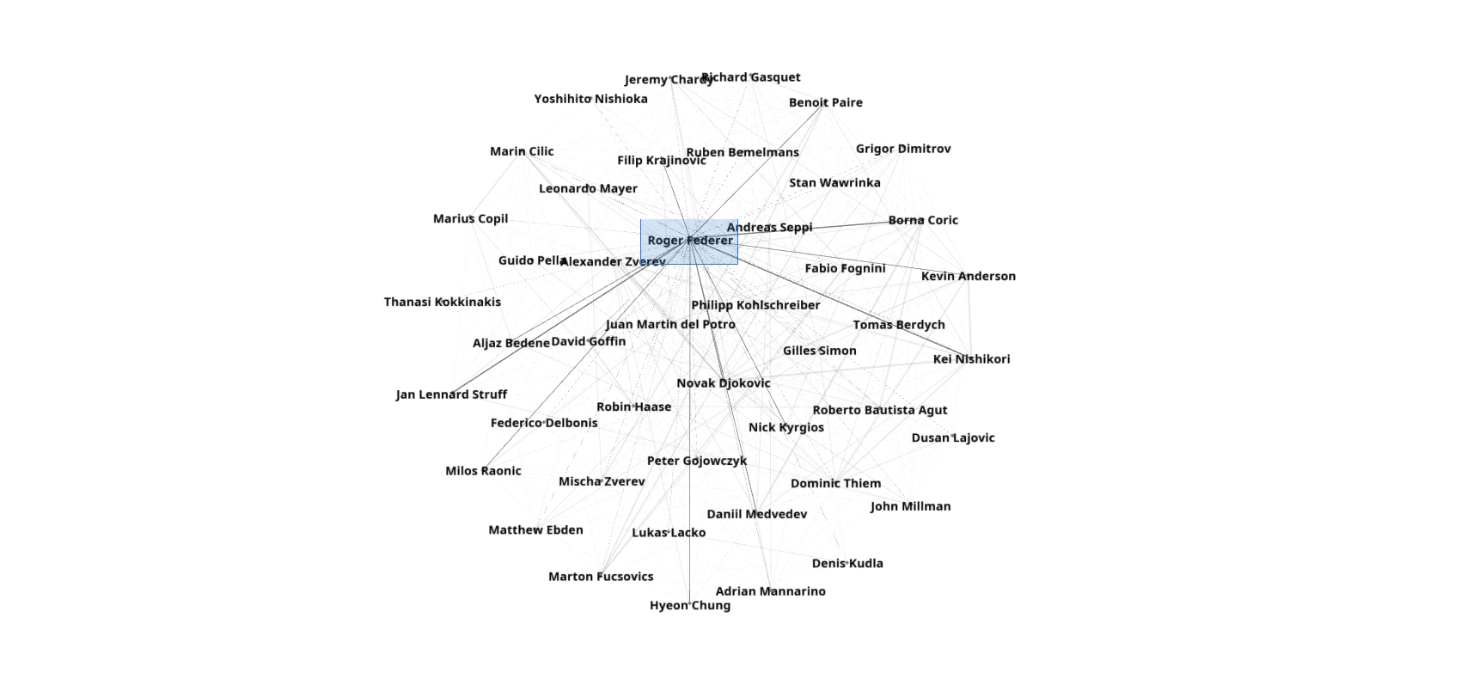
Слика 7 Его мрежа Рафаел Надал 2019.



Слика 8 Его мрежа Рафаел Надал 2019.



Слика 9 Его мрежа Рафаел Надал 2020.



Слика 10 Его мрежа Роџер Федерер 2018.



Слика 11 Его мрежа Роџер Федерер 2019.



Слика 12 Его мрежа Роџер Федерер 2020.

1. Како су посматране его мреже уграђене у мрежу тенисера?

Чворови (чланови „Велике тројке“) представљају централније чворове у мрежи и у зависности од године, њихове активности и одиграних мечева и изглед одговарајуће его мреже је различит (дато у приказима его мрежа по годинама).

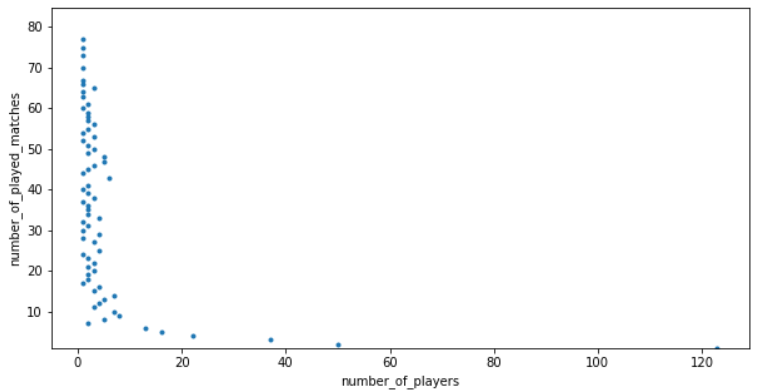
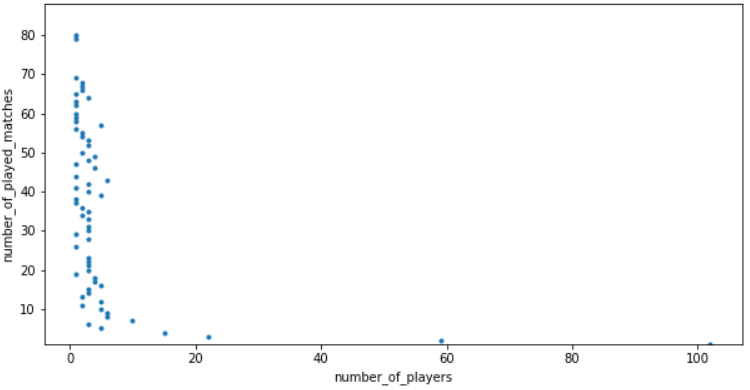
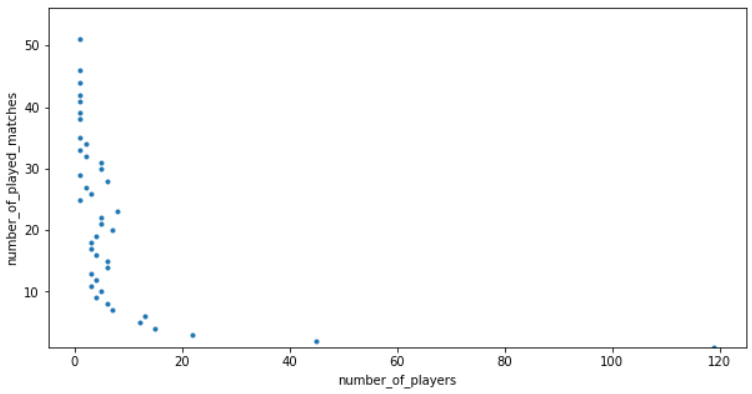
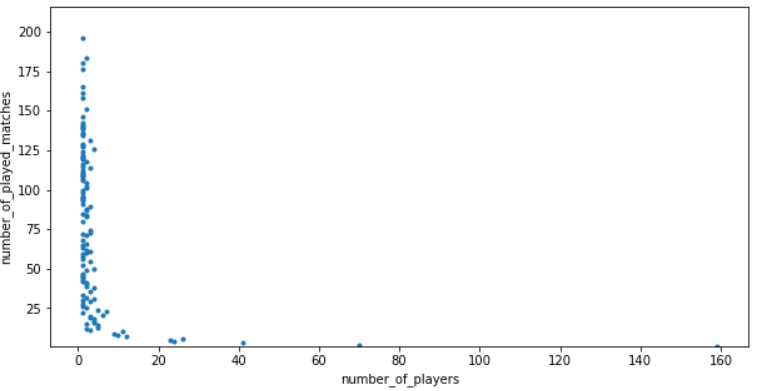
Додатак одговору у наредном питању.

1. Анализирати мрежу добијену унификацијом его мрежа чланова Велике тројке. Који проценат чворова мреже свих тенисера учествује у њој? Спровести кластерисање овако добијене мреже на три кластера и на основу добијених резултата дати интерпретацију критеријума припадности кластерима.

Број чворова унификованих мрежа и њихов удео у укупном броју чворова је 2018 – 81 19.331742243436754%, 2019. - 88 24.175824175824175%, 2020 – 57 16.52173913043478%, агрегиран – 136 23.407917383820998%.

1. Каква је дистрибуција броја тенисера у односу на број мечева који су одиграли?

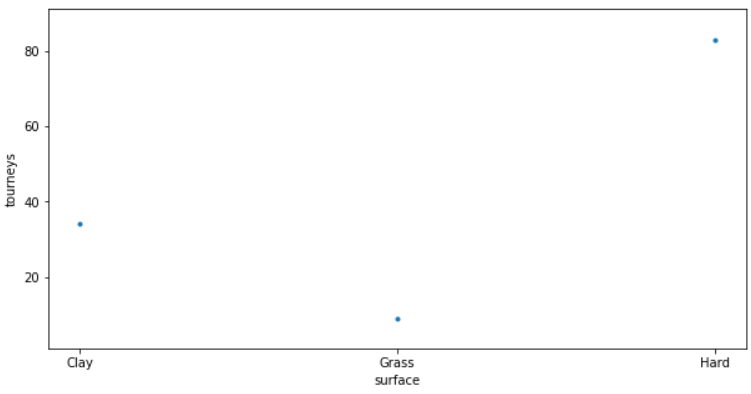
На наредним сликама је приказана корелација броја тенисера у односу на број мечева који су одиграли. Примећује се да је мали број тенисера играо велики број мечева, а већина играла мали број мечева. То је највероватније узроковано тиме што боље рангирани играчи играју велики број мечева на турнирима, док већи број слабије рангираних испада већ у квалификацијама или првим колима.

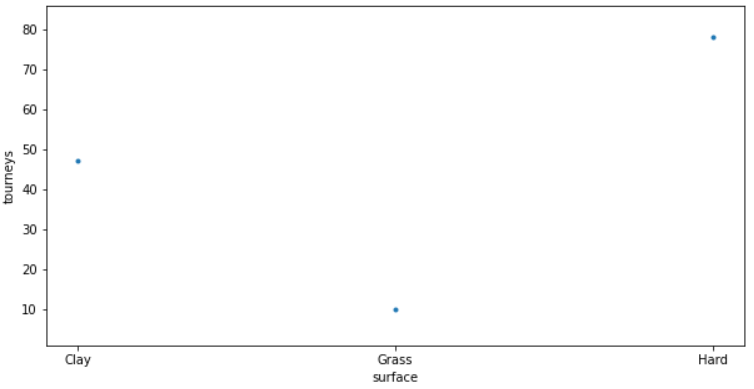
Слика 13 Дистрибуција броја тенисера у односу на број мечева 2018., 2019., 2020., агрегирано слева на десно

1. Каква је дистрибуција броја турнира у односу подлогу и годину одржавања?

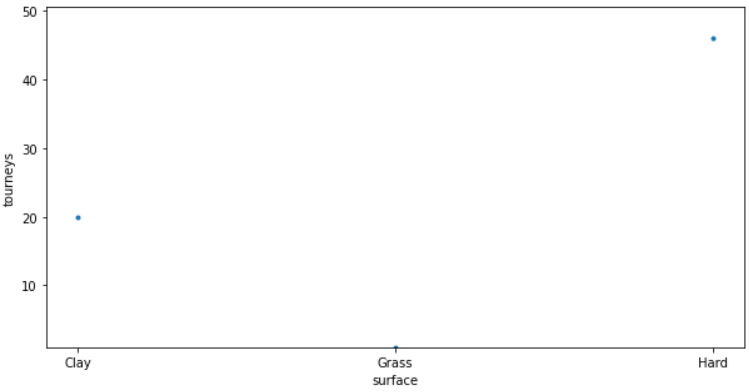
На наредним графицима (слика 4 - 7) је приказана дистрибуција броја турнира у односу на подлогу. Примећује се да се на тврдој подлози игра највећи број мечева, што осликава реалну ситуацију, како се на тврдој подлози играју два од четири гренд слем турнира као и турнири који им претоде у истом делу године. Интересантно је и приметити да је у 2020. години одиграно само два меча на травнатој подлози, што је узроковано тадашњон епдемиолошком ситуацијом због које се нису одржавали спортски догађаји.



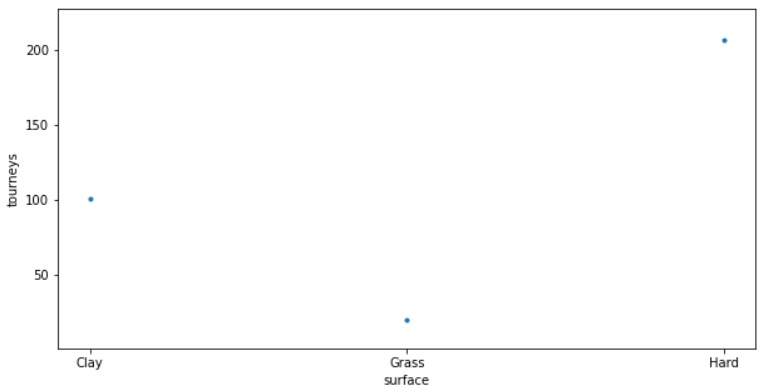
Слика 14 Дистрибуција чворова по броју турнира у односу на подлогу за 2018. годину



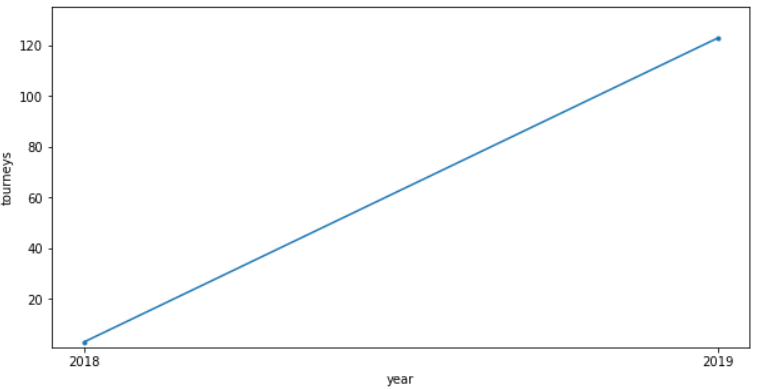
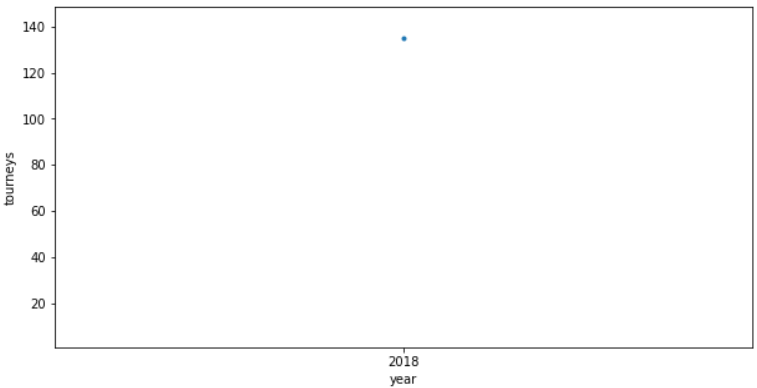
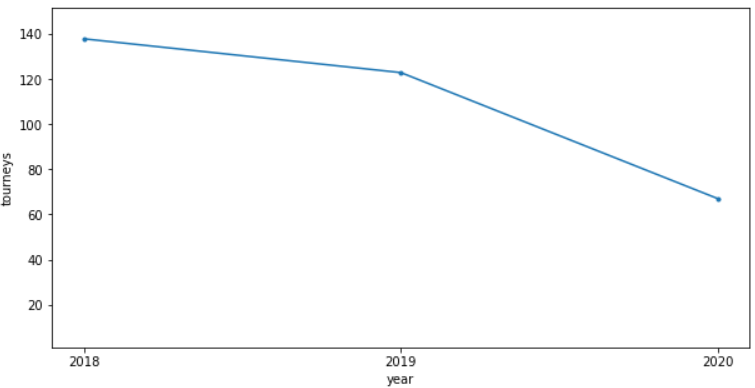
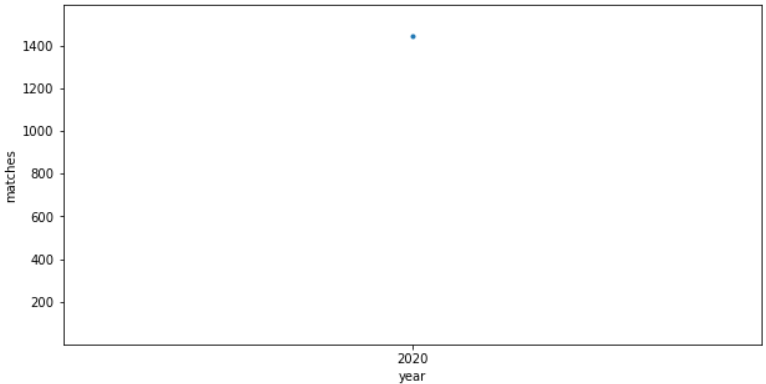
Слика 15 Дистрибуција чворова по броју турнира у односу на подлогу за 2019. годину



Слика 16 Дистрибуција чворова по броју турнира у односу на подлогу за 2019. годину



Слика 17 Дистрибуција чворова по броју турнира у односу на подлогу за агрегиране податке

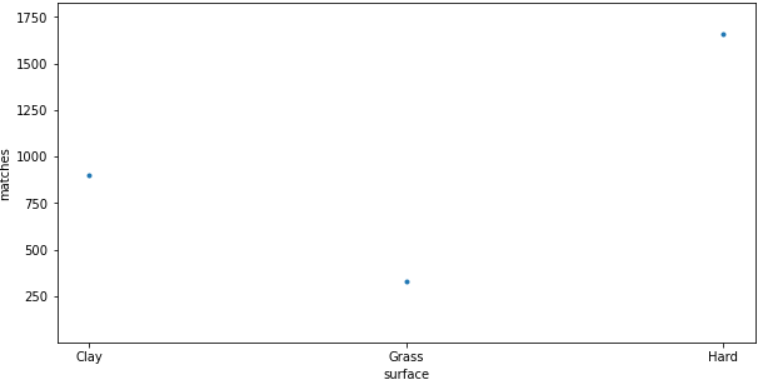
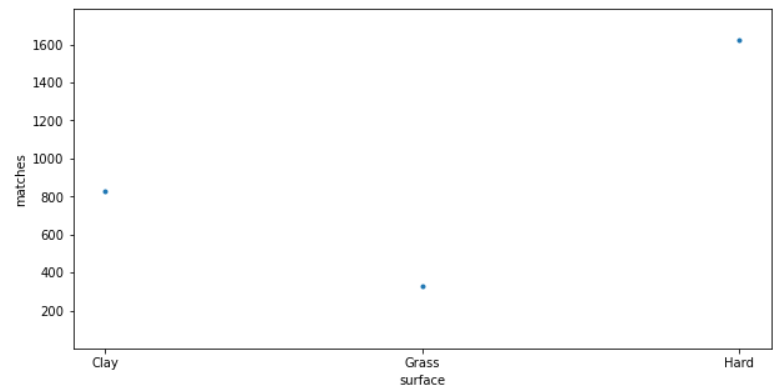
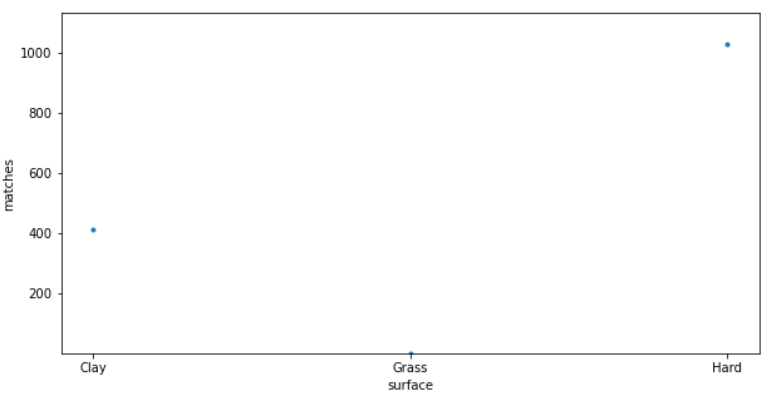
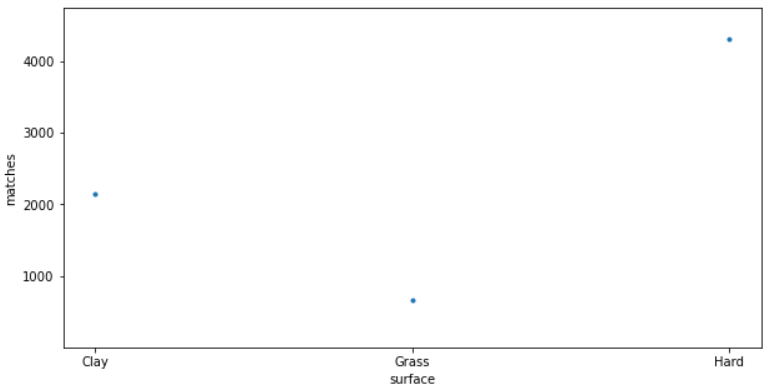
 

Слика 18 Дистрибуције чворова по броју турнира у односу годину одржавања, приказани редом с лева на десно графици за 2018., 2019., 2020. и агрегирано

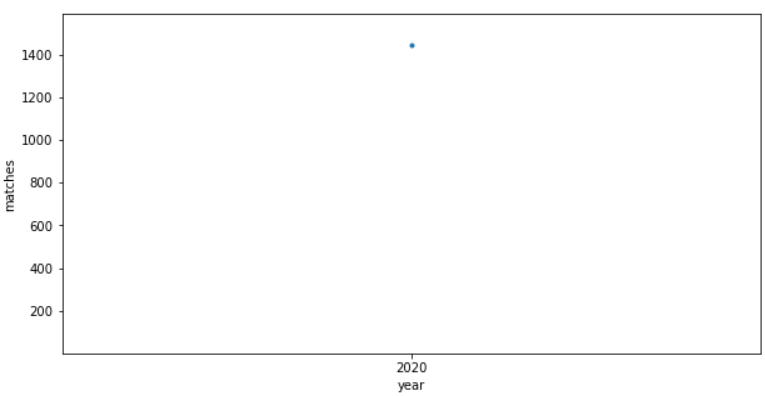
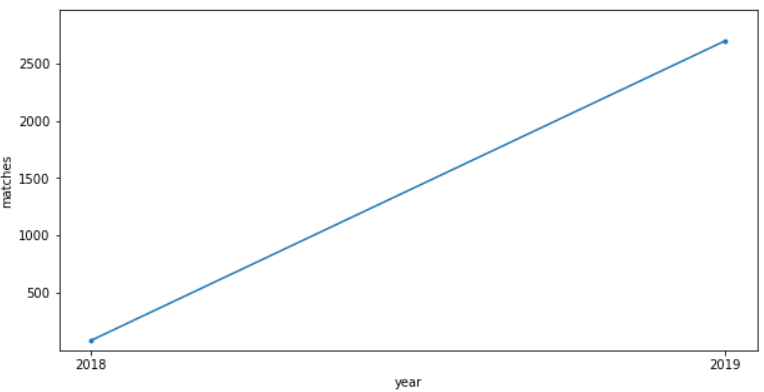
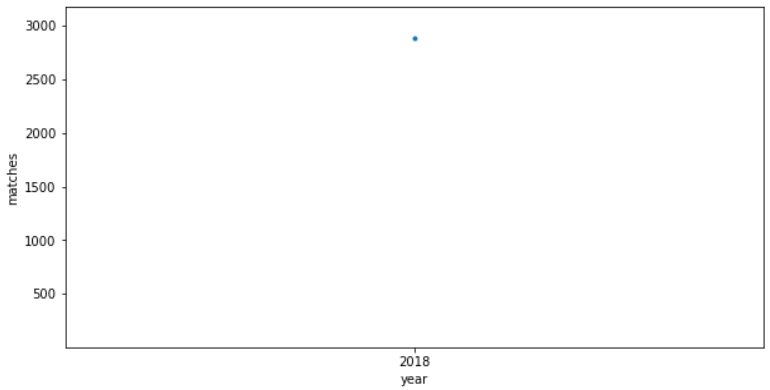
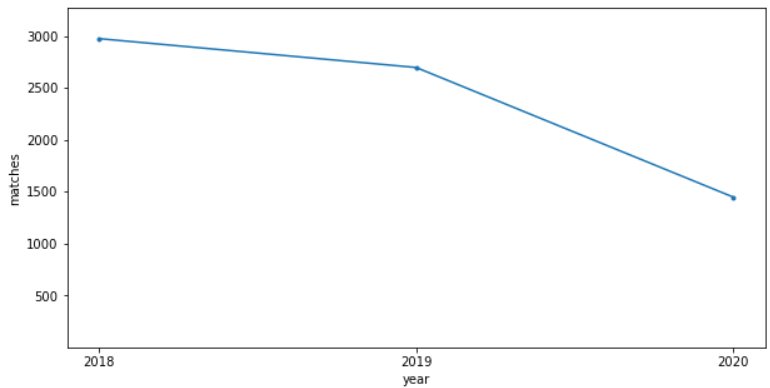
На претходним графицима је приказана дистрибуција чворова по броју турнира у односу на годину одржавања. Види се да су за 2018. и 2020. годину сви мечеви одржани у тим годинама. За 2019. годину види се да је део мечева одржан у 2018.. То се десило због тога што је тениска сезона трајала од 31. децембра 2018. до 24. новембра 2019. и три турнира су стартовала 31. децембра 2018., па због тога график овако изгледа. Такође, на графику за агрегиране податке се види мањи број мечева у 2020. години него у осталим годинама.

1. Каква је дистрибуција броја мечева у односу подлогу и годину одржавања?

На наредним сликама 9 и 10 су приказани графици дистрибуције броја мечева у односу на подлогу и годину одржавања, респективно. Примећује се да графици изгледају јако слично као графици приказани за дистрибуцију у односу на број турнира, с тим да је број одиграних мечева знатно већи од броја одиграних турнира.

Слика 19 Дистрибуције чворова по броју мечева у односу на подлогу, приказани редом с лева на десно графици за 2018., 2019., 2020. и агрегирано

Слика 20 Дистрибуције чворова по броју мечева у односу на годину одржавања, приказани редом с лева на десно графици за 2018., 2019., 2020. и агрегирано

Списак скраћеница

CSV – *Comma Separated Values*

TCP - *Transmission Control Protocol*

Списак слика

[Слика 1 Код за израчунавање језгра мреже 13](#_Toc81442098)

[Слика 2 Дистрибуција чворова по степену за 2018., 2019., 2020. и агрегирано слева на десно 15](#_Toc81442099)

[Слика 3 Его мрежа Новак Ђоковић 2018. 18](#_Toc81442100)

[Слика 4 Его мрежа Новак Ђоковић 2019. 19](#_Toc81442101)

[Слика 5 Его мрежа Новак Ђоковић 2020. 19](#_Toc81442102)

[Слика 6 Его мрежа Рафаел Надал 2018. 20](#_Toc81442103)

[Слика 7 Его мрежа Рафаел Надал 2019. 20](#_Toc81442104)

[Слика 8 Его мрежа Рафаел Надал 2019. 21](#_Toc81442105)

[Слика 9 Его мрежа Рафаел Надал 2020. 22](#_Toc81442106)

[Слика 10 Его мрежа Роџер Федерер 2018. 22](#_Toc81442107)

[Слика 11 Его мрежа Роџер Федерер 2019. 22](#_Toc81442108)

[Слика 12 Его мрежа Роџер Федерер 2020. 23](#_Toc81442109)

[Слика 13 Дистрибуција броја тенисера у односу на број мечева 2018., 2019., 2020., агрегирано слева на десно 24](#_Toc81442110)

[Слика 14 Дистрибуција чворова по броју турнира у односу на подлогу за 2018. годину 25](#_Toc81442111)

[Слика 15 Дистрибуција чворова по броју турнира у односу на подлогу за 2019. годину 25](#_Toc81442112)

[Слика 16 Дистрибуција чворова по броју турнира у односу на подлогу за 2019. годину 25](#_Toc81442113)

[Слика 17 Дистрибуција чворова по броју турнира у односу на подлогу за агрегиране податке 26](#_Toc81442114)

[Слика 18 Дистрибуције чворова по броју турнира у односу годину одржавања, приказани редом с лева на десно графици за 2018., 2019., 2020. и агрегирано 26](#_Toc81442115)

[Слика 19 Дистрибуције чворова по броју мечева у односу на подлогу, приказани редом с лева на десно графици за 2018., 2019., 2020. и агрегирано 27](#_Toc81442116)

[Слика 20 Дистрибуције чворова по броју мечева у односу на годину одржавања, приказани редом с лева на десно графици за 2018., 2019., 2020. и агрегирано 28](#_Toc81442117)

Списак табела

[Табела 1 Тенисери који имају највише сусрета са другим тенисерима 6](#_Toc81441642)

[Табела 2 Тенисери који су учесвовали на највећем броју турнира 7](#_Toc81441643)

[Табела 3 Тенисери који су се сусретали са највећим бројем тенисера 8](#_Toc81441644)

[Табела 4 Државе из којих долази највећи број играча 9](#_Toc81441645)

[Табела 5 Првих осам тенисера на ранг листи по годинама 9](#_Toc81441646)

[Табела 6 Првих осам држава по броју тенисера у првих 30 по годинама 9](#_Toc81441647)

[Табела 7 Вредности параметара за анализу кластерисања и асотативности 12](#_Toc81441648)

[Табела 8 Максималне вредности централности тенисера по годинама 14](#_Toc81441649)

[Табела 9 Вредности дистанце и дијаметра по годинама 15](#_Toc81441650)

[Табела 10 Корелација ранга и тежинског степена чвора 16](#_Toc81441651)

[Табела 11 Карактеристике его мрежа велике тројке, по годинама 18](#_Toc81441652)