DBBT
Digital Broadcasting & Broadband Technologies

Co-funded by the Erasmus+ Programme of the European Union

Metode istraživanja

Profesor dr Miroslav Lutovac

"This project has been funded with support from the European Commission. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein"

Autori

- Ko je sve autor?
- Koliko autora?
- Zahvalnica?
- Tehnička obrada, prevod, ….
- Ko je vlasnik rezultata, ko finansira, copyright?
- Urednici, pristrasnost, naklonost, favoriti?
- Recenzenti – ne odavati pre objavljivanja, konflikt interesa?

A contribution to graphlet counting using GPU

Plagijarizam

Abstract — **Graphlets are used for different kinds of analyses in social networks, bioinformatics and other areas described by large networks. They are small non-isomorphic connected subgraphs, where their number can provide a characterization of the network properties. Much of existing methods for counting the graphlets are based on direct enumeration. However, in case of large networks, this type of counting becomes computationally very demanding. Fortunately, for very sparse networks the computational cost is much less prohibitive than in dense networks, leading to the more efficient graphlet counting algorithms. This work is an algorithm.**

[2] O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes, and N. Przulj, "Topological network alignment uncovers biological function and phylogeny," *Journal of The Royal Society Interface*,

I. Introduction

NETWORK is a way of representing some set of objects, and some relationships between the objects. Networks can be applied to a wide range of domains. Technological, social, information, and biological networks are all frequently studied, and other applications are possible as

[17] C. E. Tsourakakis, "Counting triangles in real-world networks using projections," *Knowledge and Information Systems*, vol. 26, no. 3, pp. 501–520, Mar. 201

number of appearances of graphlets in the network provides a description of the network's structural properties. On a local level, counting how many times a particular node participates in each kind of graphlet induced in the network gives a topological signature of the node's neighborhood represented as a 30-dimensional vector.

We need here some graphlet definitions here, so we follow [4]. Assume, $G(V, E)$ is a graph, then V is the set of vertices and E is the set of edges. Each edge $e \in E$ can be represented by a pair of vertices (v_i, v_j) where, $v_i, v_j \in V$. A graph is called simple, if it does not contain a self loop, and at most one edge exists between two of its vertices. We consider simple, connected, and undirected graphs. A graph $G' = (V', E')$ is a subgraph of G if $V' \subseteq V$ and $E' \subseteq E$. A graph $G' = (V', E')$ is a vertex-induced subgraph of G if $V' \subseteq V$ and $E' \subseteq E$ and $\{e = (v_a, v_b) : v_a, v_b \in V', e \in E, e \notin E'\} = \emptyset$. A vertex-induced subgraph is a subset of the vertices of a graph G together with any edges whose both endpoints are in this subset. We will refer to vertex-induced subgraph as "induced subgraph". Two graphs G and G' are isomorphic, denoted by $G \cong G'$, if there exists a structure-preserving (both adjacency and non-adjacency

ORIGINALITY REPORT

43%

SIMILARITY INDEX

Plagijarizam

PRIMARY SOURCES

1	Hocevar, T., and J. Demsar. "A combinatorial approach to graphlet counting", Bioinformatics, 2014. CrossCheck	461 words — 14%
2	Bhuiyan, Mansurul A., Mahmudur Rahman, Mahmuda Rahman, and Mohammad Al Hasan. "GUISE: Uniform Sampling of Graphlets for Large Graph Analysis", 2012 IEEE 12th International Conference on Data Mining, 2012. CrossCheck	284 words — 9%
3	Brahim Betkaoui. "A framework for FPGA acceleration of large graph problems: Graphlet counting case study", 2011 International Conference on Field-Programmable Technology, 12/2011 CrossCheck	86 words — 3%

Tema za rad

a) genetski algoritmi
b) neuronske mreže
c) sistemi zasnovani na kompjuterskom rezonovanju
d) Data Mining
e) Association Rules Using Ontologies
f) Rule Schema Formalism

Tema za rad

- korišćenje postojećih softverskih rešenja za bolje planiranje
  - In **Semantic Web** field, **ontology** is the most appropriate representation to **express** the complexity of **user knowledge**, & several specification languages
  - The need of utilizing **appropriate tools** like **design for environment**, **life cycle assessment** and other environmentally sound practices
  - **Sustainability** and **environmental responsibility** is increasingly important in manufacturing and design and are likely to **influence the main priorities for advancing manufacturing** operations and technologies
  - collaboration to improve understanding and to enhance technology transfer and applications of sustainability

Tema za rad

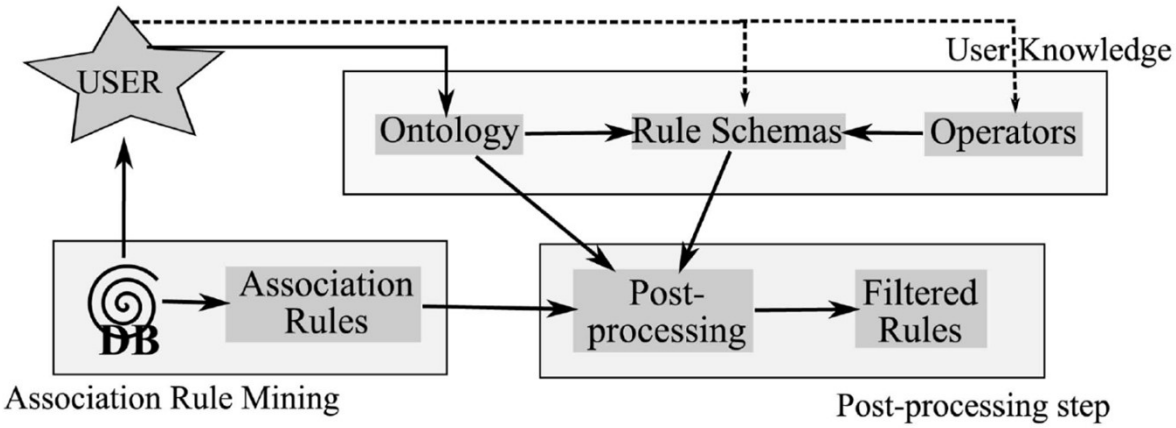
- razvoj algoritama i alata za planiranje
- razvoj tehnika za bolje korišćenje kompjuterske snage ili alata
- napredne tehnike i paradigme korišćenjem računarskih algebarskih sistema

Tema za rad

- korišćenje viših programskih jezika za
  - Modelovanje
  - Optimizaciju
  - Vizuelizaciju (3G)
- delova proizvodnog proces i integracija sa mernim i proizvodnim sistemima

IEEE transactions on knowledge and data engineering

- Knowledge-Based
- Interactive
- Postmining
- of Association Rules
- Using Ontologies



Resource Description Framework

- RDF is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model.
- It has come to be used as a general method for **conceptual description** or **modeling** of information that is **implemented in web** resources, using a variety of syntax notations and data serialization formats.

Semantic Web

- A collaborative movement led by international standards body the World Wide Web Consortium
- The standard promotes **common data formats** on the World Wide Web
- By encouraging the inclusion of semantic content in web pages, the Semantic Web aims at converting the current web, dominated by **unstructured and semi-structured** documents **into** a "**web of data**".

Ontology Web Language

- A family of knowledge representation languages for authoring ontologies or knowledge bases
- The languages are characterized by formal semantics and RDF/XML-based serializations for the Semantic Web
- OWL is endorsed by the World Wide Web Consortium (W3C) and has attracted academic, medical and commercial interest

Knowledge-Based Interactive Postmining of Association Rules Using Ontologies

- In Big Data, the usefulness of association **rules** is strongly limited by the huge amount of delivered rules
- To overcome drawback, proposed
  - Concise representations (**modelovanje podataka**)
  - Redundancy reduction (**minimizacija podataka**)
  - Postprocessings (**obrada van realnog vremena**)

Knowledge-Based Interactive Postmining of Association Rules Using Ontologies

- Methods generally based on **statistical** information **do not guarantee** that the extracted rules are interesting for the user
- crucial to help the decision-maker with an efficient **postprocessing** step in order to **reduce** the number of **rules**
- **interactive approach** to reduce and filter discovered rules

Metode istraživanja

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS **V. Milutinovic**

- Selection of the title
  - **enables an expert to figure out the essence of the basic idea and the main contribution even without reading the paper**
  - induces the reader to think deeply over the "philosophy" of the contribution described in the paper

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- Structure of the **abstract**
  1. Problem statement of the research under consideration
  2. A short list of existing solutions and what is their drawback, from the point of view of the above defined problem statement
  3. Essence of the proposed solution, and why it is expected to be better under the same conditions
  4. What type of analysis was done to show that the proposed solution is really better than any of the existing ones, from both the performance and the complexity points of view
  5. What are the major numerical highlights of the analysis
- 50-word abstract - one sentence long
- 500-word abstract - 10 sentences long

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- Structure of the **figures/tables** and the related captions
  - All details must be **clearly visible**
  - Reading only the **figure captions** of the paper can substitute the **first rough reading** of the entire paper
  - captions should include only the facts which are "local" to the figure/table, and these facts should **never be repeated again in the main body of the paper**

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- Structure of the written paper
  - a) Introduction, to include the basic facts needed to tune the reader to the paper
  - (b) Problem statement, to define precisely the problem being attacked by the research under consideration, and why is that problem important

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- Structure of the written paper
  - (c) **Existing solutions and their criticism**, to survey briefly the major existing solutions **form the open literature** and to underline their deficiencies from the point **of view of interest for this research**, which is defined in the above mentioned problem statement section;

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- Structure of the written paper
  - (d) **Proposed solution and why is it expected to be better**, to give the **essence of the proposed solution** (i.e., the essence of the idea which is to be introduced), followed by a logical/phylosophical discussion about the expected benefits stemming from the idea;
  - (e) **Conditions and assumptions** of the research to follow, to **summarize the environment of interest**.

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- **Analytical** analysis to show one or more of the following
  - **proof of validity of the major idea** of the paper
  - **calculation of initial values for simulation** analysis to follow
  - **rough estimation of the performance**;
  - **rough estimation of the complexity**;
  - **something else which is relevant**

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- **Analytical analysis**
  - **Simulation** analysis to show performance
  - **Implementation** analysis to show complexity
- Develop the skeleton on the paragraph level
  - defining all **subtitles** on lower levels
  - Specify the **first sentence of each paragraph**, which is the major one; other sentences of each paragraph are just to **explain and/or justify the statement** conveyed by the first sentence

THE BEST METHOD FOR PRESENTATION OF RESEARCH RESULTS V. Milutinovic

- Conclusion
  - **revisiting the major contribution** from the **performance/complexity** point of view;
  - stating **who will benefit** from the presented results;
  - stating what are the **newly open problems** and research avenues.
- keep in mind that some people read only the abstract and the conclusion

Razlog odbijanja rada

- The paper has 29% similarity with reference [2] and similar journal paper. Authors has mentioned the reference (“final key stream to access the main messages [2]”) without explanation of the difference between [2] and this paper. It seems that the contribution is not sufficient in this paper, and probably in the next paper they can provide adequate contribution (“the proposed algorithms may be applied on xxx bytes by using parallel processors”)

Rad ne odgovara nivou rada za konferenciju
How confident is reviewer ... 5=highest

Razlog odbijanja rada

- Rad ne odgovara nivou rada za konferenciju.
- Autori su pokazali nedovoljno poznavanje principa u oblasti koja je prikazana u radu. U svakom smislu rad ne zadovoljava potrebne kriterijume za konferencijski rad, ni u tehničkom ni u suštinskom smilu.
- ... ne bez ozbiljne prerade u osnovnom konceptu, formi i detaljima.
- autori nisu u dovoljnoj meri familijarni sa materijom koju obrađuju.
- Proračun koji autori prikazuju zasniva se na elementarnoj teoriji ... odgovara ispitnom zadatku ...
- 1 strana zauzima prikaz elementarne teorije
- ... umesto reči „pojačavač“ autori koriste „pojačalo“ ...
- Naslov rada „xxx“ ne odgovara sadržaju ... analizirana samo 1 platforma
- aspekt korišćenja softvera nije objašnjen
- Rad je u grafičkom i jezičkom smislu neuredan

Bez naučne podloge, popularno,

Razlog odbijanja rada

- The main weakness of this manuscript is that it is more suitable for the popular magazine than for the science conference
- The manuscript lacks in critical evaluation, and some parts of the evaluation are based on the authors experience not on the measurements that authors performed.
- It is not clear how the author selected the features for the evaluation among all the features

too simple hypotheses, poziv na nepostojece slike?

Razlog odbijanja rada

- The paper has two hypotheses that are too simple and the conclusion is expected (positive correlation between the developer's team size, and the total number of committers)
- The format of the paper is not as required by the conference (see the template document)
- Paper suffers from some serious issues regarding its clarity and the choice of hypotheses it consider
- … make several references to tables and figures which are not even included in the document

Bez validacije, upoređenja
Nepoštovanje recenzenta, nije uradio korekcije rada

Razlog odbijanja rada

- Ovaj rad je u identičnom obliku bio poslat na konferenciju XXX 2017 gde sam radio recenziju, pa sam i ja sada poslao identičnu recenziju
- There is a complete lack of any experimental validation and comparison of supposed enhancements to the previous state of art. The authors admit that they have done all the work without appropriate XX hardware and leave performance evaluation of their approach for the future work.

The paper lacks originality and scientific novelty.

Puno grešaka, nejasno,

Razlog odbijanja rada

- The paper seems to be unfinished, with number of errors. For instance, in Fig 1, the simple calculation is wrong, even the section numbers are wrong.
- Written English is poor, many sentences are too long and unclear.
- Maybe this is the reason why the authors contribution is not clear at the end.

Nejasan doprinos,

Razlog odbijanja rada

- It is unclear what is the contribution of the research in this paper
- Different well-known computer xxx algorithms were enabled and tested in XXX environment
- No comparison is given in the paper
- There are lack of references that deal with similar problems

Koja je svrha rada, nema pregled drugih rešenja

Razlog odbijanja rada

- It is not clearly explained what is the contribution of the paper
- What is a purpose of proposed simulation?
- Review of other methods for this or similar problems is not presented
- Acronyms ... has to be explained
- Origin of input data set is unknown

Stare reference, nevidljiv doprinos iz rada

Razlog odbijanja rada

- Paper presents the results of simulation using XXX tool for measuring the performance of zzz.
- The proposed algorithm (that can be the authors new idea) is explained using text from references [1, 4, 2, 8] known for more than 10 years
- Almost the whole first and second pages are used from known published works
- The contributions of simulation and presented performances in table form are not sufficient for accepting for this conference

Bez slika! Autori nisu implenetirali – šta su radili?

Razlog odbijanja rada

- Your article has no figure. It is necessary to insert appropriate images that will closer to ilustrate your work
- It is not clear what is the contribution of this paper
- ... are also well known, even in xxx applications
- ... authors claim in the conclusion that they even did not implemented this xxx

Profesor dr Miroslav Lutovac
mlutovac@viser.edu.rs