

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата **мр Јамал Салем Али Бзаи** за израду докторске дисертације

Одлуком бр. 5052/14-1 од 20. 09. 2018. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата **мр Јамал Салем Али Бзаи** за израду докторске дисертације и научне заснованости теме:

„Побољшање перформанси обраде великих количина података применом сличности над детектованим заједницама у мрежном окружењу“
„Enhancing Performance of Big Data applying Similarity over Detected Community“.

На основу материјала приложеног уз Захтев кандидата, после одржане јавне усмене одбране теме и докторског испита о подобности кандидата, Комисија за оцену подобности теме и кандидата подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Мр Јамал Салем Али Бзаи је рођен 01 октобра 1966 у Триполију, Либија. Уписан је на Faculty of Science, Computer Science Department, Tripoli University, Tripoli, Libya, школске 1985/86 године и дипломирао 1990. године. Мастер студије је уписао на Електротехничком факултету у Београду 1997. године и одбранио мастер рад 2000. године са темом "Analysis and evaluation of hardware & software fault tolerant architectures".

У периоду од 1991-2004. године радио је у R&D Center – Tripoli, где се бавио програмирањем у dBase 4 (Inventory system Application), развојем софтвера за финансијске примене. Држао је и курсеве из Fortran-a, Pascal-a and C-a.

У периоду од 2001-2013. године радио је на Electronic Technical College, full-time Lecturer– Tripoli - Libya, где је био предавач на предметима: Indroduction to computer science, Programming in C/C++, Pascal languages, Concepts of Operating Systems, Data structure, System analysis, Computer Foundation, Computer design, Html & PHP, Computer & numerical. У том периоду је држао курсеве и на Alnjella High Institute of General Professional, part-time Lecturer – Tripoli - Libya, Sok Algomma Higher Institute of General Professional, part-time Lecturer, Tripoli, Sok alhamis (Awata) Higher Institute of General Professional, part-time Lecturer–Sok alhamis, Tariq almatatar Higher Institute of General Professional, part-time Lecturer – Tripoli, Tripoli university, part time lecturer - Tripoli.

Докторске студије на Електротехничком факултету у Београду је уписао 2014. године.

1.2. Стечено научноистраживачко искуство

Кандидат је у јануару 2000. године одбранио мастер тезу под насловом „Analysis and evaluation of hardware & software fault tolerant architectures“ на Електротехничком факултету, универзитета у Београду код ментора проф. др Мирослава Бојовића. Ужа научна област предложене теме су рачунарски системи са смањеном осетљивошћу на отказе. Магистарски рад обухвата анализу и евалуацију хардверских и софтверских архитектура рачунарских система са смањеном осетљивошћу на отказе.

Након одбраћене магистарске тезе кандидат се бавио развојем сложених софтверских система различите намене.

Кандидат је 18.10.2018. године успешно усмено одбранио тему докторске дисертације и положио испит о подобности кандидата (докторски испит) пред Комисијом за јавну усмену одбрану теме докторске дисертације у саставу др Драган Бојић, ванредни професор, Универзитет у Београду – Електротехнички факултет (председник комисије), др Леонид Стоименов, редовни професор, Универзитет у Нишу – Електронски факултет, др Марија Рашајски, ванредни професор, Универзитет у Београду – Електротехнички факултет.

Одбраћена мастер теза:

1. Јамал Салем Али Бзаи: *Analysis and evaluation of hardware & software fault tolerant architectures*, Електротехнички факултет, Универзитет у Београду, 2000, М72

Радови на конференцијама:

1. Јамал Салем Али Бзаи, Мирослав Бојовић: *Big Data Overview, the Characteristics, and Applications*, симпозијум INFOTEN-JAHORINA, VOL 16, март 2017. М33

1.3. Оцена подобности кандидата за рад на предложеној теми

Комисија је размотрила све претходно наведене резултате досадашњег рада кандидата. Кандидат је кроз одбраћену мастерску тезу, учешће у стручним пројектима, успешну одбрану теме докторске дисертације и објављени рад показао да је стекао неопходно искуство за самосталан научно-истраживачки рад. Комисија сматра да је кандидат способан да током израде докторске дисертације оствари резултате који ће бити верификовани од стране научне и стручне заједнице. Комисија констатује да кандидат Јамал Салем Али Бзаи испуњава све потребне услове да приступи изради докторске дисертације под наведеном темом.

2. Предмет и циљ истраживања

Друштвена мрежа се може моделовати као вишеструки и хетерогени скуп података приказан графом који се састоји од скупа великог броја ивица и чворова. У случају великих количина података представљених графом, граф обично није структуриран и врло је неправилан. Примарни фокус ове тезе је на аналитичким методама које се користе за моделовање и анализу великих количина података, са нагласком на анализи неструктурираних података, који могу да чине 95% великих података.

Аналитичке методе које се могу применити на велике количине хетерогених података у неструктурираним текстуалним форматима какви постоје у друштвеним мрежама представљеним помоћу графова, представљају једно од актуелних области истраживања. Алгоритми презентирани у расположивој литератури [1,...,15] који се базирају на графичком приказу су тешки за анализу и претварање у употребљив облик за доношење одлука у управљању великим количинама података.

Предложено истраживање отвара могућности за израду нових алата за предиктивну анализу за неструктуриране велике количине података. Статистичке методе су осмишљене како би се могли доносити закључци из података о узорку. Ово истраживање предлаже развој рачунарско ефикасног алгоритма који се базира на детекцији заједнице (community detection).

2.1 Референце

- [1] S. Günnemann, B. Boden, and T. Seidl. Db-csc: a density-based approach for subspace clustering in graphs with feature vectors. In *PKDD 2011*, Springer-Verlag, 2011., pages 565–580.

- [2] C. Aggarwal, Y. Zhao, and P. Yu. Outlier detection in graph streams. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*, IEEE, 2011., pages 399–409.
- [3] V. Satuluri, S. Parthasarathy, and Y. Ruan. Local graph sparsification for scalable clustering. In *SIGMOD 2011*, ACM, 2011., pages 721–732.
- [4] Y. Zhou, H. Cheng, and J. Yu. Clustering large attributed graphs: An efficient incremental approach. In *ICDM 2010*, IEEE, 2010., pages 689–698.
- [5] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5), 2010., pages 75–174.
- [6] J. Leskovec, K. Lang, and M. Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, ACM, 2010., pages 631–640.
- [7] A. Maiya and T. Berger-Wolf. Sampling community structure. In *WWW 2010*, ACM, 2010., pages 701–710.
- [8] W. Tang, Z. Lu, and I. Dhillon. Clustering with multiple graphs. In *ICDM'09*, IEEE, 2009., pages 1016–1021.
- [9] T. Yang, R. Jin, Y. Chi, and S. Zhu. Combining link and content for community detection: a discriminative approach. In *SIGKDD '09*, ACM, 2009., pages 927–936.
- [10] S. Zhu, K. Yu, Y. Chi, and Y. Gong. Combining content and link for classification using matrix factorization. In *SIGIR '07*, ACM, 2007., pages 487–494.
- [11] D. Zhou, E. Manavoglu, J. Li, C. Giles, and H. Zha. Probabilistic models for discovering e-communities. In *WWW'06*, ACM, 2006., pages 173–182.
- [12] I. Dhillon, Y. Guan, and B. Kulis. Kernel k-means: spectral clustering and normalized cuts. In *SIGKDD '04*, ACM, 2004., pages 551–556.
- [13] A. Strehl and J. Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research*, 2003., pages 583–617.
- [14] S. van Dongen. Graph clustering by flow simulation. *PhD Thesis*, 2000.
- [15] G. Karypis and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20(1), 1998., pages 359–392.

3. Полазне хипотезе

Нагли пораст потреба за анализом великих количина података захтева развој нових ефикаснијих алгоритама. Опет, сама графичка анализа генерише много података које треба додатно анализирати. На пример, мало је консензуса о основном питању колико су велики подаци потребни да би се квалификовали као „велики подаци“. Стога постоји потреба да се у академском окружењу документује еволуција концепта и технологија великих података, генерисаних графова и снажног захтева за аутоматизацијом аналитике како би се олакшао посао доносилаца одлука.

4. Научне методе истраживања

Овај истраживачки пројекат односи се на суштинске аспекте брзог доношења одлука анализирањем графова података, како би се спречио огроман губитак услед ограничених ресурса и мањка ефикасности запослених. Истраживачки проблем заснован је на анализи

великих количина података нестандартних шаблона, који би, правилном употребом, помогао у управљању одлучивањем, као и повећању мотивације међу заједницама, што даље олакшава њихов посао применом предложеног окружења. Слично, учење би се користило у прикупљању података, у препознавању лица или идентификацији лица и у системима за доношење препорука. Такође, многи приступи машинском учењу се ослањају на неку метрику. Ово подразумева учење без надзора, као што је груписање, које групише исте или сличне објекте. Такође укључује приступе као што је алгоритам тзв. К-најближих суседа који се ослањају на особине околних објеката како би одлучили у вези особине новог објекта. Метричко учење је предложено као корак који предходи многим од ових приступа за тестирање резултата.

5. Очекивани научни допринос

Кључни допринос овог истраживања је истицање често занемарених димензија великих података представљених графовима. У конкретном случају непреклапајућег проналажења заједнице, ово подразумева да се мрежа природно дели на групе чворова са густим везама унутар и ретким везама између група. Међутим, такође су дозвољена преклапања заједница. Основни принцип је да је пар чворова вероватније повезан ако су оба чланови исте заједнице и мање вероватно повезан ако не деле заједницу. Ово се може догодити у друштвеној мрежи у којој сваки чвор представља особу, а заједнице представљају различите групе пријатеља: једну заједницу за породицу, другу заједницу за сараднике, трећу за пријатеље у истом спортском клубу и тако даље. Популарни приступ великим подацима, који је доминантно условљен маркетиншким напорима програмера великих софтверских и хардверских компанија, фокусира се на предиктивну аналитику и структуриране податке. Тиме се игнорише највећа компонента великих података, а то је неструктурираност и доступност као комбинација чворова и ивица насталих из неструктурираног текста. Процењује се да структурирани подаци припремљени за анализу формирају само мали подкуп целокупних података. Неструктурирани подаци, посебно подаци у графичком формату, представљају највећу компоненту великих података и само су делимично архивирани.

Главни научни доприноси који се очекују од предложене тезе су:

- 1) Упоредна анализа и стратегија имплементације алгоритама коришћених за графове који се јављају у различитим врстама мрежа (нпр. социјалних или пословних мрежа). Посебна пажња ће бити посвећена методама кластеровања података.
- 2) Развој алгоритма који анализира велике количине неструктурираних података представљених графовима и
- 3) Развој окружења које решава проблем анализе великих графова користећи предложени алгоритам.

6. План истраживања и структура рада

У овој тези се предлаже истраживање и реализација система који користи алгоритам за детекцију заједнице и алгоритам рангирања сличности како би се решио проблем са типично неструктурираним и изузетно нерегуларним подацима представљеним графом, какве генеришу друштвене мреже. Слично, учење је уско повезано са метричким учењем на даљину. Метричко учење представља процес учења функције удаљености објеката. Користећи неколико алата за анализу података, биће израчуната метрика за различите величине мреже. Алгоритам за откривање заједнице ће бити базиран на великим мрежама и имплементиран у дистрибуираном окружењу ради веће ефикасности извршавања алгоритма. На исти начин, алгоритам рангирања сличности ће се имплементирати на резултатима алгоритма за детекцију заједнице ради тестирања добијених резултата.

У првом поглављу биће дат преглед расположиве литературе о алгоритмима за анализу велике количине података представљене графовима. Овде ће такође бити дискутовани проблеми који се јављају приликом процесирања великих количина података. У другом поглављу ће бити дата анализа и стратегија имплементације алгоритама коришћених за графове који се јављају у различитим врстама мрежа (нпр. социјалних или пословних мрежа). У трећем поглављу биће дат преглед разних метода кластеровања података. У првом делу ће се говорити о значају груписања, а затим би се описивале различите технике кластеровања. Поглавље четири описиваће израду окружења и студија случаја имплементираних алгоритама. Истражују се различити механизми проналажења заједница који се у данашње време користе. Пето поглавље ће увести предложени алгоритам и детаљно разматрати поступак, окружење и опис предложеног метода. Овде ће такође бити описани тестирање и учење мреже. Затим ће бити дата анализа прикупљених података и предложена стратегија која повећава продуктивност анализе графова за доношење одлука помоћу предложеног окружења. Поред тога, поглавље шест даће и анализу експеримената коришћених у овом раду, као и анализу резултата. Последње поглавље приказаће перформансе предложеног алгоритама у области откривања заједнице, након чега ће бити дат закључак и даљи правци истраживања.

7. Закључак и предлог

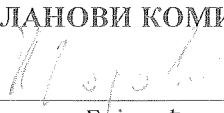
На основу свега изложеног, Комисија констатује да кандидат Мр Јамал Салем Али Бзаи испуњава све потребне услове да може приступити изради докторске дисертације под насловом **„Побољшање перформанси обраде великих количина података применом сличности над детектованим заједницама у мрежном окружењу“** (енглески **„Enhancing Performance of Big Data applying Similarity over Detected Community“**).

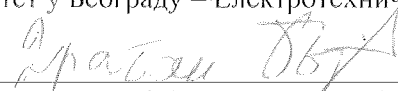
Ужа научна област предложене теме је софтверско инжењерство.

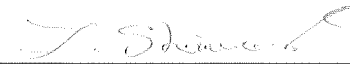
Стога комисија предлаже Научно-наставном већу Електротехничког факултета у Београду да одобри кандидату Мр Јамал Салем Али Бзаи израду докторске дисертације под горе наведеним насловом под руководством ментора проф. др Мирослава Бојовића.

У Београду, 15. новембар 2018. године

ЧЛАНОВИ КОМИСИЈЕ


др Мирослав Бојовић, редовни професор
Универзитет у Београду – Електротехнички факултет


др Драган Бојић, ванредни професор
Универзитет у Београду – Електротехнички факултет


др Леонид Стоименов, редовни професор
Универзитет у Нишу – Електронски факултет


др Марија Рашајски, ванредни професор
Универзитет у Београду – Електротехнички факултет