dr hab Anna Gambin, prof. UW
Instytut Informatyki
Uniwersytetu Warszawskiego
e-mail: aniag@mimuw.edu.pl

Warszawa, 7 wrzesień 2014

Recenzja rozprawy doktorskiej

mgra inż. Adama Kozaka

zatytułowanej:

Grafy etykietowalne i sieci Petriego w analizie procesów biochemicznych i biologicznych.

1. Problem badawczy i jego znaczenie

Rozprawa podejmuje dwa problemy badawcze, które łączy metodologia modelowania przy użyciu teorii grafów. Są to zagadnienia: analizy danych pochodzących z sekwencjonowania przez hybrydyzację (SBH) oraz modelowania komórkowych procesów metabolicznych za pomocą sieci Petriego.

Asemblacja łańcuchów DNA otrzymanych w procesie sekwencjonowania przez hybrydyzację stanowiła motywację dla zdefiniowania nowej klasy grafów. Autor rozprawy analizuje własności grafów bazowo-etykietowalnych, które reprezentują model danych SBH zawierających regularne błędy. Grafy te są podgrafami indukowanymi grafów leksykalnych, stanowiących z kolei nadklasę grafów de Bruijna, a analiza ich własności jest interesująca także z czysto matematycznego punktu widzenia. Dodatkowo rozważany jest bardziej praktyczny problem budowy biblioteki oligonukleotydowej, czyli zbioru krótkich fragmentów DNA (oligonukleotydów), dla których zminimalizowane jest prawdopodobieństwo wzajemnej hybrydyzacji.

Rozważane w rozprawie modele sieci Petriego pozwoliły na formalny opis procesów regulacji hepcydyny oraz stresu oksydacyjnego w procesie rozwoju miażdżycy. Analiza strukturalna sieci (zbióru t-niezmienników oraz ich hierarchicznego porządku) umożliwiła w obydwu przykładach sformułowanie i weryfikację biologicznie ważnych hipotez.

Obydwa problemy rozważane przez Autora rozprawy mają niewątpliwie charakter naukowy, przy czym zagadnienia analizy własności grafów bazowo-etykietowalnych możemy uznać, za badania podstawowe, natomiast problemy dotyczące analizy modeli opartych na sieciach Petriego mają potencjalnie duże znaczenie praktyczne w diagnostyce i prognostyce medycznej

2. Wkład autora

Oryginalnym wkładem Autora rozprawy jest zdefiniowanie i analiza własności nowych klas grafów motywowanych problemem SBH. Dodatkowo został zaproponowany nowy algorytm budowy bibliotek oligonukleotydów antykomplementarnych, który wykorzystuje pojęcie zdefiniowanych przez Autora grafów leksykalnych.

Wkładem Autora w części dotyczącej modelowania procesów biologicznych jest analiza strukturalna zaproponowanych sieci Petriego oraz sformułowanie wniosków biologicznych z niej wynikających.

Wyniki przedstawione w rozprawie zostały opublikowane w serii prac naukowych. Autor rozprawy jest pierwszym autorem w pracy przedstawiającej analizę grafów etykietowalnych opublikowanym w czasopiśmie *Discrete Applied Mathematics (IF=0.67),* oraz w raporcie technicznym przezentującym algorytm budowania biblioteki oligonukleotydowej.

Orginalny model sieci Petriego dla homeostazy żelaza został zaproponowany w roku 2007 (*Journal of Biomedical Informatics, IF=2,48*) natomiast rezultaty obejmujące jego rozwiniecie oraz dogłębną analizę procesów regulacji hepcydyny, w którym to projekcie brał już udział Autor rozprawy zostały opublikowane w tym samym czasopiśmie w roku 2013. Dodatkowo model sieci Petriego dla stresu oksydacyjnego został opublikowany w wydawanym przez Politechnikę Poznańską czasopiśmie *Foundations of Computing and Decision Sciences,* natomiast model sieci Petriego wykorzystany do analizy anemii ukazał się w czasopiśmie *Bioprocess and Biosystems Engineering* (IF=1,82). Wspomniane wyniki powstawały w interdyscyplinarnym zespole, w którym dr n.med. D. Formanowicz odpowiadała za sfiormułowanie modelu biologicznego, natomiast Autor rozprawy był odpowiedzialny za analizę bioinformatyczną.

3. Poprawność

Szerokie spektrum problemów rozważanych w rozprawie zaowocowało też różnorodnością osiągniętych wyników. W rozdziale czwartym są to definicje i analiza własności nowych klas grafów etykietowanych, czyli rezultaty w postaci sformułowania definicji oraz adekwatnych twierdzeń wraz z ich dowodami. Docenić należy konsekwentną notację oraz przejrzystość formułowanych wniosków. Autor odnotowuje też fakt, że zdefiniowane przez niego pojęcie grafów leksykalnych, zostało niezależnie zdefiniowane (już po zgłoszeniu pracy do czasopisma Disc. Appl. Math) przez Zhi Xu jako m-przesunięte sekwencje de Bruijna (praca została finalnie opublikowana w materiałach konferencji AFL 2011). Fakt ten nie umniejsza dokonań Autora, zwłaszcza że idea m-przesuniętych sekwencji de Bruijna, chociaż równoważna grafom leksykalnym jest rozważana w innym kontekście.

Rozdział piąty zawiera algorytm budowy bibliotek nukleotydowych wykorzystujący grafy leksykalne. Został przeprowadzony eksperyment obliczeniowy, który pozwala na wybór adekwatnych parametrów algorytmu. Zgadzam się z uwagą Autora, że wyniki eksperymentu wzbogaciłaby analiza problemu liczby powtórzeń spektrum słowa losowego, który został sformułowany w rozprawie jako Problem 5.3 pozostając otwartym. Za niedopracowaną uważam też prezentację wyników eksperymentu (Tabele 5.3 i 5.4). Ciągnące się przez 10 stron ciągi liczb, z których wiele się powtarza, są dość nieczytelne. Alternatywna reprezentacja za pomocą odpowiednich wykresów, pozwoliłaby łatwiej odczytać wyniki.

Część pracy zawierająca modele sieci Petriego (rozdział 7 i 8) ma bardziej bioinformatyczny charakter. Zaproponowane przez biologów/lekarzy modele zostały przez Autora zanalizowane, co pozwoliło na sformułowanie biologicznie interesujących wniosków. W analizie modelu regulacji hepcydyny Autor skoncentrował się na identyfikacji t-niezmienników sieci (reprezentowanych jako wektory charakterystyczne zbioru tranzycji), które zostały następnie porównane i pogrupowane w strukturę hierarchiczną (metodą UPGMA). Ciekawe i narzucające wydaje się tutaj porównanie innych metod klastrowania danych, nie opartych na korelacji Pearsona, np PCA. Dodatkowo bardzo

pożądane byłyby odwołania do literatury potwierdzającej wiedzę biologiczną (por. str 116). Podobne zastrzeżenia dotyczą modelu stresu oksydacyjnego z rozdziału 8 rozprawy.

4. Wiedza kandydata

Wyniki przedstawione w rozprawie pozwalają stwierdzić, że Autor posiada ogólną wiedzę w dyscyplinie Informatyka. W rozdziale 2 Autor dokonuje przeglądu podstawowych pojęć i zagadnień biologii molekularnej i obliczeniowej. Rozdział 3 wprowadza podstawowe problemy teorii grafów i informatyki teoretycznej, natomiast rozdział 6 stanowi wyczerpujące wprowadzenie do problematyki badawczej sieci Petriego. Bibliografia rozprawy jest obszerna i ze względu na interdyscyplinarność tematyki zawiera artykuły zarówno matematyczne jak też biologiczne i bioinformatyczne.

5. Inne uwagi[1]

Rozprawa jest napisana przejrzyście i starannie, bez błędów językowych. Jedyny wyjątek stanowią błędy interpunkcyjne, gdyż Autor bardzo często nie rozdziela przecinkiem zdań złożonych nawet w sytuacji, gdy są one połączone spójnikiem zobowiązującym do postawienia przecinka lub zaimkiem względnym "który".

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami)[2] moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problem naukowego? (wybierz jedną opcję stawiając znak **X**)

[x]	[ ]	[ ]	[ ]	[ ]
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka lub Automatyka i Robotyka?

[x]	[ ]	[ ]	[ ]	[ ]
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

C. Czy kandydat umiejętność samodzielnego prowadzenia pracy naukowej?

[x]	[ ]	[ ]	[ ]	[ ]
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*



1" Opcjonalnie

2" http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf