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: Assist. Prof. Dr. Alexander Panchenko, Skoltech

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 02-10-2019**

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

Ph.D. thesis of Sergei Ivanov deals with the representation learning problem on graphs. Graph representation leaning is a very important field of study in machine learning today due to the fact that data represented in the forms of networks are ubiquitous in nature and technics, spanning from networks of roads to networks of co-occurrences of linguistic items. Analysis of such networks has been in focus of the network science for decades even centuries with multiple real-world applications in almost all fields. More recently, with the rise of the popularity of neural networks, learning a dense representation of symbolic structures became an important task, as sparse graph-based representations are not easily usable by the modern deep learning architectures.

This Ph.D. thesis contains several new approaches to learning of graph embeddings: representations of nodes and entire graphs possess important theoretical properties and show good results in many real-world problems such as graph classification problem or influence maximization. The thesis has three main parts. In the first section, the author provides the necessary background and motivation for his next sections. In the second, an extensive review of graph representation methods is provided: neural and combinatorial embeddings are analyzed. Then a new type of embeddings is introduced based on anonymous walks, which form the central contribution of this thesis (and is based on the ICML publication of the author). The author proposed approaches to learning of graph embedding based on both combinatorial and neural paradigms. The latter is analogous to paragraph vectors used in the natural language processing. In addition to intrinsic experiments, the author shows that the proposed graph embeddings are useful in node/graph classification problem, which he tests in data sets coming from chemical engineering and neuroscience (fMRI data classification). In the third section, the author formulates and provides solutions to the problem of product recommendation on graphs. He shows that the problem is NP-hard and there are several solutions that he proposes that solve this problem efficiently in real-world settings.

The high quality of the produced work is confirmed by the fact that the results of his thesis were published in the top conferences such as ICML, SIGIR, and ICDM which are extremely selective. This showed that the international data mining/machine learning community accepted these contributions to be novel and important to share. Recently, I reviewed Ph.D. theses from Computer Science faculties of HSE and the University of Rome and should say that this dissertation meets the highest international standards. Besides, another strong point of this thesis is the author verified his results in real applications such as medical diagnostics based on fMRI images and recommendations of groups to social network users.

On the potentially low side of this thesis may be short writing. I would be happy to see more examples and more discussion of the results. However, despite this limitation, overall, the thesis is well written and presents novel and significant contributions to the field of graph representation learning. I recommend accepting the thesis.

Provisional Recommendation

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