

REFeree REPORT
on doctoral thesis of mgr inż. Ruby Khan
„Development of Methods for Identifying Key Variables in Complex Mathematical
Models of Biological Systems”

Supervisor: dr hab. inż. Krzysztof Puszyński, prof. Pol. Śl.

I. Scientific issues and subject of the dissertation

The impressive progress in biological sciences, which has been observed during the last three decades, was and still is possible mainly due to the application of appropriate mathematical and computational methods. One of the reasons for such a situation is the emergence of large biological data sets, which was initiated by a dynamic development of DNA sequencing methods in the mid-1990s. In practice, an analysis of such data is impossible without the application of algorithmic methods. Another reason is that living organisms are composed of complex systems appearing at various levels of their organization (organs, tissues, cells, etc.). Mutual interactions among components of these systems are crucial for their functioning. Hence, the knowledge about these interactions seems necessary for a deep understanding of the rules on which the functioning of the living world is based. However, these interactions are very complex; hence, an analysis of them is impossible in practice without the use of appropriate mathematical and computational methods. These observations became an impulse to apply methods from systems sciences in the area of complex biological phenomena analysis and the development of systems biology, a branch of science evolving on the border between biology, mathematics, computer science, and related sciences, where biological systems are investigated. It is worth noting that these systems have their own specificity, hence a direct application of methods initially developed for analyzing systems of other types may be insufficient in many cases. Therefore, there is a need to design new methods that take into account this specificity. A construction of such methods and their application is an intensively developed research area.

One of the many areas where systems methods are especially important is the study of processes related to the formation and development of cancer. It is well known that cancers are very complex diseases, i.e., many factors interact with each other and with the environment in which the cancer develops. Obviously, it makes the problem of designing anticancer therapies very difficult. It requires a thorough understanding of cancer processes, which is a significant challenge due to the high complexity of these processes.

Mrs. Ruby Khan, in her doctoral thesis, considers the problem of designing methods for an analysis of mathematical models of complex biological systems. Such models are the basis of an analysis conducted using formal methods. Due to the nature of interactions between elements of biological systems, graph and network-based models are especially useful. In her

dissertation, the doctoral candidate proposes a set of methods that can be used to analyze selected parameters of such models and to draw conclusions concerning properties of analyzed biological systems. The proposed set of methods was tested on data concerning, among others, ovarian cancer. The obtained results indicate that these methods make it possible to discover key elements of biological systems.

II. Analysis of the content of the thesis and the obtained results

1. The content of the thesis

The doctoral dissertation of Mrs. Ruby Khan is written in English. It consists of five chapters and a bibliography preceded by an abstract written in English and Polish, a table of contents, and a list of the author's publications and conference presentations. At the end of the dissertation, appendices are included.

The first chapter contains an introduction to the topic of the dissertation, a formulation of the research goal and hypotheses, and a very concise presentation of the methodology used. The main goal formulated by the doctoral candidate is to develop and validate innovative computational techniques for identifying key variables in complex biological systems.

The second chapter presents a literature review and a history of applying mathematical methods to complex biological systems modeling and analysis. Selected issues related to the analysis of such systems are briefly discussed here, and methods used in the research described in the thesis are also indicated.

The third chapter is devoted to the methodology used by the doctoral candidate. Here, a structure of the proposed computational pipeline is described, and the components of the pipeline are presented, i.e., acquisition and integration of biological data necessary to build and analyze the models of the studied biological systems, database implementation, network-based model construction, PageRank and random walk algorithms, Boolean modeling, and recurrent convolutional neural networks. Moreover, an application of these components to the modeling and analysis of the studied biological system is outlined.

This chapter is composed of many short subsections, whose content in many cases is of a technical nature.

The fourth chapter presents the obtained results. Here, methods used for the acquisition of biological data from various databases, the implementation of the database used for their storage, and the construction of the network model are discussed. Moreover, the results of the analysis of structural properties of the model, visualizations of the network, the analysis conducted using PageRank and random walk algorithms, and the results of Boolean modeling are presented. Furthermore, in this chapter, the results of additional experiments are also discussed. These experiments aimed to confirm that the developed computational pipeline can be used not only for the analysis of processes related to ovarian cancer but also to study various biological processes. The doctoral candidate developed and analyzed network models of the cell cycle and the MAPK signaling pathway. The results of these analyses are presented in this chapter.

Similar to chapter three, this chapter is also composed of several very short subsections, many of which have a nature similar to a technical report.

The fifth chapter contains a discussion of the obtained results and conclusions. Here, Mrs. Ruby Khan discussed, among others, the consistency of the obtained results with the knowledge about the analyzed processes available in the literature and indicated the achievement of the goal of the thesis as well as a positive verification of the research

hypotheses.

The dissertation ends with a bibliography consisting of 130 items, appendices containing program codes written in Python and additional tables.

2. Substantive and editorial comments

The doctoral thesis of Mrs. Ruby Khan is interesting and concerns an important issue from both theoretical and practical points of view. As was mentioned earlier, the analysis of complex biological systems is a very important research problem, and more effective methods for such analysis are still needed. It is worth mentioning that the results of studying such systems can have practical applications in many areas, of which medicine is one of the most important. Hence, there should be no doubt that the doctoral candidate has chosen a very important research problem and set an important goal for her thesis.

The title of Mrs. Ruby Khan's doctoral thesis reflects its content well, as she developed a method in the form of a computational pipeline for the analysis of network-based models of biological systems. In particular, this method can determine some essential properties of such models.

In general, the structure of the dissertation is correct, i.e., it contains an introduction, where the research problem is presented, and the goal as well as hypotheses are formulated. In the following chapters, a literature review, the methodology used, and the obtained results are presented. The dissertation ends with a chapter devoted to discussing the results and conclusions. However, dividing the third and fourth chapters into many short or very short subsections is a bit surprising and raises some doubts. The content of these subsections is, in many cases, very technical in nature, especially in the third chapter. Moreover, in these chapters, information is often presented in a very concise manner. It causes the main idea of the thesis to be lost somewhere in a thicket of many details, which seems unnecessary. In a doctoral dissertation, the most important are the main research ideas, not the technical details, so these ideas should be, above all, exposed. The structure of Mrs. Ruby Khan's dissertation does not facilitate the recognition and appreciation of these ideas.

The doctoral candidate properly uses the literature, however, the number of references to the literature on biological aspects of the modeled processes is a bit disappointing. This small number of such references follows from the fact that in the dissertation, there is no description of the biological processes that are modeled using the proposed method. This omission is somewhat surprising, since such a description seems to be natural and would allow for better understanding of the results of the methods used, especially in the area of modeling. If the doctoral candidate has chosen some biological processes to validate the proposed approach, these processes should be described in the dissertation, at least to some extent.

The method proposed by Mrs. Ruby Khan in her thesis is in the form of a computational pipeline. This pipeline is composed of known algorithms that are applied appropriately, which allowed the achievement of the goals of the thesis. Hence, it can be concluded that these algorithms were chosen and applied properly. The results obtained were compared with the knowledge concerning the analyzed biological processes in the literature. This comparison confirmed the effectiveness of the proposed method (the comparison is discussed in the fifth chapter of the thesis). It is worth emphasizing that in addition to testing the proposed approach on data related to ovarian cancer, which was the doctoral candidate's primary biological interest, the method was also tested on two additional biological processes, i.e., the cell cycle and the MAPK signaling pathway. In all cases, the tests' results confirmed the

method's usefulness. Although the obtained results do not contain new significant biological insights, they indicate the potential of the method to make biological discoveries. It means that the method could be used in research on various biological processes and systems, especially signalling and metabolic pathways.

In addition to the comments mentioned above, the dissertation contains several errors, omissions, or mistakes, many of them being editorial in nature. Among them, there are:

- In the dissertation, there are many references to program codes included in the Appendix. It seems that the intention of the author was that the codes should explain how various phases of the modeling and analysis are realized. However, these issues should be explicitly explained in the main text of the dissertation since not all aspects of the modeling and analysis are obvious, and the reader should not be expected to deduce them from the program codes.
- on page 16, at the end of the second paragraph, there is an unfinished sentence, „Using iterative modeling”;
- in the dissertation, there are no direct references to some figures and tables (by their numbers), e.g., to tables 3.1, 4.7, 4.10, 4.12, 4.13, 4.14, 4.20, 4.21, and figures 1 (this figure should be numbered 3.1), 4.7, 4.8, 4.12, 4.31, 4.32;
- the caption for Figure 4.1 should be below the figure, not above it;
- there are three captions for Figure 4.4, and it has three numbers, i.e., 4.2, 4.3, 4.4, despite that it is one figure;
- on page 65, there are incorrect references to Figures 1 and 2 – in the fourth chapter, there are no figures with such numbers – these references should concern the appropriate parts of Figure 4.4.;
- in the last line on page 71, there is the text „Listing 4.1: Network Filtering Implementation” – it is not known what it refers to;
- Figure 4.7 is unreadable and very difficult to interpret;
- on page 72, there is a formula (4.1) that is not commented on in the text, and it is not known what the meaning of the symbols in it is, nor how the formula should be interpreted;
- subsections 4.2.3 and 4.2.4 have the same title, and the latter contains only one sentence ended by a colon after which the following subsection begins;
- subsections 4.4.3 and 4.4.5 have the same title;
- subsection 4.4.4 is empty;
- there is no description of Figure 4.12 in the text – the reader has to guess on their own why it is included in the dissertation and what is presented in it;
- there is no discussion of the networks presented in Figures 4.13 and 4.14 – what are the differences between the network representations shown in these figures, what important information they provide, and how they are related to the networks presented in Figures 4.10 and 4.11;
- there are five captions for Figure 4.19, and it has five numbers (4.15 – 4.19) – if it is one figure, it should have one number and there should be one caption explaining the content of its individual parts, but if there are four separate figures, then they should not have a common number nor a common caption;
- the meaning of the curves in Figures 4.19, 4.23, and 4.24 should be explained (in each of these figures, there are two curves);
- in Figures 4.25 – 4.29, visualizations of the analyzed network obtained using various methods are shown, but there is no explanation what the differences between these visualizations are, what their characteristic features are and to what extent they proved to be useful for the conducted analyses;

- Figures 4.26 – 4.29 are unreadable;
- it is not known what the meaning of the yellow circles in Figure 4.29 is;
- In Table 4.18, there is information about three vertices, despite the fact that on page 90 it is written that in this table there should be information about all vertices;
- subsection 4.11 „Random Walk Analysis” should be included in section 4.10 „Results of Random Walk and Network Analysis”;
- in Table 4.20 the order of vertices Wip1, PIP3, and ATM is different than in Figure 4.31 (why?);
- it is not clear what the difference is between the information presented in the charts shown in Figures 4.33 and 4.34;
- values of coefficient C (equation (4.3)) for the vertex groups described in Table 4.22 are not provided – they could be included in this table;
- it is not clear what is shown in Figure 4.35;
- instead of TNFa, IKKKa, IKKa there should be TNF α and IKK α (in several places);
- in Figure 4.46, the normalized values of vertex visit counts are shown, but they are in the range from 0 to 12, so the question arises, how was the normalization performed? ;
- subsection 5.0.1 should have number 5.0;
- Appendix A is titled „Supplementary Tables”, but it contains primarily program codes, so the title does not reflect the content of the appendix;
- In the dissertation there is no comparison with other computational methods for similar analyses. Many such methods exist. Such a comparison would allow to evaluate better and appreciate the significance of the method proposed by the doctoral candidate.

The above comments do not significantly influence a positive assessment of the thesis. The doctoral candidate proposed a computational method that allows the development of models of complex biological systems and their analysis in a largely automated manner. This method can be a valuable tool for studying complex biological systems of various types, among others, those related to disease processes occurring in the human body. Therefore, it can be one of the tools used for drug or more effective therapies design for various diseases, e.g., cancers. Reading the dissertation leads to the conclusion that the doctoral candidate is very well versed in the field of methods for modeling and analysis of complex biological systems and that she possesses the ability to conduct research independently.

3. Summary

In my opinion, Mrs. Ruby Khan, in her doctoral thesis, presented interesting results concerning the modeling and analysis of complex biological systems. The proposed computational pipeline is an interesting combination of many known methods in a way that allows for conducting analyses of complex biological processes largely in an automated manner. This method is an essential contribution in the field of systems biology and can be a valuable tool for research on biological systems as well as a starting point for further development of methods used in this area. The doctoral candidate demonstrated both knowledge of biomedical engineering and the ability to conduct research independently. It is also worth noting that parts of the results of the doctoral candidate’s research related to the topic of her thesis have been published in two journals from the JCR list, i.e., *Genes* and *Scientific Reports*, which indicates that these results have been appreciated by the research community.

III. Conclusion

The doctoral dissertation of Mrs. Ruby Khan contains original and interesting scientific results concerning the methods for modeling and analysis of complex biological systems. In my opinion, these results make a significant contribution to the discipline of biomedical engineering. Moreover, in my opinion, the requirements for doctoral dissertations specified in Polish law (Ustawa z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce, Dz. U. z 2024 r. Poz. 1571, z późn. zm.) have been fulfilled. Therefore, I recommend the admission of Mrs. Ruby Khan to the next stages of the procedure for awarding the scientific degree of doctor.