

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ (QUALITY ASSESSMENT
REPORT ABOUT PHD THESIS)

I ПОДАЦИ О КОМИСИЈИ (DATA ABOUT COMMITTEE MEMEBRS)		
1. Датум и орган који је именовео комисију (The date and body that named the committee):		
Наставно научно веће Универзитета у Новом Саду, Природно-математичког факултета на 18.седници одржаној 16.03.2023. године		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> (Committee members):		
1.	Говедарица др Мирó	редовни професор
	презиме и име	звање
	Факултет Техничких Наука	Геоинформатика 26.04.2012.
	установа у којој је запослен-а	ужа научна област и датум избора
		председник
		функција у комисији
2.	Месарош др Минучер	ванредни професор
	презиме и име	звање
	Природно-математички факултет	Друштвена географија 16.07.2019.
	установа у којој је запослен-а	ужа научна област и датум избора
		ментор
		функција у комисији
3.	Брдар др Сања	виши научни сарадник
	презиме и име	звање
	Институт БиоСенс	Информационе технологије 28.02.2023.
	установа у којој је запослен-а	ужа научна област и датум избора
		ментор
		функција у комисији
4.	Арсеновић др Даниела	ванредни професор
	презиме и име	звање
	Природно-математички факултет	Друштвена географија 11.07.2019.
	установа у којој је запослен-а	ужа научна област и датум избора
		члан
		функција у комисији
5.	Тешендић др Данијела	ванредни професор
	презиме и име	звање
	Природно-математички факултет	Информациони системи 17.11.2020.
	установа у којој је запослен-а	ужа научна област и датум избора
		члан
		функција у комисији
6.	Papadopoulos dr Apostolos N.	associate professor
	surname and name	title
	Faculty of Sciences Aristotle University of Thessaloniki	Data mining 04.05.2016.
	the institution	scientific area and election date
		member
		committee function
II ПОДАЦИ О КАНДИДАТУ (ABOUT THE CANDIDATE)		

1. Име, име једног родитеља, презиме (Name, name of one parent, surname):

Оливера (Небојша) Мулић (Нововић)

2. Датум рођења, општина, држава (Date of birth, municipality, state):

19.05.1991. Нови Сад, Србија

3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив (Name of the Faculty, name of the last studied program and academic title):

Факултет Техничких Наука, Мастер студије, Мастер инжењер Геодезије

Faculty of Technical Sciences, MSc studies, MSc in Geodesy

4. Година уписа на докторске студије и назив студијског програма докторских студија (The year when PhD studies are enrolled and name of the program):

2016, Доктор наука – Геонауке, PhD Geosciences

III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ (TITLE OF THE PHD THESIS):

„Примена Big Data аналитике за истраживање просторно временске динамике људске популације“
“Big Data analysis applied in space time human dynamics research”

IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ (OVERVIEW OF THE PHD THESIS):

Преглед докторске дисертације:

Докторска дисертација је написана на енглеском језику, са кључном документацијском информацијом писаном на српском и енглеском језику, као и кратким апстрактном писаним на српском и енглеском језику и продуженим апстрактном писаним на српском језику, ћириличним писмом. Дисертација садржи укупно 124 (нумерисане) странице, 7 табела, 43 слике и 2 прилога. Прилози су: кратка биографија кандидата на једној страници, и продужени апстракт на српском језику на 16 страница. Садржај докторске дисертације је распоређен у 8 издвојених поглавља:

1. Увод (стр. 1 – 14): У уводном поглављу је дат кратак преглед теме и истраживања у области , дефинисане су хипотезе докторског истраживања и дат је опсежан преглед релевантних извора података као и кратак преглед организације по поглављима и научни допринос.
2. Наука о мрежама у истраживањима људске популације (стр. 15 – 31): **У другом поглављу је дат преглед стања у области и релевантне литературе.** Извори литературе су релевантни за тему и наведени извори других аутора су адекватно цитирани. Преглед области је дат кроз три потпоглавља која описују истраживање динамике људске популације, теорију графова и основу науке о мрежама, посебно су издвојени граф алгоритми и алгоритми за кластеризацију.
3. Дистрибуирана аналитика података (стр. 33 – 43): У трећем поглављу су описане напредне методе дистрибуираног процесирања података уз коришћење *Big Data* технологије *Apache Spark* и *Hadoop*.
4. Мреже конективности кроз време (стр. 45 – 53): У четвртом поглављу описан је развој еволутивних мрежа повезаности кроз призму аналитике телеком података. Приказани су и описани експериментални резултати.
5. Детекција заједница у телеком мрежама конективности (стр. 55 - 64) : У петом поглављу је представљен теоријски концепт кластеризације над телеком подацима конективности, као и експериментални резултати и резултати везани за ефикасност процесирања у различитим

хардверским окружењима.

6. Динамика људске популације кроз призму телеком конективности (стр. 65 – 93): У шестом поглављу представљени су и описани резултати истраживања везано за евалуацију и квантификацију динамике људске популације коришћењем података о телеком интеракцијама људи и семантици простора. За сврху истраживања примењени су концепти из теорије графова, статистике, географске анализе простора, машинског учења и *Big Data* аналитике.
7. Мреже мобилности у урбаним просторима (стр. 95 – 110): У седмом поглављу представљена је примена методологије развијене у току докторског истраживања над новим сетом података који је везан за мобилност корисника у урбаном простору сакупљеним кроз платформу „*Foursquare*“.
8. Закључци (стр. 111 – 113): У осмом поглављу представљени су закључци везани за резултате истраживања, описана је могућност практичне примене методологије, научни допринос резултата и дати су предлози за будућа истраживања. Резултати јасно показују да је методологију могуће применити над подацима из различитих извора који имају структуру мреже и при томе тумачити резултате на начин прилагођен семантици података и простора.

У дисертацији је наведено 140 извора литературе који се односе на научне радове, књиге, техничку документацију и веб изворе (стр. 115 – 124). Прилог А: Кратка биографија кандидата на једној страници. Прилог Б: Продужени апстракт на српском језику представља кратак преглед докторске дисертације са издвојеним кључним резултатима.

Overview of the PhD thesis:

PhD thesis is written in English language, with key documentation and short abstract written both in English and Serbian language, while extended abstract is written in Serbian language, in Cyrillic letter. The thesis contains overall 124 pages, 7 tables, 43 figures, and 2 appendices. Appendices are short biography of the candidate (1 page) and extended abstract in Serbian language (16 pages). Main content of the PhD thesis is organized in 8 Chapters:

1. Introduction (pp. 1 – 14): This Chapter contains a short overview of the topic and related research. In this Chapter the main hypotheses are defined and a detailed overview of the data sources is presented. Also, thesis structure and contribution are presented in this Chapter.
2. Network science for studying human dynamics (pp. 15 - 31): The second Chapter contains related literature overview. Selected literature is important for the topic and sources from other authors are adequately cited. The overview is organized in 3 subsections which describe human dynamics research, graph theory and Network Science, graph algorithms and clustering algorithms.
3. Distributed data analysis (pp. 33 – 43): In the third Chapter, the advanced technologies for distributed and large scale data processing Apache Spark and Hadoop are presented.
4. Evolving connectivity networks (pp. 45 – 53): In this Chapter the time evolving connectivity networks derived from telecom data are presented with the discussion of experimental results.
5. Community detection in telecom connectivity networks (pp. 55 – 64): In this Chapter, clustering of telecom connectivity networks is presented with experimental results and results from performance testing in different processing environments.
6. Human dynamics evaluated through telecom connectivity networks (pp. 65 – 93): In this Chapter are presented the results related to human dynamics evaluated through telecom connectivity and spatial semantics. Diverse methods are applied such as graph theory, statistics, geographical analysis, Machine Learning and Big Data analysis.
7. Mobility networks in urban spaces (pp. 95 – 110): In this Chapter is presented the application of methodology developed with regard to telecom data over new data set that contains users mobility in urban spaces collected through Foursquare platform.
8. Conclusions (pp. 111 – 113): In this Chapter the main conclusions from the research are presented,

the means of methodology usage are described, scientific contribution of the results is discussed and suggestions for future research are provided. The results clearly show that methodology could be used with data from different sources if they have the network structure, and that the results can be further explained in the context of spatial semantics.

PhD thesis contains 140 references among which there are books, technical reports, scientific papers, and other web sources (pp. 115 – 124). Appendix A: Short biography of the candidate. Appendix B: Extended abstract in Serbian language that contains short summary of the main results and key findings from the research.

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ (ASSESSMENT OF SPECIFIC PARTS OF THE PHD THESIS):

Комисија сматра да је наслов докторске дисертације „Примена Big Data аналитике за истраживање просторно-временске динамике људске популације“, „Big Data analysis applied in space-time human dynamics research“ јасно дефинисан и приказује суштину проблематике обрађене у току истраживања.

У првом поглављу дат је кратак увод у област истраживања динамике људске популације. Описани су релевантни извори података. Такође, у овом делу наведен је предмет истраживања дисертације, постављени су циљеви и дефинисане су хипотезе. На самом крају поглавља дат је опис научног доприноса и описана је структура саме дисертације.

У другом поглављу дат је приказ основних концепата науке о мрежама и теорије графова са посебним нагласком на алгоритме над графовима и алгоритме за кластеризацију графова који су коришћени у току истраживања. Такође, преглед релевантне литературе у области динамике људске популације је дат у другом поглављу, литература је добро одабрана и адекватно цитирана.

Посебно значајно за ову дисертацију је што се бави динамиком људске популације кроз призму дигиталних података, као што су подаци о телеком интеракцијама и подаци са друштвених мрежа, што представља посебан допринос науци као и области истраживања динамике људске популације јер такав приступ није раније примењиван. Такви подаци су захтевни за анализу и интерпретацију због своје величине и комплексности па је за њихову употребу неопходно применити напредне *Big Data* технологије што је у овом истраживању и урађено и представља посебан допринос науци и области. Детаљан опис коришћених *Big Data* технологија је дат у трећем поглављу.

Комисија закључује да су поглавља Увод, Наука о мрежама у истраживањима људске популације и Дистрибуирана аналитика података (прво, друго и треће поглавље) концептуално и садржајно одговарајућа за дефинисану тему истраживања. Поглавља садрже детаљно наведену релевантну литературу у истраживачкој области. Кроз прва три поглавља дат је опширан увод у истраживање, дефинисане су хипотезе, дат је преглед области и значајних теоријских концепата и технологије која је примењена у експерименталном делу истраживања.

Четврто поглавље описује развој телеком мрежа конективности кроз време. Јасно је приказано, у нумеричким резултатима као и у географском приказу на мапама, да постоји изражена периодичност у формирању мрежа конективности у зависности од типа дана. Такође, показано је кроз експерименталне резултате да се својства графова који се односе на мрежу конективности могу користити као показатељи неких других догађаја у простору који су индуковани људском динамиком. Један део резултата који су описани у четвртом поглављу је објављен у конференцијском зборнику радова [4].

У петом поглављу су детаљно објашњени теоријски концепти кластеризације графова и специфичности конкретне алгоритма који је примењиван у истраживању представљеном у дисертацији, *Louvain* алгоритам. Метода кластеризације графа је примењена над телеком мрежом конективности да би се детектовале заједнице са снажном унутрашњом повезаношћу, што представља посебан поглед на људску динамику јер на тај начин детектовани кластери представљају просторне целине формиране од људских

интеракција. Такође, методологија описана у петом поглављу има велик практичан значај јер су коришћене напредне методе за *Big Data* аналитику у рачунарским експериментима. Део резултата описаних у петом поглављу ове дисертације је објављен у конференцијском зборнику радова [3].

Шесто поглавље описује динамику људске популације посматрану кроз призму телеком мреже конективности. У шестом поглављу приказани су резултати студије која обједињује целокупан теоријски и експериментални рад у оквиру докторске дисертације. Ова студија обухвата анализу локалних и глобалних својстава графова, утицај типа локације на динамику људске популације и методе ефикасног процесирања великих података. Поред тога, додатно је проширен оквир методологије применом машинског учења у експериментима, коришћен је регресиони модел да би се предвиђала својства графа на основу просторних особина локације које су дефинисане класама употребе земљишта. Овакав приступ у истраживању динамике људске популације представља иновацију и научни допринос у области сам по себи, а експериментални резултати потврђују валидност примене методологије и технологије у пракси. Резултати приказани у шестом поглављу су објављени у међународном научном часопису [1] и цитирани су већ 17 пута до сада.

У седмом поглављу описани су резултати везано за истраживање још једног важног аспекта људске динамике, а то је мобилност. За ову анализу употребљени су подаци са друштвене мреже „*Foursquare*“ за десет светских градова у периоду од две године. У овој студији је кретање корисника између две локације посматрано као грана у графу, док су чворови графа локације које су посећене. Број мобилности између две локације представља тежински коефицијент гране. Над графовима мобилности примењен је метод кластеризације како би се детектовале које структуре у граду се групишу заједно и под којим условима. Резултати кластеризације су анализирани у контексту семантички јединствених просторних целина.

Комисија оцењује да су поглавља Мреже конективности кроз време, Детекција заједница у телеком мрежама повезаности, Динамика људске популације кроз призму телеком конективности, Мобилне мреже у урбаним просторима (четврто, пето, шесто и седмо поглавље) јасно и прегледно написана, да је структура текста добро конципирана и да су резултати истраживања оригинални, примењиви у пракси и релевантни у области. Кандидаткиња Оливера Мулић је студиозно и објективно, користећи напредне технологије за аналитику великих података, детаљно анализирала податке и интерпретирала резултате у одговарајућем контексту.

У осмом поглављу су представљени закључци из целокупне студије која је изведена у току докторског истраживања. Потврђене су хипотезе које су дефинисане у уводном делу тезе и продискутовани су главни закључци везано за истраживање. У докторској дисертацији је представљен нови приступ у истраживању динамике људске популације и нови концепт „динамичности“ који је више виртуелни него физички. За моделовање људских интеракција коришћена је теорија графова и напредне технологије аналитике великих података. Представљена је методологија која користи граф аналитику, машинско учење и науку о подацима да би се извукло ново знање из различитих извора података који описују људску динамику.

Комисија закључује да су остварени планирани циљеви истраживања у оквиру докторске дисертације, да су постигнути значајни научни резултати у области, као и да је закључак рада јасно написан.

Након поглавља Закључак следи поглавље које садржи сву цитирану литературу. У оквиру целе дисертације наведено је 140 извора других аутора.

Комисија сматра да је литература цитирана на одговарајући начин, да је избор литературе добар и да наведене референце представљају значајне научне доприносе у области.

English translation:

The Committee agrees that the title of the PhD thesis „Примена Big Data аналитике за истраживање просторно-временске динамике људске популације“, „Big Data analysis applied in space-time human dynamics research“ is well defined and aligned with the core problem that was addressed in this research.

In the first chapter, a brief introduction to the field of human dynamics research is given. Relevant data sources are described. Also in this part, the main topic of the PhD research is explained, goals are set and hypotheses are defined. At the very end of the chapter, a description of the scientific contribution is given and the structure of the dissertation itself is described.

The second chapter presents the basic concepts of Network Science and graph theory with a special emphasis on graph algorithms and clustering algorithms that were used during the research. Also, an overview of the relevant literature in the field of human dynamics is given in the second chapter, the literature is well selected and adequately cited.

It is particularly important for this dissertation that it explores the human population dynamics through the prism of digital data, such as data about telecom interactions and data from social networks, which represents a special contribution to science and human dynamics research because such approach has not been applied before. Such data are demanding for analysis and interpretation due to their size and complexity, so for their use it is necessary to apply advanced Big Data technologies, which was done in this research and represents a special scientific contribution. A detailed description of the Big Data technologies used in this research is given in the third chapter.

The Committee concludes that the chapters Introduction, Network Science for studying human dynamics and Distributed data analytics (chapters one, two and three) are conceptually and content-appropriate for the defined research topic. The chapters contain cited relevant literature in the research area. Through the first three chapters, a comprehensive introduction to the research is given, hypotheses are defined, an overview of the field and significant theoretical concepts and technology applied in the experimental part of the research is given.

The fourth chapter describes the development and evolution of telecom connectivity networks over time. It is clearly presented in the numerical results and in the geographical maps, that there is a pronounced periodicity in the formation of connectivity networks depending on the type of day. Also, it was shown through experimental results that the properties of the graphs related to the connectivity network can be used as indicators of some other events in space that are induced by human dynamics. A part of the results described in the fourth chapter was published in the conference proceedings [4].

In the fifth chapter, the concepts of clustering and specially the Louvain clustering algorithm are explained in detail. The graph clustering method was applied over the telecom connectivity network to detect communities with strong internal connectivity, which represents a special scientific contribution because the clusters detected in this way are spatial entities formed by human interactions. Also, the methodology described in the fifth chapter has great practical importance because advanced methods for Big Data analytics were used in the experiments. A part of the results described in the fifth chapter was published in the conference proceedings [3].

The sixth chapter describes the human dynamics observed through the prism of the telecom connectivity network. In this chapter are presented the combined results of the entire theoretical and experimental research work done within the doctoral study. This study includes the analysis of local and global properties of graphs, the influence of location type on human dynamics, and methods of efficient processing of large-scale data. In addition, the methodology was further expanded by utilizing machine learning in the experiments. The regression model was used to predict the properties of the graphs based on the spatial properties of the location defined by the land use classes. Such an approach in the research of human dynamics represents an innovation and a scientific contribution itself, and the experimental results confirm the validity of applying the methodology

and technology in practice. The results presented in the sixth chapter were published in a scientific journal [1] and have been cited 17 times so far.

The seventh chapter describes the results related to the research of another important aspect of human dynamics, which is mobility. For this analysis, data from the social network "Foursquare" was used for ten worldwide cities over a period of two years. In this study, the user's movement between two locations is viewed as a branch in the graph, while the nodes of the graph are the locations visited. The number of movements between two locations represents the weight coefficient of the branch. Clustering was applied to the mobility graphs to detect which structures in the city cluster together and under what conditions. The clustering results were analyzed in the context of semantically unique spatial units.

The Committee agrees that the chapters Evolving connectivity networks, Community detection in telecom connectivity networks, Human dynamics evaluated through telecom connectivity networks, Mobility networks in urban areas (fourth, fifth, sixth and seventh chapters) are clearly and well written, that the structure of the text is well conceived and that the research results are original, applicable in practice and relevant in the field. Candidate Olivera Mulić studiously and objectively, using advanced technologies for Big Data analytics, analyzed the data in detail and interpreted the results in the appropriate context.

In the eighth Chapter are presented the conclusions from the entire study which was carried out during the doctoral research. The hypotheses defined in the introductory part of the thesis were confirmed and the main conclusions related to the research were discussed. In the PhD thesis are presented a new approach to the research of human dynamics and a new concept of dynamics that is more virtual than physical. Graph Theory and advanced Big Data technologies were used to model human interactions. Unique methodology is presented, that uses graph analytics, Machine Learning, and Data Science to extract new knowledge from diverse data sources related to human dynamics.

The Committee concludes that the planned goals of the research within the doctoral dissertation have been achieved, that significant scientific results have been accomplished, and that the conclusion of the work is clearly written.

After concluding chapter there is an unnumbered chapter containing all cited literature. Within the entire dissertation, 140 sources of other authors are listed.

The Committee agrees that literature is cited in an appropriate manner, that the selection of literature is good and that the cited references represent significant scientific contributions in the field.

VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ

(LIST OF PUBLISHED OR ACCEPTED FOR PUBLISHING PAPERS THAT ARE RESULT OF THE PHD RESEARCH):

1. **Novović, O.,** Brdar, S., Mesaroš, M., Crnojević, V., & N. Papadopoulos, A. (2020). Uncovering the relationship between human connectivity dynamics and land use. *ISPRS International Journal of Geo-Information*, 9(3), 140. **(M22)**
2. **Novović, O.,** Grujić, N., Brdar, S., Govedarica, M., & Crnojević, V. (2020). Clustering foursquare mobility networks to explore urban spaces. In *Trends and Innovations in Information Systems and Technologies: Volume 3 8* (pp. 544-553). Springer International Publishing. **(M33)**

3. Truić, C. O., **Novović, O.**, Brdar, S., & Papadopoulos, A. N. (2018). Community detection in who-calls-whom social networks. In *Big Data Analytics and Knowledge Discovery: 20th International Conference, DaWaK 2018, Regensburg, Germany, September 3–6, 2018, Proceedings 20* (pp. 19-33). Springer International Publishing. (M33)
4. **Novović, O.**, Brdar, S., & Crnojević, V. (2017, April). Evolving connectivity graphs in mobile phone data. In *NetMob, The main conference on the scientific analysis of mobile phone datasets* (pp. 73-75). (M33)
5. Brdar, S., **Novović, O.**, Grujić, N., González-Vélez, H., Truić, C. O., Benkner, S., ... & Papadopoulos, A. (2019). Big data processing, analysis and applications in mobile cellular networks. *High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet*, 163-185. (M14)

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА (CONCLUSIONS AND RESULTS OF THE RESEARCH):

На основу резултата истраживања, приликом којег су реализовани постављени циљеви и задаци уз примену научне методологије и коришћењем многобројне научне литературе из мултидисциплинарних истраживања везаних за динамику људске популације, аналитику великих података, аналитику над графовима, алгоритме и рачунарске технологије, изведени су закључци који потврђују иницијално дефинисане хипотезе. Кратко ћемо се осврнути на конкретне хипотезе и образложити на који начин су потврђене.

1. Обрасци понашања везаних за људску динамику су снажно повезани са временом и типом дана, као и сатима у току дана и радним, одн. нерадним данима – у оквиру ове дисертације посматрани су обрасци у динамици људске популације зависно од доба дана, типа дана, сати са појачаном активношћу, радних и нерадних сати у току дана, ноћних сати, викенда и празника. На основу анализираних података генерални закључак је да је динамика људске популације снажно условљена временом. Начин на који су формиране мреже конективности из телеком саобраћаја изразито се разликују у појединим временским интервалима.
2. Обрасци у људској динамици су географски дефинисани и груписани унутар просторних јединица са сличном семантиком локације – кроз различите рачунарске огледе над подацима закључено је да су просторне локације са различитом семантиком повезане са различитим облицима људске динамике и понашања и да су стварале различите „дигиталне отиске“ у подацима на јасан и недвосмислен начин.
3. Људска динамика и интеракције, представљене кроз конективност и мобилност мреже могу се даље анализирати помоћу теорије графова и кластеризације за издвајање корисних својстава на нивоу градова и његових целина – у оквиру дисертације је на иновативан начин представљено како се мреже људских интеракција забележених кроз телеком саобраћај могу ефикасно моделовати уз помоћ теорије графова, јер управо таква структура, структура графа најбоље одговара реалној мрежи људских интеракција. Оно што је карактеристично за ту мрежу је да је она просторно дефинисана па се самим тим све што се квантификује кроз анализу графова може и мапирати у географском простору. Показано је да се применом методе кластеризације над телеком графовима конективности долази до просторних целина које имају особину снажне унутрашње повезаности а самим тим и изражен капацитет за проток информација.
4. Начин коришћења и функционалност локације снажно утичу на предвиђање својстава конективности и мобилности која су у корелацији са људском динамиком – у оквиру истраживања описаног у докторској дисертацији дефинисана су три међусобно повезана

концепта. Семантика локације, динамика људске популације и својства мреже одн. графова. Показано је да семантика локације условљава динамику људске популације која се може моделовати путем мреже конективности. Коришћењем теорије графова квантификована су својства мреже конективности, а самим тим и динамике људске популације. Експериментално је показано да се семантика просторне локације може користити за предвиђање својстава мреже, а имплицитно и динамике људске популације.

5. Технологије аналитике великих података могу се користити за дизајнирање и развој ефикасних токова обраде података – у оквиру ове дисертације је показно да постоји велика потреба за применом напредних технологија за аналитику података, јер су кориснички генерисани подаци којима је дефинисана динамика људске популације велики по обиму и комплексни по структури. Кроз рачунарске огледе је показано које су предности коришћења напредних технологија и квантификована је њихова ефикасност кроз посебне огледе за мерење перформанси.

Комисија позитивно оцењује дефинисане хипотезе, резултате и закључке истраживања.

English translation:

Based on the research results, it is concluded that the initially defined hypotheses are confirmed. The research is conducted regarding well established scientific methodology, using large body of literature from multidisciplinary research related to graph analysis, human dynamics, data mining, Big Data analytics, Data Science, Computer Sciences and many more. We would briefly provide an overview of the hypothesis and their conclusions.

1. Patterns in human dynamics and behavior are strongly correlated with daytime and day type, distinguishing between hours within a day, but also between working days and non-working days. Within this dissertation, patterns in human dynamics were observed depending on the time of day, type of day, hours with increased activity, working and non-working hours during the day, night hours, weekends, and holidays. Based on the analyzed data, the general conclusion is that human dynamics is strongly conditioned by time. The way in which connectivity networks were formed differs significantly in certain time intervals.
2. Patterns in human dynamics are geographically mapped and clustered within spatial units with similar location semantic – through various computer experiments, it was concluded that spatial locations with different semantics are associated with different forms of human dynamics and behavior and that they created different "digital footprints" in the data in a clear and unambiguous way.
3. Human dynamics and interactions, represented through connectivity and mobility networks, can be further analyzed by means of graph theory and clustering to extract informative properties at the level of cities and its units – within the dissertation, it was presented in an innovative way how the networks of human interactions evaluated through telecom traffic can be effectively modeled with the help of Graph Theory, because exactly such a structure, the structure of the graph, best suits the real network of human interactions. What is characteristic of that network is that it is spatially defined, so everything that is quantified through graph analysis can also be mapped in geographic space. It has been shown that the application of the clustering method over telecom connectivity graphs leads to spatial entities that have the characteristic of strong internal connectivity and, therefore, a pronounced capacity for information flow.
4. Land use type and location functions strongly impact the prediction of connectivity and mobility properties that are correlated with human dynamics – within the scope of the research, three interrelated concepts were defined. Location semantics, human population dynamics and network properties, respectively graphs. It is shown that the semantics of the location conditions the human dynamics, which can be "registered" through the network of connectivity. Using Graph Theory, the properties of the connectivity network and thus the dynamics of the human population were quantified. It has been

experimentally demonstrated that spatial location semantics can be used to predict network properties, and implicitly, human population dynamics.

5. Big data technologies need to be utilized for designing more efficient data processing workflows – within this dissertation, it is shown that there is a great need for the application of advanced technologies for data analytics, because the user-generated data that defines the dynamics of the human population is large in volume and complex in structure. The advantages of using advanced technologies were demonstrated through computer experiments and their effectiveness was quantified through special experiments to measure performance.

The Committee positively evaluates the defined hypotheses, results, and conclusions of the research.

**VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА
(ASSESSMENT OF THE PRESENTATION AND EXPLANATION OF THE RESULTS):**

Кандидаткиња Оливера Мулић је на јасан и одговарајући начин изложила, протумачила и представила резултате спроведеног истраживања кроз детаљан приказ примене теоријских концепата и експерименталних резултата. На основу резултата су изведени закључци који одговарају предмету студије и пружају одговоре на постављене циљеве истраживања. Структура рада поклапа се са планом и дефинисаним циљевима у пријави теме. Наведена литература коришћена у раду је савремена, обимна и релевантна за тему истраживања. Резултати су научно оправдани и подржани бројним литературиним изворима. На основу начина приказивања и тумачења података, може се констатовати да рад садржи оригиналне научне резултате који задовољавају захтеве нивоа докторске дисертације.

English translation:

The PhD candidate Olivera Mulić presented, explained, and discussed the results of the research in a good manner regarding theory and practical experiments. Based on the results, conclusions were drawn that correspond to the subject of the study and provide answers to the research objectives. The structure of the paper matches the plan and defined objectives. The cited literature used in the paper is modern, extensive and relevant to the topic of the research. The results are scientifically justified and supported by numerous literature sources. Based on the way the data is presented and interpreted, it can be concluded that the research work contains original scientific results that meet the requirements of the doctoral dissertation level.

**IX КОНАЧНА ОЦЕНА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ
(FINAL NOTE ABOUT PHD THESIS):**

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме?
Is dissertation well aligned with the description in the topic submission?

Комисија сматра да је дисертација кандидаткиње Оливере Мулић написана у складу са предлозима и образложењима изложеним у пријави теме.

English translation:

The Committee agrees that the dissertation of candidate Olivera Mulić was written in accordance with the proposals and explanations presented in the topic submission.

2. Да ли дисертација садржи све битне елементе?
Does dissertation contain all necessary elements?

Комисија оцењује да дисертација садржи све битне елементе оригиналног научног рада: јасно дефинисану тему истраживања, одговарајућу методологију уз употребу напредних технологија за аналитику великих података, преглед постојеће литературе у мултидисциплинарним областима које обухватају динамику људске популације, геонауке и географске анализе, рачунарске науке, посебне области у науци о подацима и науци о мрежама. Такође, резултати су представљени на јасан начин и њихово значење и значај су продискутовани у складу са захтевима научног рада.

English translation:

The Committee positively assesses the dissertation and concludes that it contains all the essential elements of an original scientific work: a clearly defined research topic, an appropriate methodology with the use of advanced Big Data analytics technologies, a review of existing literature in multidisciplinary areas that include human dynamics, Geosciences and geographic analysis, Computer Sciences, special areas in Data Science and Network Science. Also, the results are presented in a clear way and their meaning and importance are discussed in accordance with the requirements of scientific work.

3. По чему је дисертација оригиналан допринос науци?
According to what is the dissertation original scientific contribution?

Главни допринос ове дисертације састоји се из две целине. Прва целина се односи на могућности примене анализе података из извора који до сада нису разматрани у истраживањима динамике људске популације, као што су телекомуникациони подаци и подаци са друштвених мрежа. Друга целина се односи на технички део везан за анализу великих и комплексних података применом метода *Big Data* технологија и алгоритама за анализу графова и машинског учења. Ове две целине заједно представљају веома значајан допринос науци, с обзиром да у литератури која је испитана није пронађен свеобухватан пример који демонстрира симултану употребу нових извора података као и примену напредних софтверских технологија и алгоритама за анализу великих података. Стога, ова студија даје нови увид и нова сазнања везана за динамику људске популације и представља добар основ за даља истраживања у тој области и друга слична мултидисциплинарна истраживања. Такође, добијени резултати у овој студији представљају прво детаљно истраживање у области динамике људске популације у коме се користе напредни алати за процесирање великих количина података као и алгоритми из теорије графова и машинског учења.

Напомена

Докторска дисертација је у библиотеци Природно-математичког факултета прошла проверу оригиналности применом софтвера iThenticate за детекцију сличности који је показао да „индекс сличности“ (Similarity Index) износи 50%. Применом софтвера утврђено је да од 120 извора са којима

је утврђено преклапање, у 96% извора је пронађено мање од 1% подударности. Извори у којима је пронађено више од 1% подударности (извори 1, 2, 3, 4 и 5 у извештају сличности) представљају научне радове где је кандидаткиња Оливера Мулић први аутор или један од аутора. Анализа сличности је рађена над корпусом од 32 859 речи из докторске дисертације почевши од другог поглавља тезе, што даје 16 453 речи где је пронађена подударност са другим изворима. Важно је истаћи да се преглед литературе у области, као и приказ теоријских основа везаних за Теорију графова и Науку о мрежама налази у другом поглављу тезе где је детектовано 87% подударности, што оставља свега 13% подударности за остала поглавља где су представљени резултати и главни доприноси истраживања. Такође, од укупног корпуса у коме је пронађена сличност, 77% се односи на радове где је кандидаткиња Оливера Мулић први аутор или један од аутора. Сва литература других аутора која је коришћена у току истраживања је у тези наведена и адекватно цитирана.

На основу свега наведеног у извештају, комисија закључује да је ова докторска дисертација оригинално дело и даје позитивну оцену за приказ и тумачење резултата истраживања.

English translation:

The main contribution of this dissertation consists of two parts. The first part refers to the possibilities of applying the analysis of data from sources that have not been considered so far in human dynamics research, such as telecommunication data and data from social networks. The second part refers to the technical objective related to the analysis of large and complex data using the Big Data technologies and algorithms for graph analysis and Machine Learning. Together, these two parts represent a very significant contribution to science, given that no comprehensive example was found in the reviewed literature demonstrating the simultaneous use of new data sources as well as the application of advanced software technologies and algorithms for the analysis of very large data. Therefore, this study provides new insight and new knowledge related to human population dynamics and represents a good basis for further research in that area and other similar multidisciplinary research. Also, the results obtained in this study represent the first detailed research in the field of human population dynamics using advanced tools for processing large scale data, algorithms from Graph Theory, and Machine Learning.

Note

The PhD thesis was checked for originality in the library of the Faculty of Science using iThenticate software for detecting similarities, which showed that the "Similarity Index" is 50%. Using the software, it was determined that out of 120 sources where an overlap was found, in 96% of the sources less than 1% match was found. Sources where more than 1% match was found (sources 1,2,3,4 and 5 in the similarity report) represent scientific papers where candidate Olivera Mulić is the first author or one of the authors. Similarity analysis was performed on a corpus of 32,859 words from the doctoral dissertation starting from the second chapter of the thesis, which gives 16,453 words where a match with other sources was found. It is important to point out that the review of the literature in the field, and theoretical foundations related to Graph Theory and Network Science, is contained in the second chapter of the thesis, where 87% matches was detected, which leaves only 13% matches for other chapters where the results and main research contributions were presented. Also, of the total corpus in which similarity was found, 77% refer to papers where candidate Olivera Mulić is the first author or one of the authors. All literature from other authors that was used during the research is listed and adequately cited in the thesis.

Based on everything stated in the report, the Committee concludes that this doctoral dissertation is an original work and gives a positive assessment for the presentation and interpretation of research results.

4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?
What are the flows of the dissertation and how do they affect the results of the research?

Увидом у докторску дисертацију и детаљним прегледом свих резултата комисија није уочила недостатке који би утицали на резултате истраживања и закључке рада.

English translation:

Upon inspection of the doctoral dissertation and a detailed review of all results, the Committee did not notice any deficiencies that would affect the research results and the conclusions of the work.

X ПРЕДЛОГ (THE PROPOSAL):

На основу наведеног, комисија предлаже:
Based on all stated facts, committee proposes:

**а) да се докторска дисертација прихвати, а кандидату одобри одбрана;
(to accept the phd thesis and allow candidate to defend the thesis)**

б) да се докторска дисертација врати кандидату на дораду (да се допуни односно измени);
(to return the thesis to the candidate for minor/major edits)

в) да се докторска дисертација одбије.
(to reject the phd thesis).

Место и датум:

Place and date: Нови Сад, 12.05.2023.

1. Говедарица др Миро, редовни професор

_____, председник

2. Месарош др Минучер, ванредни професор

_____, ментор

3. Брдар др Сања, виши научни сарадник

_____, ментор

4. Арсеновић др Даниела, ванредни професор

_____, члан

5. Тешендић др Данијела, ванредни професор

_____, члан

6. Papadopoulos dr Apostolos N., associate professor

_____, member