

POLITECHNIKA POZNAŃSKA

Wydział Informatyki i Telekomunikacji

Instytut Informatyki



**METODY PORÓWNYWANIA MODELI  
ZŁOŻONYCH SYSTEMÓW BIOLOGICZNYCH  
WYRAŻONYCH ZA POMOCĄ SIECI PETRIEGO**

Bartłomiej Szawulak

Promotor: prof. dr hab. inż. Piotr Formanowicz

Poznań 2024

# Streszczenie

Biologia systemów reprezentuje szybko rozwijającą się dziedzinę nauki, korzystającą z jednoczesnego rozwoju biologii i informatyki. Obserwowany jest stały wzrost liczby modeli przedstawiających skomplikowane procesy biologiczne. Do przeprowadzenia ich analizy konieczne są dedykowane narzędzia i algorytmy. W zależności od formy reprezentacji matematycznej użytej dla danego modelu, dostępny będzie inny zbiór dostępnych metod analitycznych. W części przypadków możliwa jest transformacja modelu do innej postaci w celu przeprowadzenia dostępnej dla niej analizy. Niestety często jest to obciążone pewnym kosztem precyzji wyniku lub czasochłonności analizy. Powoduje to zapotrzebowanie na dedykowane metody dla danej reprezentacji matematycznej.

Problemem badawczym niniejszej rozprawy jest opracowanie metod służących do ustalania stopnia podobieństwa pomiędzy modelami systemów biologicznych wyrażonych za pomocą sieci Petriego. Sieci Petriego służą do modelowania równoległych procesów dyskretnych, które mogą rywalizować o dostęp do zasobów. W swojej klasycznej formie są one przedstawiane jako dwudzielny, skierowany graf ważony, w którym występują dwa rodzaje wierzchołków. W biologii systemowej sieci Petriego znajdują zastosowanie w modelowaniu sieci metabolicznych, ścieżek sygnałowych oraz innych złożonych systemów biologicznych. W dotychczasowych podejściach analizy modeli przedstawionych za pomocą sieci Petriego widoczny jest brak metod dokonujących oceny podobieństwa dwóch modeli. Zostało to zaobserwowane nie tylko dla systemów biologicznych, ale i w innych obszarach zastosowań sieci Petriego. Do momentu rozpoczęcia badań będących podstawą niniejszej rozprawy, opublikowana była tylko jedna metoda porównywania przepływów w sieciach metabolicznych wyrażonych za pomocą sieci Petriego. Wraz z rosnącą liczbą dostępnych modeli i technik transformacji grafów do postaci sieci Petriego, potrzeba porównywania tych modeli staje się coraz bardziej istotna. Dostęp do odpowiednich metod analizy był dotąd ograniczony.

Rozprawa koncentruje się na opracowaniu metod porównywania klasycznych sieci Petriego dedykowanych do modeli systemów biologicznych. W tym celu wykorzystano obszerną literaturę naukową zajmującą się problemem porównywania grafów i przeanalizowano dostępne w niej metody pod względem ich adaptacji do sieci Petriego, jak i zastosowania w kontekście biologicznym. Metody porównywania sieci Petriego przedstawione w ramach tej rozprawy przedstawiają zróżnicowane podejścia do oceny podobieństwa. Z tego powodu są one dedykowane dla sieci o różnych rozmiarach, strukturze, rozmiarze zbioru niezmienników, oraz kontekstu biologicznego. Metody oparte o zdefiniowane w ramach tej rozprawy wierzchołki rozgałęziające, oraz dedykowany do sieci Petriego wariant grafletów reprezentują podejścia stworzone do działania na dużych sieciach zawierających kilka tysięcy wierzchołków poprzez analizę rozkładu małych podsieci w nich występujących. Reprezentują one podejścia za pomocą których nie jest możliwe określenie izomorfizmu (identyczności strukturalnej), a jedynie ich przybliżonego podobieństwa. Podejściem, które pozwala na określenie izomorfizmu pomiędzy po-

równywanymi sieciami Petriego jest metoda bazująca na dekompozycji do podsieci wybranego typu, dla których wyznaczane są wspólne podsieci. Opracowano kilka algorytmów pozwalających na zastosowanie metody do różnych typów podsieci używanych w modelach biologicznych, jak np. podsieci indukowane przez zbiory MCT i conADT. Metoda ta przez parametryzację pozwala na modyfikację uzyskanej wspólnej podsieci do postaci pożądanej w danej aplikacji biologicznej. Przeprowadzono także modyfikacje metody porównywania modeli ścieżek metabolicznych wyrażonych za pomocą sieci Petriego przez dopasowanie występujących w nich t-niezmienników odpowiadających przepływowi. Dostosowaną ją do działania na t-komponentach, p-niezmiennikach i sub-, sur-niezmiennikach. Wyniki porównano ze zwracanymi przez oryginalny algorytm.

Dla modeli systemów biologicznych wyrażonych za pomocą sieci Petriego nie opracowano powszechnie uznanych norm reprezentacji modelowanych procesów. Z tego powodu szczególną uwagę poświęcono problemowi różnic mających swoje źródła w poziomie szczegółowości i stylach reprezentacji autorów porównywanych modeli.

Niniejsza praca stanowi wprowadzenie do dalszych badań nad porównywaniem złożonych modeli systemów biologicznych.

# Abstract

System biology represents an expanding field of science that benefits from the concurrent development of biology and computer science. There is a continuous increase in the number of models describing complex biological processes. Analyzing these models requires specialized tools and algorithms. Depending on the mathematical representation used for a given model, different analytical methods are available. In some cases, it is possible to transform the model into a different type to conduct available analyses, though this often comes with a downgrade of precision or increase of time complexity. This creates a demand for dedicated methods for specific mathematical representations.

The research problem addressed in this dissertation is the development of methods for determining the degree of similarity between biological system models represented using Petri nets. Petri nets are used to model concurrent discrete processes that may compete for access to resources. In their classical form, they are represented as a bipartite, directed weighted graph with two types of vertices.

In systems biology, Petri nets are applied to model metabolic networks, signaling pathways, and other complex biological systems. Previous approaches to analyzing Petri net models have lacked methods for assessing the similarity between two models. This gap has been observed not only in biological systems but also in other areas where Petri nets are applied. Before the start of the research that is a foundation of this dissertation, only one method for comparing flows in metabolic networks represented using Petri nets had been published. Given the continuously growing number of available models and the techniques for transforming graphs of various types into Petri net form, the ability to compare them is a desired analytical method that has previously been limited.

This dissertation focuses on developing methods for comparing classical Petri nets dedicated to biological system models. To achieve this, extensive scientific literature on graph comparison was reviewed, and available methods were analyzed for their adaptation to Petri nets and their application in a biological context. The Petri net comparison methods presented in this dissertation offer diverse approaches to assessing similarity. They are designed for networks of different sizes, structures, minimal invariant set sizes, and biological contexts. Methods based on branching vertices defined in this dissertation, as well as a Petri net variant of graphlets, represent approaches created to work with large networks containing several thousand vertices by analyzing the distribution of small subnets within them. These methods provide approximate similarity and can't confirm an isomorphism. An approach that allows for determining isomorphism between compared Petri nets is based on subnet decomposition of a chosen type, for which common subnets are identified. Several algorithms have been developed to apply this method to various types of subnets used in biological models, such as subnets induced by MCT and conADT sets. This method allows for parameterization to modify the resulting common subnet to the desired form for a given biological application. Additionally, modifications have been made to the method for comparing metabolic pathway models represented using

Petri nets by matching their t-invariants that correspond to metabolic flows. It has been adapted to work with t-components, p-invariants, and sub- and sur-invariants, and the results were compared with those from the original algorithm.

For biological system models represented using Petri nets, there are no widely accepted norms for the representation of modeled processes. Because of that, particular attention has been given to issues arising from differences in the level of detail and the representation styles of the authors of compared models.

This work serves as an introduction to further research on comparing complex biological system models.