Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Sergei Ivanov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Combinatorial and Neural Graph Vector Representations

Supervisor: Prof. Evgeny Burnaev

Name of the Reviewer: Andrzej Cichocki

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **31-10-2019** **Date: DD-MM-YYYY**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

Review of the PhD thesis of Mr Sergei Ivanov "Combinatorial and Neural Graph Vector Representations"

The topic of the thesis is strictly related to machine learning, graph theory and representation learning. The Author of the thesis developed new graph embedding methods that could be potentially efficient in several machine learning tasks. The developed embedding methods can be theoretically used in any graph related problem, where vector representations are useful, especially standard machine learning problems on graphs such as classification of graphs.

The research results of this thesis are interesting and important because they have many potential applications for some real-life problems like disease diagnostics or drug discovery. Graph representations are ubiquitous and can be seen as a way to model relationships in the complex multidimensional data. In fact, graph representations may lead to optimized or better solution for such problems that can be modeled with graphs.

This topic of research is recently quite hot and competitive and many machine learning methods on graphs were developed in the last decade, such as graph classification that can model many real-life applications such as drug discovery. Solving them with higher accuracy has been important and led to many new solutions in representing graphs in vector space. The topic has opened several quite efficient solutions that can be potentially used in practice, but the author the thesis searched even better and alternative solutions.

The author of the thesis carefully verified and validated his theoretical and experimental (by computer simulation) results by thoroughly evaluating performance his method for real-world datasets and compared them with state of the art methods. For theoretical parts, he provide rigorous proofs that have been assessed and verified by the reviewers of the top international conferences.

Moreover, all his papers are supported by software open-source the code and he provided data necessary to reproduce the experiments. The code is mostly written in Python and C++ and the code is available in github.

In my opinion the most important achievement and new results are related to use anonymous walks (AW) for graph representation. Anonymous walks were available before but they have not been used for representing graphs. The Author proposed and developed two original new schemes for using AW for graph representations. The theoretical developments of AW were also developed and proved. Furthermore, the author of the thesis investigated and formulated problem of product recommendation in graphs and proved its complexity, developed and studied theoretically new algorithm, compared experimentally with other related algorithms. The most original and interesting idea is the development of anonymous walk embeddings.

The author of the thesis published his innovative results in international conferences and related workshops, including top conferences in AI and ML, especially ICML, ACM SIGR and ICDM. The list and quality of publications are quite impressive, although so far no any work has been

published in any high impact factor Journal. Furthermore, the author have ongoing research work on graph embedding, which was not included to the thesis, because it was not strictly connected to the main topic of the thesis directly and another his paper is currently under review for prestigious conference ICLR19 (https://openreview.net/pdf?id=rJlUhhVYvS).

Critical comments and future research directions. None of the graph embedding can be a silver-bullet for all possible graph problems. The developed by the Author of the thesis graph embedding for graph classification method showing that it can be potentially useful to solve other problems such as node classification and in applications to medical diagnostics. However, these potential applications were not yet fully explored. For each new application and new graph problem, corresponding embedding methods should be extensive tested and compared with other known graph representation methods. For the future work, I think, developing graph embedding method end-to-end for reinforcement learning framework that operates on graph combinatorial problem would be a very important and perspective topic of research, which could be applied to solve many problems without the hard-to-get labeled data.

In conclusion, Sergei Ivanov performed the following significant and innovative tasks:

- Proposed two new schemes to use anonymous walks to learn combinatorial and neural graph embedding.
- Provided theoretical background of the embedding such as sampling complexity, bounds or the number of anonymous walks.
- Performed evaluation on commonly used data sets and showed empirical superiority of his method in comparison to state-of-the-art solutions in terms of the accuracy.
- Developed the idea of anonymous experiments to the context of graph representation.
- Collected the data and performed evaluation tests for medical diagnostics using fMRI images.
- Formulated a combinatorial optimization problem of product recommendation in networks.
- Proposed a solution that allows to combine his neural embeddings for recommendation task in graphs.

Provisional Recommendation

x ☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*