

Bírálat

Farkas Illés

Átfedő modulok molekuláris biológiai kölcsönhatási hálózatokban
című MTA doktori értekezéséről

A jelölt doktori értekezésben a modern hálózat kutatás megközelítésével vizsgálta az élő rendszerek molekuláris biológiai hálózatait, illetve új eszközök biztosítása révén hozzájárult ahhoz, hogy további vizsgálatokat lehessen végezni ezen a téren. Az értekezés témájának aktualitása részben a különböző molekuláris biológiai hálózatok elérhetővé válásából fakad. Másrészt, a skálafüggetlen hálózatok univerzális tulajdonságainak felismerése révén szintén az elmúlt években vált általánosan elterjedté a korábban csak matematikai modellekben használt hálózat kutatás módszertana. A dolgozat témája tehát kétséget kizáróan korszerű és érdeklődésre számottartó.

A jelölt tudományometriai mutatói nemcsak, hogy megfelelnek az MTA doktora cím követelményeinek, hanem magasan túlteljesítik azt mind impakt faktor, mind hivatkozások tekintetében. A részletesebb vizsgálat egy kicsit árnyaltabb képet mutat. Az értekezés alapjául 12 publikáció szolgál. Ezek között szerepel egy Nature cikk, amely eddig több mint 1500 hivatkozást kapott. Ez nemzetközi szinten is kiemelkedőnek számít, és az egyik legtöbbet hivatkozott Magyarországon készült publikáció. Szerepel még a listában két magas impakt faktorú cikk, amelyek azonban nem tartalmaznak önálló kutatási eredményt. Ebből az egyik a Cell újságban közölt kitekintés cikk, míg a Science Signalling újságban megjelent közlemény egy Csermely Péter által tartott előadás nem referált kivonata. A jelöltnek van további 7 cikke, amelyek elkészítésében a Ph.D. fokozat megszerzése órá részt vett, ezek más témákban íródtak. A megadott 12 cikk közül a jelölt 5-ben első szerző, egyben megosztott első szerző. Azonban csak egyben utolsó szerző, ami egy nagydoktori esetén jóval alatta van az átlagnak.

Összességében az értekezés igényesen elkészített, jól olvasható mű. Bár a hálózat kutatás legvonzóbb tulajdonsága, hogy segítségével egészen eltérő típusú hálózatokat lehet vizsgálni, az értekezés a biológiai hálózatokat elemzésére szorítkozik. Tematikailag így sokkal egységesebb lett az értekezés, és ez remek lehetőséget biztosít a biológiai hálózatok sokszínűségének a szemléltetésére is. Maga a dolgozat elsősorban egy fizikus szemszögéből íródott, ennek megfelelően nagyobb hangsúly esik az eredmények megértéséhez szükséges biológiai ismeretek bevezetésére. Alapvetően a dolgozat szerintem nagyon sikeresen megoldja azt a nehéz feladatot,

ami a dolgozatban tárgyalt kutatások határterület jellegéből adódik, hogy a dolgozat mind fizikusok, mind biológusok számára érthető legyen.

A bevezetés a statisztikus fizika és a biológia határterületén található tudományterület bemutatásával indul. Ezt a fejezetet a dolgozatban tárgyalt kérdések alfejezet követi, amelyben a kérdések felvetése helyett inkább egy rövid áttekintést kapunk a dolgozatban leírt eredményekről. Az értekezés legnagyobb része a kutatási előzményeket és az eredményeket mutatja be három fejezetre tagolva. Ezek közül a legnagyobb terjedelmű a fehérje-fehérje kölcsönhatási hálózatok moduljairól szóló rész, ami több mint 60 oldal. Ebben a fejezetben kerülnek bevezetésre a hálózatok jellemzésére szolgáló alapvető fogalmak és modellek. Ezt követi a transzkripció és transzláció szabályozási hálózatok moduljairól szóló fejezet, ami összesen 26 oldal. A harmadik, jelátviteli hálózatokról szóló rész mindössze 12 oldal, és itt az alkalmazásokon van hangsúly. Ezt a fejezetet aránytalanul rövidnek éreztem, szerintem itt a jelátviteli hálózatokról is, illetve az ezekhez kapcsolódó eredményekről is érdemes lett volna bővebben írni.

Az eredményeknél a jelölt nagy hangsúlyt fektet arra, hogy saját hozzájárulását mindig pontosan megjelölje, mindegyik alfejezetben külön-külön. Ugyanakkor számomra szokatlan volt, hogy nincsenek igazi kutatási kérdések megfogalmazva, illetve az eredmények jelentőségének és a további kérdések diszkussziója szinte teljesen hiányzik.

A tézisekben illetve a dolgozatban megadott “Új tudományos eredmények” egy részét nem tudom elfogadni annak, mivel vagy nem új, vagy nem tudományos eredmény. Ezek közé tartozik például már ismert algoritmusok beprogramozása, web oldal üzemeltetése, adat hibák kiszűrése, különböző forrásból származó bejegyzések neveinek átalakítása. Tisztában vagyok azzal, hogy ezek rendkívül fontos és időigényes részei a számítógépes biológiai, bioinformatikai munkáknak, azonban ezek alapvetően technikai jellegű tevékenységek, amelyek önmagukban nem tekinthetők tudományos eredményeknek, különösen nem egy nagydoktori szintjén.

A jelölt eredményeinek nagy részét csapatmunkában érte el. Ezek közül kiemelkedik a CPM (Clique Percolation Method) hálózati modulkereső algoritmus kidolgozása, amely egymással átfedő és a környezethez képest nagy belső élsűrűséggel rendelkező hálózati modulokat azonosít. A módszer egyik nagy előnye sok más klaszterező eljáráshoz képest, hogy lehetővé teszi átfedő modulok azonosítását is. A módszert kiterjesztették irányított és súlyozott gráfokra, és alkalmazták a biológiai hálózatok mellett kutatók együttműködési, web oldalak, illetve szó asszociációs hálózatok elemzésére is.

Szintén nagy jelentőségűnek gondolom a Signalink adatbázis megalapozását, illetve elindítását, amelynek révén a különböző jelátviteli folyamatokról kaphatunk pontosabb képet. Ennek segítségével azonosítottak például új komponenseket is ezekben a hálózatokban, illetve elemezték a jelátviteli hálózatok közötti áthallást biztosító kapcsolatokat. Ezek nagyon fontosak lehetnek a jelátviteli folyamatok működésének pontosabb megértésében, illetve felhasználhatók lehetnek gyógyszertervezésben is.

Az értekezés és a tézisek alapján új tudományos eredménynek fogadom el a következőket:

1. Részt vett a CPM módszer kidolgozásában, illetve annak kiterjesztésében irányított és súlyozott hálózatokra, továbbá a módszer teszteléséhez szükséges adatok összegyűjtésében. A kidolgozott módszer segítségével fehérje-fehérje kölcsönhatási hálózatokat tanulmányoztak élesztőben. A jelölt hozzájárulásával meghatározták az átfedő modulokat, és elemezték a modulok legfontosabb tulajdonságait, többek között az átlagos modul számot, a modulok jellemző méretét és méret eloszlását, definiálták a modulok kapcsolati számát és meghatározták ezek eloszlását. A jelölt végezte a modulokhoz kapcsolódó statisztikusan szignifikáns biológiai funkciók meghatározását.

2. Az irányított CPM módszert alkalmazta a jelölt transzkripció szabályozó hálózatokra. A csúcspontok relatív fokszáma és modul száma alapján azonosított fő irányító csúcspontokat és alközpontokat ezekben a hálózatokban.

3. A jelölt vezetésével vizsgálták az emberi mikroRNS-ek szerepét a transláció csendesítési hálózat elemzésével. Azonosították a hálózat moduljait, és definiálták a modulokban lévő mikroRNS-ek fontosságát. Ennek alapján megjósolhatóak azok a mikroRNS-ek, amelyek eltávolítása nagyobb eséllyel okoz fenotípus változást.

4. Két fős csapatban elindította a Signalink adatbázist, és leírta a kézi adatgyűjtés kritériumait. Elemezte a jelátviteli hálózatok közötti áthallást. Részt vett a Signalink adatbázis és a PathwayLinker web oldalainak kidolgozásában.

Összességében az eredményeket érdemes lett volna kicsit kevésbé a technikai feladatokra koncentrálni megírni.

Az értekezés olvasása közben a következő kérdések fogalmazódtak meg bennem:

1. A jelenlegi biológiai hálózatokban nagyon sok fals pozitív, tehát hibás, illetve fals negatív, tehát hiányzó találatot találunk. Hogyan befolyásolja a kapcsolatok hiánya, illetve a hamis kapcsolatok megléte a hálózatok általános jellemzőit, illetve a modulok megjelenését és eloszlását?

2. A fehérje kölcsönhatási hálózatok esetén meg lehet különböztetni asszociációt, illetve direkt kölcsönhatást. Hogyan változnak a modulok, ha csak direkt kölcsönhatásokat vizsgálunk? Így is leírhatóak-e az azonos funkcionális csoportban található fehérjék, mint átfedő klikkek?

3. A dolgozatban négyfajta biológiai hálózatot elemzett. Érdeemes lett volna ezeknek a hálózatoknak az összehasonlítását elvégezni. Vannak-e jelentős eltérések például a hálózatok általános paramétereinek, illetve a modulok méretének és eloszlásának tekintetében? Tudná-e értelmezni az eltérések biológiai következményeit?

4. A SignaLink adatbázis, amely elindításában a jelölt fontos szerepet játszott, az elmúlt években sokat bővült. Lehet-e ezt az adatbázist is elemezni modulok szintjén? Jellemző-e a jelátviteli útvonalakra is az átfedő modulok megléte, illetve ezek összefüggésben hozhatóak-e a meglévő útvonalakkal, vagy pedig továbbra is döntően lineáris kaszkádnak kell tekintetünk a jelátviteli útvonalakat?

5. Mivel a dolgozattól hiányzik egyfajta kitekintés, ezért ezzel kapcsolatban a következő kérdéseket fogalmazom meg: Hogyan látja a jelölt a hálózatkutatás biológiai rendszerekre való alkalmazásának jövőjét? Vannak-e még megválaszolatlan kérdések a területen? Vannak-e olyan molekuláris biológiai hálózatok, amelyek átfedő modulokkal nem vizsgálhatóak? Vannak-e olyan előrejelzései a modelleknek, amelyek a biológusok kísérletes igazolására várnak?

Összefoglalásképpen megállapító, hogy Farkas Illés értekezése kiemelkedően magas szintű kutatási eredményeken alapul, amelyek eléréséhez a jelölt eltérő mértékben, de aktívan hozzájárult. A jelölt bővítette a modern hálózatkutatás eszköztárát, és ezek felhasználásával új tudományos eredményeket ért el. Ugyanakkor eddigi iskolateremtő munkássága elmarad a nagydoktori szintjén elvárhatótól, amit az mutat, hogy eddig csak egy cikkben volt utolsó szerző. Ezt a hiányérzetet erősíti az értekezésben az elért eredmények jelentőségének, és további kutatási irányok diszkussziójának hiánya. Ennek ellenére, figyelembe véve a jelölt magasán kiemelkedő publikációs eredményességét, javaslom a disszertáció nyilvános vitára bocsájtását, valamint sikeres védés

esetén az MTA doktori fokozat odaítélését.

Budapest, 2016. április 26.

A handwritten signature in blue ink, consisting of stylized cursive letters that appear to be 'DZ'.

Dosztányi Zsuzsanna, PhD