Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Saeed Osat

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Percolation on complex networks and its applications

Supervisor: Assistant Professor Vladimir Palyulin

Name of the Reviewer: Assistant Professor Maksim Kitsak

I confirm the absence of any conflict of interest	[signature] Date: 22-09-2021

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

Report on the Ph.D. thesis titled "PERCOLATION ON COMPLEX NETWORKS AND ITS APPLICATIONS" by Saeed Osat.

The present thesis is based on seven original research works that Saeed published during his Ph.D. studies. The thesis discusses the percolation properties of real and synthetic networks and their applications to neuronal avalanches, epidemic spreading, diffusion, and routing problems.

The thesis is well-written and is in the coherent academic treatise format. The thesis' structure is adequately summarized in the Table of Contents, page vi of the document. For brevity, I refrain from typing the detailed summary of the thesis and instead focus on methods, results, and the impact of Saeed's works that are presented herein.

It is well-accepted that big data often comes in the form of networks or graphs. One example is a social network, where individuals are abstracted as nodes or vertices, and interactions between individuals are abstracted as links or edges. A relatively new domain in network and data sciences is that of multiplex networks. Interactions comprising a network can be of different types. A colloquial example, to this end, is a transportation network where nodes are regions of interest and links correspond to different transportation modes (bus, train, airline, personal auto) connecting them.

The core of Saeed's Ph.D. work is the development of novel percolation techniques for the analysis of large multiplex networked systems. Historically, the term percolation originated in Polymer physics, where it was introduced to study the properties of porous materials. Percolation has been reinvented in studies of large networks, where it has been extensively used in (cyber)security, epidemic spreading, network control, and communication applications. In most general terms, network percolation refers to the analysis of the network under the removal or addition of its elements (nodes or links). In this Ph.D. research, Saeed has proposed novel percolation techniques and applied them to multiplex networks.

K-core percolation has been found to be an instrumental tool in a variety of applications, ranging from data visualization to identification of influential spreaders. In his work published in *Physical Review E* (publication 4, page iv and Chapter 3 of the thesis), Saeed extended the notion of k-core percolation to Gk-core and developed mathematical models to study them in random mono-layer networks. Gk-core can be viewed as the generalized measure quantifying the robustness of networks to various types of adverse events.

In his related work (publication 6, page iv, and Chapter 7 of the thesis), Saeed has extended the notion of k-core percolation from conventional to multiplex networks. His theoretical and computational methods developed in this work set the foundation for the mathematical analysis of multiplex networks. In particular, Saeed was able to explain the roles of node connectivity patterns and correlations between network layers on the resulting structure of k-cores. These results, published in *Physical Review Research*, are of great importance for epidemic spreading and peer influence studies.

Of central importance to both epidemic studies and studies of social influence is the optimal attack problem, where one is interested in finding the minimal set of nodes or links, that if removed, will lead to the collapse of a network into fragmented pieces. The optimal attack problem is of direct relevance to the optimal vaccination/immunization problem, where one is interested in vaccinating the population with limited resources to achieve the maximal suppression of the viral process. In a seminal work published in the prestigious *Nature Communications* journal (publication 1 page iv and Chapter 6 of the thesis), Saeed has developed the generalization of optimal percolation problem to multiplex systems, demonstrating that "multiplex networks have considerably smaller sets optimal sets compared to their single-layer based network representations.

Arguably, the two state-of-the-art models of real networks are the Stochastic Block Model (SBM) and the Random Hyperbolic Graph (RHG). While introduced with different applications in mind, both the SBM and the RHG models have become popular modeling tools due to their ability to reproduce complex features of real networks, including heterogeneous degree distributions, topological clustering, and community structure. In his recent work (publication 3, page iv and Chapter 5 of the thesis), Saeed demonstrates that there is a deep connection between the two. His results, published in *Physical Review Letters*, demonstrate that the "community structure of a network can be used as a coarse version of its embedding in a hidden space with hyperbolic geometry." This finding is extremely important as it connects two challenging data science problems: community finding and geometric representation of large networks.

Network observability is the problem of measuring network state using a minimal number of sensors. In his *Physica A* publication (publication 2, page iv and Chapter 4 of the thesis), Saeed managed to extend the notion of network observability to 2-layer multiplex networks and developed a mathematical framework for the analysis of the macroscopic cluster of mutually observable nodes. These findings are of direct relevance to the control theory of networked systems.

While equally important in practice, negative results are often under-reported compared to positive results due to the perceived lack of impact. One of Saeed's works (publication 5, page iv and Chapter 2 of the thesis) can be regarded as a negative result. One of the common beliefs is that the appearance of power-law avalanches in the dynamics of neuronal networks is the signature of criticality: neuronal networks are believed to exist in a "critical" state, in between a phase where activity rapidly dies out and a phase where activity is amplified over time. Together with collaborators, Saeed has disproved this claim by introducing a counter-example. In a simple avalanche model, Saeed demonstrated that power-law distributions for neuronal avalanches might emerge even when the system is not in its critical regime.

Last but not least is Saeed's study of super-diffusion in multiplex networks (publication 5, page iv of the thesis). This work demonstrates that Machine Learning can classify and predict the presence of super-diffusion in multiplex networks with high accuracy.

In relation to Saeed's thesis, I only have several clarifying optional remarks that in no way diminish the quality of Saeed's work. I do, nevertheless, believe that addressing my comments might improve the accessibility of the thesis to non-specialists.

1. On page 4, k-core is defined as "A maximal subset of nodes where each node has at least degree k." An inexperienced reader may interpret k-core as a set of nodes with degrees larger or equal than k. To avoid such misconception, I recommend adding "at least degree k within the subset."
2. Please define phases, phase transitions, and their types before discussing "Percolation on monoplex networks," page 14
3. Likewise, I recommend creating a dedicated section discussing the basic formalism of generating functions in application to computing network's connected component.

4. The thesis routinely operates with the notion of the Giant Connected Component (GCC). GCCs are often confused with Largest Connected Components (LCC)s. To avoid this, I recommend carefully defining both terms in the thesis.
5. Please include the definition of the excess degree in Table 1.1
6. Please define the susceptibility for Fig. 1.2(c). Explain the role of susceptibility in phase transition studies on complex networks, page 17.
7. Clarify the "theoretical prediction" for Fig. 1.3
8. Add a 1-2 sentence clarification for the analogy between the ICM and bond percolation.
9. Consider adding a dedicated section, defining ordinary percolation, k-core percolation, core percolation, and Gk-core percolation along with the algorithms used to implement them on networks.
10. Network observability needs to be defined prior to page 27. What are the different phases for observability? Please clarify on Page 27. Related to It, define the quantity plotted in Fig. 1.8

In summary, Saeed published seven works in top physics and interdisciplinary journals, resulting in more than 100 citations per Google Scholar in a period of fewer than five years. His research impact is significantly higher than expected at his academic age. I would like to emphasize Saeed's communication and collaboration skills, which are remarkable for his academic age: his publications are the results of 3 distinct long-distance scientific collaborations with teams of top experts in network science from Cyprus, Portugal, and the United States. I am fully satisfied with the quality of his research. I do recommend his thesis for the defense.

Provisional Recommendation

☒ *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*