

УНИВЕРЗИТЕТ У БЕОГРАДУ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ



АНАЛИЗА СОЦИЈАЛНИХ МРЕЖА

Пројектни задатак

Професори:

др Јелица Протић

доц. др Марко Мишић

Студенти:

Бодин Матејић 2019/3145

Исидора Смиљанић 2018/3218

Београд, август 2020.

Садржај

1. УВОД	3
2. ЦИЉ.....	3
3. ИСТРАЖИВАЧКИ ПРОБЛЕМ	3
3.1. УПОРЕДНА АНАЛИЗА НАУЧНЕ ПРОДУКЦИЈЕ ИЗ ОБЛАСТИ РАЧУНАРСТВА НА УНИВЕРЗИТЕТУ У БЕОГРАДУ	3
3.2. СКУП ПОДАТАКА ЗА АНАЛИЗУ	4
3.3. МОДЕЛОВАЊЕ МРЕЖЕ	5
4. ПОСТУПАК И РЕЗУЛТАТИ.....	6
5. ИСТРАЖИВАЧКА ПИТАЊА И ОДГОВОРИ.....	7
5.1. Који је просечан број коаутора по сваком аутору?	7
5.2. Ко су најпродуктивнији научници из области рачунарства и ком факултету и катедри припадају?... 7	
5.3. Које катедре и факултети су најпродуктивнији посматрајући засебно научну продукцију у часописима и на конференцијама?	8
5.4. Које заједнице или научно истраживачке групе се могу уочити приликом анализе мреже?	9
5.5. У којој мери аутори имају тенденцију да пишу публикације са истим коауторима?	10
5.6. Који аутори представљају центре окупљања у оквиру својих катедри и факултета?	11
5.7. Који аутори представљају језгро мреже?	11
5.8. Да ли аутори са различитих факултета и катедри међусобно сарађују и у којој мери?	11
5.9. Каква је сарадња међу ауторима са истог факултета?	11
5.10. Ко су аутори који повезују различите групе у оквиру мреже?	11
5.11. Колика је густина сваке од моделованих мрежа?	12
5.12. У којој мери су мреже повезане и централизоване?	12
5.13. Које су просечне дистанце у оквиру моделованих мрежа?	12
5.14. Каква је дистрибуција чворова по степену и да ли прати неку закономерност?	13
5.15. Да ли мрежа коаутора исказује особине малог света?	13
5.16. Да ли постоје разлике између факултета и катедри у смислу обима и учестаности публикација у часописима и на конференцијама?	13
5.17. Да ли постоји разлика у просечном броју аутора по радовима у часописима и на конференцијама?	14
5.18. У којим часописима и на којим конференцијама се у просеку највише објављује?	14
5.19. Да ли се издвајају и које су групе (заједнице) часописа у оквиру мреже часописа? Да ли аутори имају тенденцију да објављују у часописима сличног профила?	15
5.20. У којим годинама су аутори били најпродуктивнији по факултетима и катедрама?	16
6. ЗАКЉУЧАК.....	17

1. Увод

У оквиру овог документа дат је извештај о изради пројектног задатка на предмету Анализа социјалних мрежа (13M111АСМ) у школској 2019/2020. години. Овај документ садржи резултате спроведене анализе и писана објашњења уочених феномена.

Обрада примарног скупа података и анализа моделиране социјалне мреже је рађена у програмском језику Python и помоћу MS Excel алата, а визуализација је одрађена коришћењем алата Gephi.

2. Циљ

Циљ пројектног задатка на предмету Анализа социјалних мрежа је практична примена стеченог знања из предмета на примеру једног конкретног истраживачког проблема. Кроз задати истраживачки проблем, извршено је прикупљање, обрада и прелиминарна анализа (сировога) скупа података, издвојени су неопходни подаци и моделиран је проблем мрежом одговарајућег типа. Моделирана мрежа је анализирана помоћу алата за обраду социјалних мрежа (библиотека NetworkX, Gephi) и извршена је визуализација мреже. Добијени резултати изложени су у форми одговора на постављена истраживачка питања.

3. Истраживачки проблем

У оквиру ове секције дата је поставка истраживачког задатка и објашњење неких корака приликом рађења анализе, као и ближе објашњење коришћеног скупа података.

3.1. Упоредна анализа научне продукције из области рачунарства на Универзитету у Београду

Тема пројектног задатка је квантитативна и квалитативна анализа научне продукције из области рачунарства на Универзитету у Београду (УБ) на три најважнија факултета у његовом саставу који се баве рачунарством, тачније научних радова запослених на катедрама факултета УБ којима је рачунарство примарна област.

Циљ анализе је да се кроз одговарајуће библиометријске и наукометријске анализе, као и кроз анализу колаборационе, социјалне мреже утврди стање рачунарских наука на појединачним факултетима, као и ниво сарадње између запослених са истих и различитих факултета који се баве облашћу рачунарства

У оквиру Универзитета у Београду (УБ) постоје три факултеа који су матични за једну од четири области рачунарства. То су Математички факултет (МАТФ) за област рачунарских наука (computer science), Електротехнички факултет (ЕТФ) за област рачунарске технике (computer engineering) и софтверског инжењерства (software engineering) и Факултет организационих наука (ФОН) за област информационих система (information systems). У оквиру наведених факултета постоје катедре које се баве поменутим научним областима. То су Катедра за софтверско инжењерство (ФОН_СИ), Катедра за информационе технологије (ФОН_ИТ) и Катедра за информационе системе (ФОН_ИС) са Факултета организационих наука, Катедра за рачунарство и информатику (МАТФ_РТИ) са Математичког факултета и Катедра за рачунарску технику и информатику (ЕТФ_РТИ) са Електротехничког факултета.

Поред обавеза у настави, запослени на универзитету имају обавезу писања одговарајућих научних публикација. Напредовање у академској каријери сваког запосленог у многоме зависи од квалитета и обима научне продукције појединца, али утиче и на видљивост, репутацију и квалитет установа (факултета) којима они припадају, као и самог универзитета.

Највећи део научне продукције се одвија кроз публикације као што су књиге, дисертације, радови у научним часописима и радови на међународним и домаћим научним конференцијама. Међутим, у смислу евалуације, највише се цене радови у научним часописима и радови на научно-стручним конференцијама.

3.2. Скуп података за анализу

У оквиру овог пројектног задатка анализирана је научна продукција из области рачунарства на Универзитету у Београду која се односи на радове у часописима и радове на научно-стручним конференцијама. Подаци за анализу (примарни скуп података) дати су у виду одговарајућих Excel табела.

За прикупљање података о научним публикацијама су коришћени подаци о ауторима који су се водили као запослени у наставним звањима или били ангажовани као лабораторијско особље на претходно поменутим катедрама у августу 2018. године, што је прикупљено са интернет презентација ових факултета. Подаци су доступни у оквиру Excel датотеке *UB_cs_authors.xlsx*. Остали аутори научних публикација који се јављају у оквиру скупа података нису од интереса за анализу.

Примарни скуп података за анализу (енг. *primary dataset*) је добијен прикупљањем из индексне базе научних радова *Scopus* закључно са августом 2018. године. У скупу података се налази 1290 радова прикупљених на основу доступних података о запосленима на раније поменути три факултета УБ. Подаци су одређеној мери пречишћени, али одређени подаци могу бити непотпуни или у неодговарајућем облику.

Подаци се састоје од следећих колона:

- УБ запослени – име запосленог коме је придружена одговарајућа публикација у *Scopus* бази
- Наслов – наслов публикације
- Година – година издавања
- Аутори – аутори рада
- Број цитирања – број цитата других аутора према *Scopus* бази
- Тип рада – тип публикације
- Cite factor – *CiteScore* библиометријски фактор
- SJIR – *SCImago Journal Rank* библиометријски фактор
- SNIP – *Source Normalized Impact per Paper* библиометријски фактор
- Име документа – име документа (часописа или зборника радова) у коме је публикација објављена

На основу примарног скупа података формиран је секундарни скуп података (енг. *secondary dataset*) који представља пречишћену верзију података за анализу. Пречишћавање је извршено према потребама задатка и циљевима истраживања. У анализи су од интереса само публикације типа *Article*, *Conference Paper*, *Article in Press*, *Book Chapter*, док су занемарене. Приликом пречишћавања су изостављени сви непотребни подаци.

3.3. Моделовање мреже

Секундарни скуп података је искоришћен за моделовање одговарајућих социјалних мрежа. Моделоване су мреже коаутора (на нивоу катедри, факултета и целог универзитета) као и мрежа часописа. Имплементирани мреже су неусмереног и тежинског типа што је у складу са постављеним истраживачким питањима и циљевима. Примарна мрежа за анализу је мрежа коаутора на нивоу универзитета, а преостале су искоришћене у функцији одговарања на одређена истраживачка питања.

У оквиру мреже коаутора, запослени на факултетима УБ који се баве рачунарством представљају чворове мреже, а веза између чворова је успостављена уколико су два запослена написала барем један научни рад заједно.

У оквиру мреже часописа, часопис представља чвор мреже, а веза између њих је успостављена уколико постоји бар један аутор који је објављивао у оба часописа. Зборници радова са конференција нису узимани у обзир приликом моделовања ове мреже.

4. Поступак и резултати

Примарни скуп података – листа релевантних аутора запослених на УБ (*UB_cs_authors.xlsx*) и база научних радова прикупљена из индексне базе *Scopus* (*UB_cs_papers_scopus.xlsx*) је пречишћен коришћењем Python скрипте *cleaner.py*. Ова скрипта уклања дубликате и непотпуне улазе у примарном скупу, исправља погрешне дијакритичке знаке (кукице на словима) и парсира ауторе научних радова - препознаје различите формате њихових имена, издваја само ауторе од интереса за анализу и исписује њихова пуна имена. Пречишћавањем се формира секундарни скуп података (*UB_cs_papers_cleaned.xlsx*) који садржи само неопходне податке из примарног скупа.

Графови су креирани скриптама *graph_authors.py* и *graph_journals.py* из секундарног скупа података коришћењем Python библиотеке NetworkX.

Анализа мрежа одрађена је визуелизацијом графова у програму *Gephi* као и коришћењем скрипте *network_analysis.py* која анализирањем секундарног скупа података и генерисаних графова креира Excel фајлове (*analysis_authors.xlsx*, *analysis_modules.xlsx*, *analysis_journals.xlsx*) у којима су сумиране различите потребне информације и резултати анализе.

Детаљни извештаји о централностима, повезаности и степенима графова извезени из *Gephi*-ја као и сами *.gephi* фајлови су дати унутар фолдера *graphs* појединачно за сваки од анализираних графова.

5. Истраживачка питања и одговори

Приликом обраде примарног и секундарног скупа података као смернице су коришћена претходно дефинисана истраживачка питања. У оквиру ове секције дати су одговори на та питања.

5.1 Који је просечан број коаутора по сваком аутору?

Просечан број коаутора по аутору је суштински просечан тежински степен мрежних графова:

- за Универзитет у Београду износи: 12,38983051.
- за Електротехнички факултет износи: 16,05.
- за Математички факултет износи: 3,076923077.
- за Факултет организационих наука износи: 16,51282051.

За више информација погледати фајл *analysis_authors.xlsx*.

5.2. Ко су најпродуктивнији научници из области рачунарства и ком факултету и катедри припадају?

Као што се види из табеле 5.2.1 најпродуктивнији научници су Владан Девеџић са укупно 137 радова и Јелена Јовановић са укупно 106 радова, обоје са катедре за софтверско инжењерство ФОН-а. Након њих иде Сања Вранеш са 67 радова са катедре за рачунарску технику и информатику ЕТФ-а.

Autor	Modul	Broj radova	Broj koautora	Prosečan broj koautora po radu
Vladan Devedzic	FON_SI	137	61	0.45
Jelena Jovanovic	FON_SI	106	53	0.50
Sanja Vranes	ETF_RTI	67	4	0.06
Dusan Starcevic	FON_IT	57	55	0.96
Miroslav Minovic	FON_IT	46	64	1.39
Milo Tomasevic	ETF_RTI	45	50	1.11
Veljko Milutinovic	ETF_RTI	42	80	1.90
Milos Milovanovic	FON_IT	39	62	1.59
Predrag Janicic	MATF_RTI	37	16	0.43

Табела 5.2.1. Најпродуктивнијих научници из области рачунарства на УБ

За више информација погледати фајл *analysis_authors.xlsx*.

5.3. Које катедре и факултети су најпродуктивнији посматрајући засебно научну продукцију у часописима и на конференцијама?

Најпродуктивнија катедра по броју научних радова и у часописима и на конференцијама је Катедра за рачунарску технику и информатику Електротехничког факултета са 140 научних публикација у часописима и 153 научних публикација на конференцијама. Подаци су приказани у табели 5.3.1.

Katedra	Broj radova	Časopisi	Konferencije
ETF_RTI	311	140	153
FON_SI	247	107	118
MATF_RTI	173	115	55
FON_IT	126	70	46
FON_IS	53	35	17

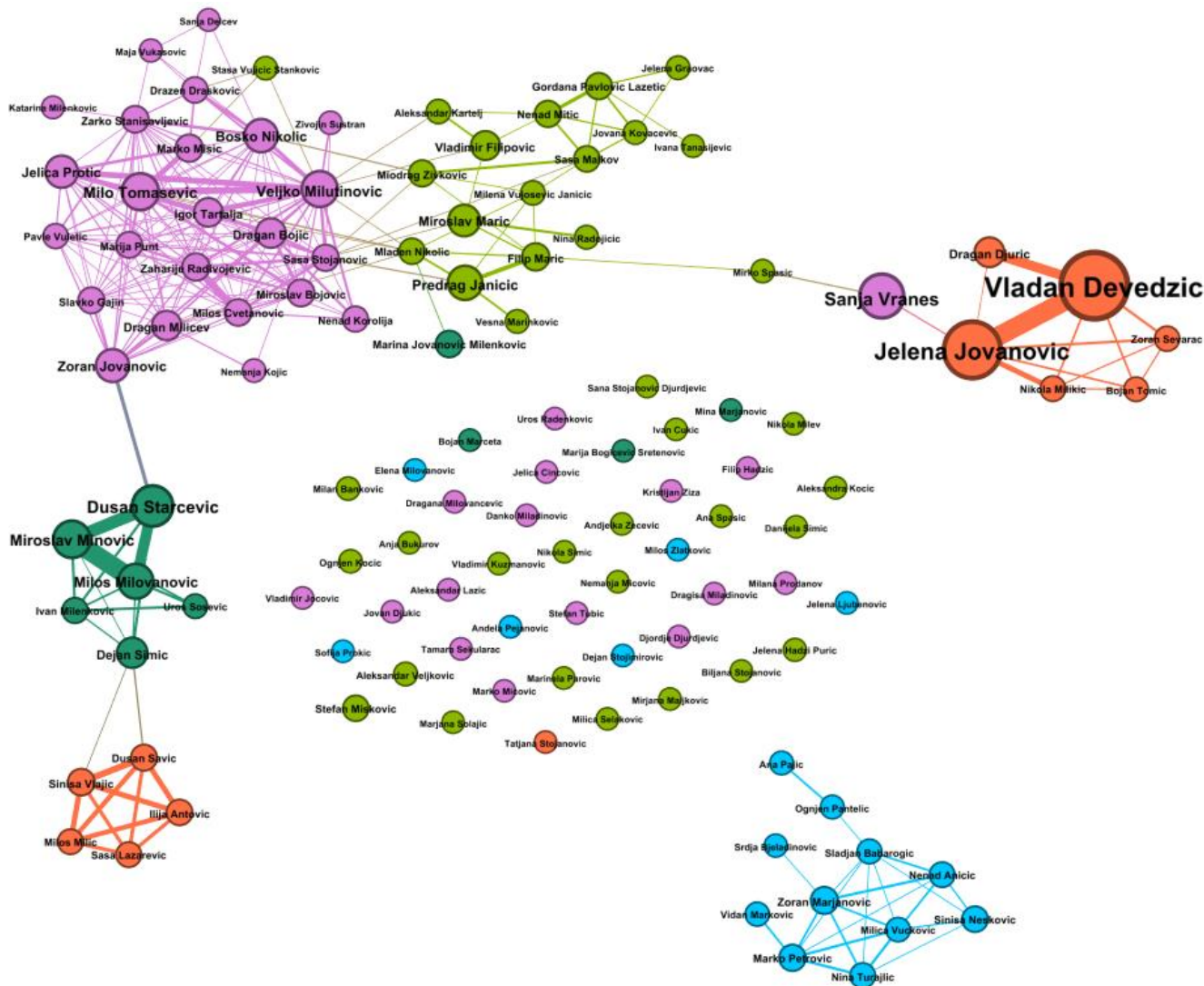
Табела 5.3.1 Најпродуктивније катедре по броју научних радова

Најпродуктивнији факултет, такође у обе категорије, је Факултет организационих наука са 210 научних публикација у часописима и 181 научне публикације на конференцијама. Подаци су приказани у табели 5.3.2.

Fakultet	Broj radova	Časopisi	Konferencije
FON	424	210	181
ETF	311	140	153
MATF	173	115	55

Табела 5.3.2 Најпродуктивнији факултети по броју научних радова

За више информација погледати фајл *analysis_modules.xlsx*.



Слика 5.4.2 Мрежа коаутора обојена по катедрама којима припадају аутори

5.5. У којој мери аутори имају тенденцију да пишу публикације са истим коауторима?

Са слике 5.4.1, приложене у претходном одговору, на основу тежина грана, односно дебљина веза које су сразмерне броју заједничких радова између два научника (чвора) и величине чвора која је сразмерна броју објављених радова, можемо да закључимо да што је научник продуктивнији то има већу тенденцију да пише публикације са истим коауторима, као и да се унутар мреже јављају мање групе научника који често међусобно сарађују.

5.6. Који аутори представљају центре окупљања у оквиру својих катедри и факултета?

Центре окупљања на својим катедрама и факултетима представљају аутори који су највише повезани унутар својих мрежа тј. они који имају највећу централност по блискости, малу ексцентричност и највише коаутора, односно велики степен чвора.

На Електротехничком факултету је то Вељко Милутиновић, на Математичком факултету Филип Марић, а на Факултету организационих наука су то Владан Девеџић, Дејан Симић и Слађан Бабарогић са својих катедри.

5.7. Који аутори представљају језгро мреже?

Са слике 5.4.1 можемо да уочимо да језгро мреже представљају аутори са Електротехничког факултета зато што су добро повезани и густо међусобно и са осталим компонентама.

5.8. Да ли аутори са различитих факултета и катедри међусобно сарађују и у којој мери?

Аутори са различитих факултета и катедри понекад сарађују, али у много мањој мери него унутар факултета, односно катедра.

Сарадња између катедри и факултета се види посматрањем чворова на границама комуна који обично имају велику релациону централност.

5.9. Каква је сарадња међу ауторима са истог факултета?

Аутори најчешће бирају за коаутора научника са истог факултета што се на графу види као велика густина веза у заједницама унутар мреже.

5.10. Ко су аутори који повезују различите групе у оквиру мреже?

Са слике 5.4.2 се могу видети аутори коју повезују различите катедре:

- ЕТФ_РТИ и МАТФ_РТИ – већи број аутора од којих највише веза са другом катедром имају Вељко Милутиновић, Саша Стојановић и Младен Николић.
- ЕТФ_РТИ и ФОН_ИТ – Зоран Јовановић и Душан Старчевић.
- ЕТФ_РТИ и ФОН_СИ – Сања Вранеш и Јелена Јовановић.
- МАТФ_РТИ и ФОН_ИТ – Младен Николић и Марина Јовановић Миленковић.
- ФОН_ИТ и ФОН_СИ – Дејан Симић, Душан Савић и Синиша Влајић

5.11. Колика је густина сваке од моделованих мрежа?

Мрежа коаутора на Универзитету у Београду – *authors_graph.gexf* – 0,037

Мрежа коаутора на Електротехничком факултету – *authors_graph_etf.gexf* – 0,188

Мрежа коаутора на Математичком факултету – *authors_graph_matf.gexf* – 0,036

Мрежа коаутора на Факултету организационих наука – *authors_graph_fon.gexf* – 0,081

Мрежа часописа – *journals_graph.gexf* – 0,1

5.12. У којој мери су мреже повезане и централизоване?

Као што се може видети са слике 5.4.1 мрежа коаутора на Универзитету у Београду се састоји из две главне неповезане компоненте – једну коју чине аутори са катедра ЕТФ_РТИ, МАТФ_РТИ, ФОН_ИТ и ФОН_СИ и другу коју чине аутори са катедре ФОН_ИС.

Поред ове две компоненте постоји одређени број аутора који нису имали коауторства те сви они чине засебне неповезане компоненте.

Детаљни извештаји о централности и повезаности сваког графа могу се наћи у фолдерима *Centrality Report*, *Modularity Report* и *Connected Components Report*, који се налазе унутар њихових појединачних фолдера.

5.13. Које су просечне дистанце у оквиру моделованих мрежа?

Мрежа коаутора на Универзитету у Београду – *authors_graph.gexf* - 3,766

Мрежа коаутора на Електротехничком факултету – *authors_graph_etf.gexf* - 1,504

Мрежа коаутора на Математичком факултету – *authors_graph_matf.gexf* - 2,574

Мрежа коаутора на Факултету организационих наука – *authors_graph_fon.gexf* – 1,8

Мрежа часописа – *journals_graph.gexf* - 2,252

5.14. Каква је дистрибуција чворова по степену и да ли прати неку закономерност?

Просечан степен чвора:

- Мрежа коаутора на Универзитету у Београду – 4,288
- Мрежа коаутора на Електротехничком факултету – 7,35
- Мрежа коаутора на Математичком факултету – 1,385
- Мрежа коаутора на Факултету организационих наука – 3,077
- Мрежа часописа – 24,033

Дистрибуција чворова у великој мери прати power law расподелу - што чвор има више веза пре ће креирати нове везе, односно нови чворови креирају везе према постојећим чворовима са вероватноћом која је пропорционална броју веза које чвор већ има. Ово је у складу са особинама реалних мрежа наспрам насумично генерисаним мрежама.

Детаљнији извештаји о дистрибуцији чворова по степену за сваки граф могу се наћи у фолдеру *Degree Report* који се налази унутар њихових фолдера.

5.15. Да ли мрежа коаутора исказује особине малог света?

Особина малог света је да просечна дужина пута у мрежи расте логаритамски са порастом величине мреже, односно са значајним порастом величине мреже било која два чвора мреже ће и даље бити повезана у релативно мало корака.

Кад се узме у обзир да је вредност дијаметра мреже 11 и да је просечна дистанца у мрежи коаутора 3.766, можемо закључити да ова мрежа исказује особине малог света.

5.16. Да ли постоје разлике између факултета и катедри у смислу обима и учестаности публиковања у часописима и на конференцијама?

Катедра ЕТФ_РТИ има приближно исти број публиковања у часописима и на конференцијама (140 наспрам 153), и таква приближна мала разлика је и на свим осталим катедрама осим на катедри МАТФ_РТИ где је број публикација у часописима 115, а на конференцијама 55.

Што се тиче разлика на нивоу факултета, ЕТФ има 13 више публикација на конференцијама, ФОН 29 више у часописима, док МАТФ има 60 публикација више у часописима.

Тачне вредности су приказане у табелама 5.3.1. и 5.3.2. а више информација можете наћи у фајлу *analysis_modules.xlsx*.

5.17. Да ли постоји разлика у просечном броју аутора по радовима у часописима и на конференцијама?

Просечан број аутора по часопису је 1,479657388, а по конференцији је 1,553524804, по чему можемо да закључимо да не постоји велика разлика.

За више информација погледати фајл *analysis_journals.xlsx*.

5.18. У којим часописима и на којим конференцијама се у просеку највише објављује?

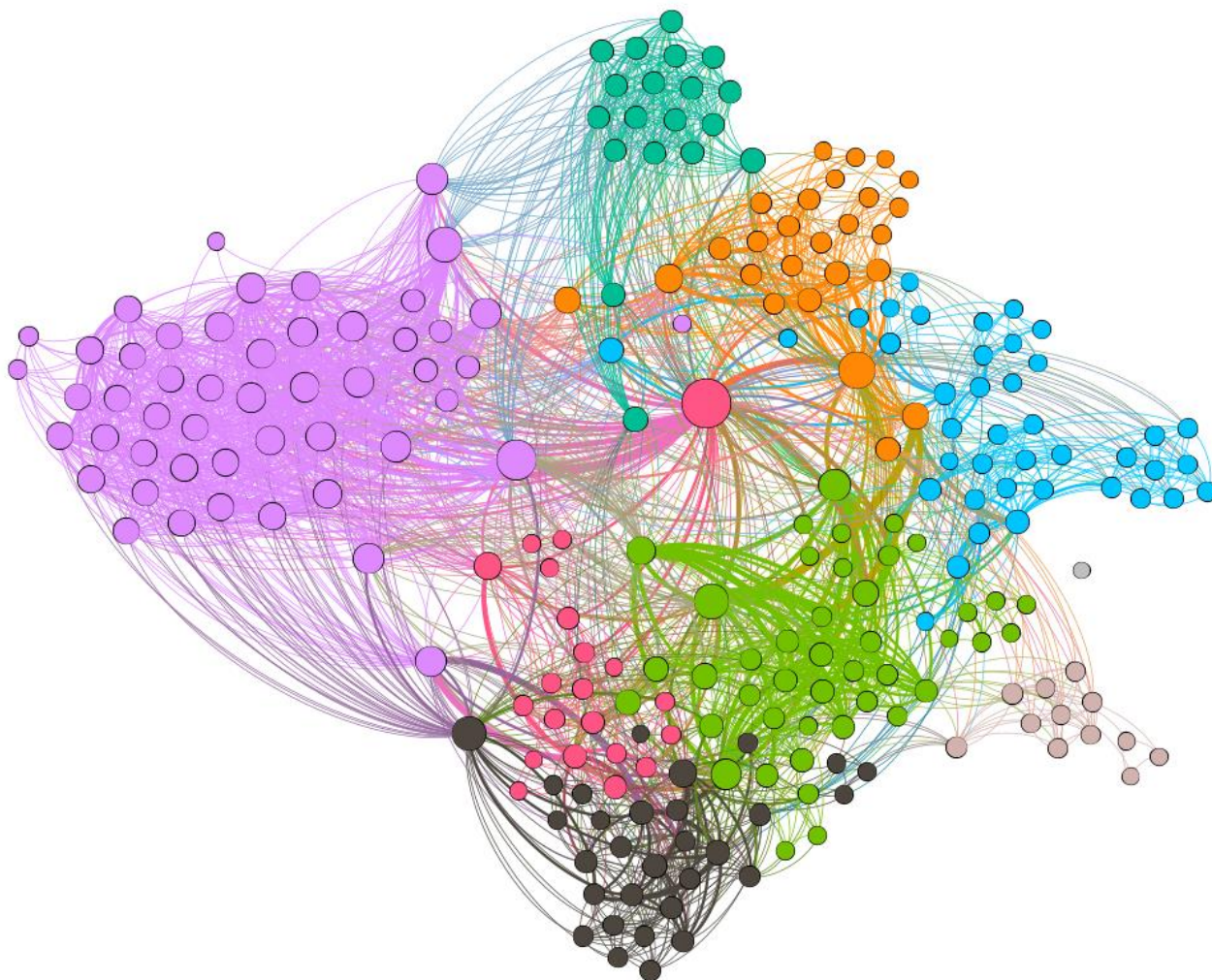
Као што се види из табеле 5.18.1 просечно се највише објављује у часопису Computer Science And Information Systems и на конференцији Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics).

Dokument	Tip rada	Broj radova	Broj autora	Prosečan broj autora po radu
Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics)	Conference Paper	76	104	1.37
Acm International Conference Proceeding Series	Conference Paper	20	32	1.60
Computer Science And Information Systems	Article	19	27	1.42
Yugoslav Journal Of Operations Research	Article	13	17	1.31
Ceur Workshop Proceedings	Conference Paper	13	22	1.69
Expert Systems With Applications	Article	13	18	1.38
International Journal Of Engineering Education	Article	12	26	2.17
Advances In Computers	Article	9	36	4.00
24Th Telecommunications Forum, Telfor 2016	Conference Paper	9	15	1.67

Табела 5.18.1 Часописи и конференција на којима се највише објављује

За више информација погледати фајл *analysis_journals.xlsx*.

5.19. Да ли се издвајају и које су групе (заједнице) часописа у оквиру мреже часописа? Да ли аутори имају тенденцију да објављују у часописима сличног профила?



Слика 5.19.1 Мрежа часописа

У мрежи часописа чвор представља часопис док тежина везе означава број аутора који су објавили радове у оба часописа.

Из приложене слике 5.19.1 можемо уочити да се формирају одређене густо повезане групе часописа у којима исти аутори најчешће објављују, што значи да аутори имају тенденцију да објављују у часописима сличних профила.

5.20. У којим годинама су аутори били најпродуктивнији по факултетима и катедрама?

У табели 5.20.1 су дате количине објављених радова по годинама за сваку катедру, док су у табели 5.20.2 дате количине објављених радова по годинама за сваки факултет.

Катедра за софтверско инжењерство са Факултета организационих наука (ФОН_СИ) је била најпродуктивнија 2012. године, Катедра за информационе технологије са Факултета организационих наука (ФОН_ИТ) је била најпродуктивнија 2013. године, а Катедра за информационе системе (ФОН_ИС) са Факултета организационих наука 2010. године. Факултет организационих наука је збирно најпродуктивнији био 2012. године.

Катедра за рачунарство и информатику (МАТФ_РТИ) са Математичког факултета и уједно и Математички факултет је био најпродуктивнији 2012. године.

Катедра за рачунарску технику и информатику (ЕТФ_РТИ) са Електротехничког факултета, па самим тим и Електротехнички факултет је такође био најпродуктивнији 2012. године.

ETF_RTI		MATF_RTI		FON_SI		FON_IT		FON_IS	
Godina	Broj radova	Godina	Broj radova	Godina	Broj radova	Godina	Broj radova	Godina	Broj radova
2012	31	2012	24	2012	25	2013	17	2010	8
2015	27	2017	22	2009	20	2012	15	2012	7
2013	26	2014	17	2004	20	2016	12	2016	6
2017	23	2013	15	2005	17	2009	11	2014	5
2016	21	2016	11	2006	16	2014	10	2017	4
2011	20	2015	11	2015	15	2018	8	2015	4
2018	16	2018	10	2007	15	2011	8	2011	4
2014	16	2006	9	2010	14	1999	7	1985	3
2005	13	2010	8	2008	13	2017	6	2018	2
2009	12	2009	7	2017	10	2010	6	2013	2

Табела 5.20.1 Количине објављених радова по годинама за сваку катедру

ETF		MATF		FON	
Godina	Broj radova	Godina	Broj radova	Godina	Broj radova
2012	31	2012	24	2012	47
2015	27	2017	22	2009	32
2013	26	2014	17	2013	27
2017	23	2013	15	2010	27
2016	21	2016	11	2016	25
2011	20	2015	11	2014	24
2018	16	2018	10	2015	23
2014	16	2006	9	2004	23
2005	13	2010	8	2011	22
2009	12	2009	7	2017	20

Табела 5.20.2 Количине објављених радова по годинама за сваки факултет

За више информација погледати фајл *analysis_modules.xlsx*.

6. Закључак

Приликом израде овог истраживачког задатка прошли смо кроз поступак припреме података за формирање мреже (такозвано чишћење), креирали смо пет различитих графова и визуелизовали мреже на начин погодан за даљу анализу социјаних мрежа. Обновили смо појмове као што су централност, дијаметар, просечна путања чвора, мали свет, power law итд. Цео поступак помогао нам је да лакше уочимо најпродуктивније факултете, као и ауторе, уочимо ауторе који највише сарађују унутар својих катедри као и оне који сарађују са ауторима са других факултета, најплодоносније године, најпопуларније часописе итд.