Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

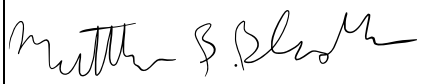
Name of Candidate: Sergei Ivanov

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: Matthew Blaschko

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 02-11-2019**

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

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

Thesis quality and overall structure of the dissertation:

This dissertation comprises four chapters. The first chapter motivates the study of graph representations from the perspective of real-world applications, and subsequently develops main results related to the graph isomorphism problem. These definitions are largely used in the subsequent two methodological chapters in which the main contributions are presented. Connections are made between the complexity theory of graph isomorphism and the algorithmic solutions that are applied in practice.

Chapter 2 is the first main methodological chapter, and in my assessment some of the strongest theoretical contributions contained within the thesis. Based on the result that anonymous walks representations are sufficient to recover graph topology [53] and therefore can be used for graph isomorphism (Section 2.4.1), these are used to build practical representations of graphs that are then used for other tasks. Mr. Ivanov stated the idea quite nicely in his own words:

> However, at least we are interested in the case of data-generating processes that assign the same labels to topologically same graphs, in which case our embeddings should be the same for isomorphic graphs. This leads to the idea that vectors that can recover isomorphism of two graphs are powerful enough to work well in many practical applications, of which graph classification is a prominent example. (p. 27)

Up to Section 2.4, previous results in graph classification are presented, though these could equally have been considered for Chapter 1. Section 2.4 closely follows the presentation in the original ICML paper and is well organized. Section 2.4.1 presents much deeper results about the relationship between the anonymous walks representation and graph isomorphism, and did not appear in the ICML 2018 paper. Section 2.5 presents the combinatorial and related neural models, while Sections 2.6 and 2.7 show empirical performance on benchmark and neuro-imaging datasets, respectively.

Chapter 3 marks a move away from graph isomorphism and graph classification strategies, and towards problems related to influence maximization. The classical influence maximization algorithm is submodular, and therefore a simple greedy approach gives optimal approximation guarantees. By contrast, Mr. Ivanov presents in Section 3.2 the CAIM problem (which presumably stands for "Content Aware Influence Maximization"), in which user attributes associated to each node in a graph must be selected to maximize influence. He subsequently shows that the resulting problem is neither submodular nor supermodular, indicating that previous greedy strategies used for submodular problems are inapplicable. Furthermore, he shows in Section 3.2.2 that the problem is not only NP-hard, but NP-hard to approximate within a factor n^(1-eps) for any eps>0. This motivates a heuristic but well motivated algorithm in Section 3.2.3 which is subsequently benchmarked empirically. Section 3.3 explores a different problem of influencer recommendation in social networks, proposing a classification based embedding algorithm in Sec 3.3.2.

Chapter 4 summarizes the main contributions of the work and outlines areas of future research. Mr. Ivanov highlights the use of pretrained graph models, and of learning based strategies for solving NP-hard problems.

Overall, the work is reasonably structured, though Mr. Ivanov has opted for a small number of relatively long chapters. Additionally, a number of typos and language issues are prevalent throughout the thesis, and in particular in Chapters 1 and 2. Special attention should be paid to spelling (please run English spell-check software) and the use of articles (a, the, pluralization).

The relevance of the topic of dissertation work to its actual content

The content of the dissertation matches well to the topic. Mr. Ivanov has done an excellent job of connecting theoretical contributions in computer science to a series of real world problems related to graph representations.

The relevance of the methods used in the dissertation

The dissertation excellently combined a strong theoretical approach, thorough attention to the relevant literature, and integration into deep learning models which frequently define the state of the art in machine learning today. In particular, Mr. Ivanov has done a very nice job of making connections to the classical graph theory literature as well as kernel methods in developments leading to deep learning models.

The scientific significance of the results obtained and their compliance with the international level and current state of the art

The results in this thesis compares favorably to international norms and state of the art. This is evidenced also by the publication quality (see below).

The relevance of the obtained results to applications (if applicable)

Machine learning methods on graphs are of fundamental importance. Mr. Ivanov has crafted a strong strategy of focusing on theoretical methods for graph isomorphism in developing subsequent methods. As such, the methods are immediately relevant to a huge number of problems that rely on graph representations. Similarly in Chapter 3, the relevance of influence in networks is well established and highlighted. These relationships are explicitly stated in the thesis, in the first chapter as well as methodological chapters.

The quality of publications

In my estimation, the work at ICML 2018 represents some of the highest quality contributions of the thesis. All together, the doctoral research has resulted in 5 strong publications:

1. Ivanov S, Karras P. Harvester: Influence optimization in symmetric interaction networks. In2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA) pp. 61-70. IEEE, 2016.

2. Ivanov S, Theocharidis K, Terrovitis M, and Karras P. ”Content recommendation for viral social influence.” In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 565-574. ACM, 2017.

3. Ivanov S, and Burnaev E. ”Anonymous Walk Embeddings.” In International Con- ference on Machine Learning, pp. 2191-2200, 2018.

4. Sharaev M, Artemov A, Kondrateva E, Ivanov S, Sushchinskaya S, Bernstein A, Cichocki A, Burnaev E. Learning connectivity patterns via graph kernels for fmri- based depression diagnostics. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) pp. 308-314. IEEE, 2018 .

5. Ivanov S, Durasov N, Burnaev E. Learning node embeddings for influence set completion. In2018 IEEE International Conference on Data Mining Workshops (ICDMW) pp. 1034-1037, IEEE, 2018.

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*