Warszawa, 14 maja 2023

dr hab. Michał Pilipczuk

Instytut Informatyki
Uniwersytet Warszawski
ul. Banacha 2, 02-097 Warszawa
michal.pilipczuk@mimuw.edu.pl

Recenzja w postępowaniu habilitacyjnym dr inż. Krzysztofa Turowskiego

Dr inż. Krzysztof Turowski złożył wniosek o przeprowadzenie postępowania w sprawie nadania stopnia doktora habilitowanego i w ramach tego wniosku przedstawił osiągnięcie naukowe: cykl powiązanych tematycznie prac pod tytułem

Structural analysis and compression for duplication random graph models.

W skład tego cyklu wchodzi pięć opublikowanych prac naukowych:

[A1] Krzysztof Turowski, Wojciech Szpankowski,
Towards Degree Distribution of a Duplication-Divergence Graph Model,
Electronic Journal of Combinatorics, 28(1) (2021), P1.18.

[A2] Alan Frieze, Krzysztof Turowski, Wojciech Szpankowski,
Degree Distribution for Duplication-Divergence Graphs: Large Deviations,
46th International Workshop on Graph-Theoretic Concepts in Computer Science, WG 2020,
Lecture Notes in Computer Science 12301, pages 226-237.

[A3] Alan Frieze, Krzysztof Turowski, Wojciech Szpankowski,
The concentration of the maximum degree in the duplication-divergence models,
27th International Conference of Computing and Combinatorics, COCOON 2021,
Lecture Notes in Computer Science 13025, pages 413-424.

[A4] Philippe Jacquet, Krzysztof Turowski, Wojciech Szpankowski,
Power-Law Degree Distribution in the Connected Component of a Duplication Graph,
31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, AofA 2020,
Leibniz International Proceedings in Informatics (LIPIcs) 159, pages 16:1-16:14.

[A5] Krzysztof Turowski, Abram Magner, Wojciech Szpankowski,
Compression of Dynamic Graphs Generated by a Duplication Model,
Algorithmica 82(9) (2020), pages 2687-2707.

Tematyka osiągnięcia naukowego. Wszystkie prace składające się na osiągnięcie naukowe dotyczą konkretnego modelu grafu losowego, wprowadzonego przez Solégo i Pastora-Satorrasa: *modelu duplikacyjnego.* W tym modelu, aby skonstruować duży graf losowy G_t (o t wierzchołkach), zaczynamy z ustalonego małego grafu G_{t_0} (o t_0 wierzchołkach) i wykonujemy $t - t_0$ rund duplikacji i mutacji. W każdej rundzie, wpierw duplikujemy jeden wierzchołek grafu (wybrany losowo z jednostajnym prawdopodobieństwem), a następnie losowo mutujemy jego sąsiedztwo: każda ze skopiowanych krawędzi zostaje zachowana niezależnie z prawdopodobieństwem p, oraz łączymy nowy wierzchołek z każdym z uprzednich niesąsiadów niezależnie z prawdopodobieństwem $\frac{r}{n}$, gdzie n to obecna liczba wierzchołków. W ten sposób, w oczekiwaniu dodajemy r zupełnie nowych krawędzi incydentnych ze skopiowanym wierzchołkiem. Liczby p i r są parametrami modelu i w zależności od nich uzyskane grafy losowe będą miały różne własności strukturalne.

Model duplikacyjny jest umotywowany konkretnymi zagadnieniami z nauk przyrodniczych, gdzie przykładowo biologiczne procesy prowadzące do powstawania sieci połączeń wydają się być dobrze opisywane przez sekwencje rund duplikacji i mutacji. W literaturze można spotkać również prace próbujące eksperymentalnie —

przy pomocy symulacji komputerowych — oszacować i opisać zachowanie grafów losowych konstruowanych zgodnie z modelem duplikacyjnym. Prace przedstawione w osiągnięciu naukowym podchodzą do problemu od strony czysto matematycznej: podstawowym celem badawczym było dogłębne zbadanie rozkładu probabilistycznego ilościowych parametrów grafów losowych konstruowanych zgodnie z modelem duplikacyjnym. Wśród rozważanych parametrów znalazły się:

- stopień ustalonego wierzchołka $\deg_t(s)$,
- średni stopień grafu $D(G_t)$,
- maksymalny stopień grafu $\Delta(G_t)$,
- odsetek wierzchołków danego stopnia.

Analiza miała na celu nie tylko wyznaczenie asymptotyki wartości oczekiwanej, ale również głębsze zrozumienie rozkładu, przykładowo na ile rozważane zmienne losowe koncentrują się wokół wartości oczekiwanych.

Omówienie prac zawartych w osiągnięciu naukowym. Praca [A1] jest pierwszą, wstępną pracą w cyklu. Dotyczy ona asymptotyki wartości oczekiwanej i wariancji stopnia ustalonego wierzchołka $\deg_t(s)$ oraz średniego stopnia $D(G_t)$ w modelu duplikacyjnym. W obu przypadkach, definicja modelu duplikacyjnego pozwala na zapisanie prostego równania rekurencyjnego uzależniającego wartość oczekiwaną (odpowiednio, drugi moment) rozważanej zmiennej losowej w czasie $t + 1$ od wartości dla czasu t. Następnie, autorzy analizują to równanie rekurencyjne przy pomocy klasycznych metod matematyki dyskretnej, uzyskując zarówno (bardzo skomplikowane) wzory jawne jak i (mniej skomplikowane i bardziej intuicyjne) szacowania asymptotyczne. W szczególności, w tych oszacowaniach można dostrzeć jakościowe własności modelu duplikacyjnego, przykładowo różną asymptotykę oczekiwanych stopni starszych i nowszych wierzchołków. Cała praca [A1] jest zatem bardzo techniczna, ale koniec końców koncepcyjnie dość prosta: rekurencje pojawiają się naturalnie i jedyne co trzeba zrobić, to „zakasać rękawy" i je umiejętnie przeanalizować.

Praca [A2] rozważa te same zmienne co praca [A1] — stopień ustalonego wierzchołka $\deg_t(s)$ i średni stopień $D(G_t)$ — ale tym razem celem jest oszacowanie koncentracji wokół wartości oczekiwanej. W tym celu autorzy analizują funkcję tworzącą momenty i podobnie jak w podejściu z pracy [A1], piszą naturalne równanie rekurencyjne dla tej funkcji celem uzyskania oszacowań asymptotycznych. Te oszacowania można potem użyć do szacowania ogonów rozważanych zmiennych, podobnie jak to ma miejsce w dowodzie nierówności Chernoffa. Koniec końców praca pokazuje dość ścisłe (z dokładnością do czynników polilogarytmicznych od t) oszacowania dla ogonów rozważanych zmiennych (zarówno ogonów powyżej wartości oczekiwanej jak i poniżej). Pozwala to na dobre zrozumienie ich rozkładów i jest pomocne do dalszej teoretycznej analizy modelu duplikacyjnego.

Praca [A3] dotyczy analizy maksymalnego stopnia $\Delta(G_t)$ grafu wylosowanego zgodnie z modelem duplikacyjnym. Okazuje się, że przy założeniu $p \in (\frac{1}{2}, 1)$, $\Delta(G_t)$ z dużym prawdopodobieństwem jest rzędu t^p, z dokładnością do czynników polilogarytmicznych od t. Wydaje się, że od strony matematycznej jest to najciekawsza praca w cyklu, gdyż trzeba zastosować inne podejście: maksymalny stopień grafu już nie daje się opisać tak łatwo przy pomocy równań rekurencyjnych jak np. średni stopień. Dowód opiera się na bardzo uważnym śledzeniu wzrostu stopnia $\deg_t(s)$ każdego ustalonego wierzchołka s. Rozkład probabilistyczny $\deg_t(s)$ jest już zrozumiany dzięki pracom [A1] i [A2], natomiast kluczowe nowe obserwacje osadzają się na tym, że również różnice postaci $\deg_{t'}(s) - \deg_t(s)$ dla $t < t'$ można dobrze zrozumieć i oszacować wykładniczo ich koncentrację (przy pomocy narzędzi Chernoffowskich). Pozwala to na dobre oszacowanie prawdopodobieństwa, że w danym momencie czasu stopień każdego wierzchołka nie będzie za duży, jak również prawdopodobieństwa, że będzie istniał wierzchołek o w miarę dużym stopniu — co pozwala na oszacowanie maksymalnego stopnia zarówno z dołu jak i z góry. Metoda użyta w rozumowaniu jest moim zdaniem technicznie ciekawa i pomysłowa; tak więc wysoko oceniam matematyczny wkład pracy [A3].

Praca [A4] dotyczy szacowania rozkładu stopni w grafach losowych konstruowanych zgodnie z modelem duplikacyjnym, przy wyzerowanym parametrze r i przy małym parametrze p (dokładniej, $p \in (0, e^{-1})$). Wiadomo było, że w takich grafach znacząca większość wierzchołków jest izolowana, więc rozważamy rozkład stopni po usunięciu izolowanych wierzchołków. Wcześniejsza praca Jordana wykazała, że jeśli przez a_k oznaczymy frakcję wierzchołków stopnia k (dla dużych t), to asymptotycznie a_k szacuje się z góry przez k^{-q} dla każdego $q < \beta(p)$ i z dołu przez k^{-q} dla każdego $q > \beta(p)$, gdzie $\beta(p) \in (2, 3)$ jest pewną stałą zależną od p; ale dokładne zrozumienie asymptotyki a_k nie było znane. Praca [A4] wypełnia tą lukę wykazując, że wartości a_k istotnie zachowują się asymptotycznie istotnie jak $C \cdot k^{-\beta(p)}$ dla pewnej konkretnej stałej C, a więc spełniają prawo wykładnicze (*power law*) z wykładnikiem $\beta(p) \in (2, 3)$, co potwierdza wcześniejsze prace eksperymentalne. Rozumowanie opiera się na rozważeniu funkcji tworzącej $A(z)$ ciągu a_k, dla której odpowiednie równanie różniczkowe zostało zauważone przez Jordana. Następnie autorzy stosują metody analizy zespolonej do oszacowania współczynników rozwinięcia analitycznego $A(z)$, czyli wyrazów a_k.

Wreszcie, w pracy [A5] rozważany jest zdegenerowany przypadek modelu duplikacyjnego: $p = 1$ oraz $r = 0$, co oznacza że każda runda polega po prostu na zduplikowaniu losowo wybranego wierzchołka. W związku z tym, każdy graf G_t skonstruowany zgodnie z takim modelem można uzyskać z G_{t_0} poprzez skopiowanie każdego wierzchołka G_{t_0} pewną liczbę razy. A więc i cały G_t można opisać przy pomocy krotki składającej się z t_0 dodatnich liczb całkowitych, mówiących ile kopii każdego wierzchołka G_{t_0} posiada G_t. Centralnymi wynikami pracy [A5] są oszacowania na entropię $H(G_t)$, gdzie G_t jest traktowany jako graf etykietowany, oraz na tzw. entropię strukturalną $H(S_t)$, gdzie G_t jest traktowany jako graf nieetykietowany; sprowadza się to do entropii rozkładu klasy izomorfizmu G_t, oznaczanej jako S_t. Wykazane zostało, że $H(G_t)$ jest rzędu $\Theta(t)$, zaś $H(S_t)$ jest rzędu $\Theta(\log t)$ (praca zawiera dużo dokładniejsze oszacowania). Tak diametralna różnica nie jest zaskakująca: wszak G_t jest zawsze bardzo prostym grafem, którego klasa izomorfizmu S_t jest opisywalna przy pomocy krotki liczb całkowitych o długości t_0, co uznajemy za stałą. Formalny dowód wykorzystuje lemat z pracy Łuczaka i innych dotyczący porównywania entropii i entropii strukturalnej dla modeli grafów losowych, oraz standardowe metody matematyki dyskretnej. Oprócz wyników dotyczących entropii, praca zawiera również wyniki algorytmiczne, dotyczące zwartego zapisu (kompresji) rozważanych grafów losowych. Muszę przyznać, że praca [A5] nie urzekła mnie: rozważany skrajny przypadek $p = 1$ oraz $r = 0$ zupełnie wypacza rozumienie modelu duplikacyjnego jako procesu konstruującego losowy graf, gdyż otrzymywane grafy są skrajnie proste, składają się ze stałej liczby wierzchołków skopiowanych wiele razy. W związku z tym, choć pytanie o entropię rozkładu takich grafów jest dobrze zadanym problemem matematycznym, jego motywacja stoi pod znakiem zapytania.

Pozostały dorobek. Oprócz prac przedstawionych w osiągnięciu naukowym, dr inż. Turowski jest również współautorem kilkunastu innych prac o różnorodnej tematyce.

- Najbliższe problematyce osiągnięcia naukowego są dwie prace dotyczące estymacji parametrów modeli grafów losowych na podstawie danych, do których próbujemy je przypasować, oraz rekonstruowania ewolucji grafów budowanych przy pomocy różnych przyrostowych modeli grafów losowych. W gruncie rzeczy, te dwie prace mogłyby być włączone do osiągnięcia naukowego.
- Pod względem metodologicznym niedalekie są również dwie prace dotyczące losowych drzew, rozważające szacowania na entropię i entropię strukturalną w pewnym konkretnym modelu konstrukcji losowego drzewa, jak również jedna praca dotycząca entropii rozkładu wielomianowego Dirichleta.
- W dorobku Habilitanta można również znaleźć szereg prac teoriografowych, dotyczących różnego rodzaju kolorowań grafu oraz często dotykających aspektów algorytmicznych. Najwcześniejsza jest seria prac dotycząca kolorowań szkieletowych, co było tematem zarówno pracy magisterskiej jak i doktorskiej Habilitanta. Późniejsze prace dotyczą takich tematów jak uogólnione liczby kolorujące, gry chromatyczne, czy kolorowania rozpiętościowe.
- Wreszcie są dwie prace algorytmiczne: jedna dotycząca złożoności i aproksymacji dla pewnej klasy problemów szeregowania, druga dotycząca złożoności problemu rekonstrukcji hipergrafu.

Znaczenie i ocena osiągnięcia habilitacyjnego. Cykl prac przedstawionych jako osiągnięcie naukowe przedstawia dogłębną i kompleksową analizę modelu duplikacyjnego. Z tego co rozumiem, podstawowym celem tej analizy było zrozumienie od strony teoretycznej na ile grafy konstruowane przy pomocy modelu duplikacyjnego spełniają różnego rodzaju własności obserwowane w praktyce, m.in. *power law*. Ten cel został w gruncie rzeczy osiągnięty. W ten sposób, matematyczny opis parametrów ilościowych grafów konstruowanych przy pomocy modelu duplikacyjnego stanowi istotny wkład w dziedzinę. Uwagę zwraca zwłaszcza spójność tematyczna osiągnięcia — Habilitant rzeczywiście "rozpracował" całe zagadnienie — choć uzasadnione wątpliwości może budzić wąskość tego zagadnienia.

Jeśli chodzi o aparat matematyczny, wszystkie prace bazują na klasycznych metodach matematyki dyskretnej i dyskretnego rachunku prawdopodobieństwa, i moim zdaniem prezentują dobry lub bardzo dobry poziom techniczny. Zwłaszcza wybijają się praca [A3], gdzie metoda szacowania przyrostów stopni użyta to śledzenia maksymalnego stopnia jest naprawdę pomysłowa, oraz praca [A4], używająca ciekawych metod analizy zespolonej. Ogólnie, choć matematyka uprawiana w pracach jest bardzo techniczna, to przedstawiony wkład jest solidny zarówno pod względem idei jak i wykonania.

Nie pracując w dziedzinie grafów losowych jest mi trudno ocenić motywację i znaczenie modelu duplikacyjnego w szerszym kontekście. Patrząc z perspektywy osoby spoza działki, model ten na pewno ma sens i jest ciekawy, a jego teoretyczna analiza wymaga zaawansowanego aparatu matematycznego. Pewną miarą znaczenia wyników może być tutaj cytowalność prac Habilitanta, która niestety jest dość mierna. Według Google Scholar, każda z prac z osiągnięcia naukowego była cytowana tylko kilka lub kilkanaście razy, przy czym są to prawie wyłącznie autocytowania. Choć prace te były publikowane niedawno — w latach 2020 i 2021 — wskazuje to na niszowość wyników uzyskanych przez Habilitanta. Również konferencje i czasopisma, w których wyniki były publikowane, choć są dobre, to nie należą do najwyższej półki.

Uwagę zwraca fakt, że współautorem wszystkich prac zawartych w osiągnięciu naukowym jest prof. Wojciech Szpankowski, mentor Habilitanta z okresu jego zatrudnienia na Purdue University. Jednak oświadczenia współautorskie jednoznacznie pokazują, że we wszystkich pracach zawartych w osiągnięciu wkład merytoryczny Habilitanta był kluczowy oraz został oceniony jako większościowy. Tym samym, przedstawione osiągnięcie w zdecydowany sposób podkreśla samodzielność naukową Habilitanta. Uważam to za mocny punkt wniosku.

Znaczenie i ocena całości dorobku. W pozostałym dorobku Habilitanta można znaleźć pięć prac dotyczących podobnej tematyki i używających podobnej metodologii. Wszystkie one prezentują podobny poziom co prace zawarte w osiągnięciu naukowym: są to solidne wyniki, ale nie należące do najwyższej półki zarówno pod względem motywacji jak i miejsc, gdzie zostały opublikowane.

Inne prace Habilitanta reprezentują szerokie spektrum dziedzin. Z wyłączeniem badań dotyczących kolorowań szkieletowych, należy je zapewne uznać za wyniki incydentalnej współpracy z innymi badaczami. Wśród tych pozycji znów brak jest prac przełomowych czy wybitnych, a wiele artykułów ma jednak charakter przyczynkowy. Ale można również znaleźć prace dobre i zawierające solidny wkład matematyczny; przykładem może być tutaj praca dotycząca problemów szeregowania. Podobnie jak w przypadku prac zawartych w osiągnięciu naukowym, pozostałe artykuły Habilitanta są również mało widoczne w środowisku: liczby ich cytowań są bardzo niskie.

Patrząc całościowo na dorobek Habilitanta wydaje się, że najtrafniejszym przymiotnikiem opisującym go jest „solidny". Na pewno brakuje prac wybijających się i rzeczywiście posuwających całą dziedzinę badań do przodu. Ale również nie brakuje prac dobrych i bardzo dobrych, badających ciekawe problemy, operujących poważnym aparatem matematycznym, oraz zawierających nowe pomysły i rozwiązania.

Uwagę zwracają dwa cykle prac — jeden dotyczący kolorowań szkieletowych i drugi dotyczący modelu duplikacyjnego — będące podstawami odpowiednio pracy doktorskiej i wniosku habilitacyjnego. Choć skoncentrowanie badań na konkretnych zagadnieniach jest bez wątpienia chwalebne, to w mojej opinii problematyka poruszana w obu cyklach jest bardzo wąska oraz, niestety, niszowa; tyczy się to zwłaszcza kolorowań szkieletowych i pokrewnych prac. Tego typu zawężenie wysiłków nie prowadzi w dłuższej perspektywie do zdrowego rozwoju naukowego, który musi cechować się podejmowaniem nowych wyzwań. Sugerowałbym w przyszłości bardziej ambitne atakowanie problemów ważnych i szerzej zakrojonych, gdyż jak pokazują choćby prace [A3] oraz [A4], Habilitant zdecydowanie ma predyspozycje do uprawiania matematyki po prostu ciekawszej.

Działalność naukowa. Dr inż. Turowski obronił pracę doktorską na Politechnice Gdańskiej w roku 2015 i pracował na stanowisku podoktorskim na Purdue University w USA (dwa okresy zatrudnienia, w sumie prawie dwa lata). Od 2019 roku pracuje jako adiunkt na Wydziale Matematyki i Informatyki Uniwersytetu Jagiellońskiego. W międzyczasie miał dwuletnią przerwę w pracy naukowej, podczas której pracował jako inżynier oprogramowania w Google. To ciekawa ścieżka kariery naukowej, podczas której bez wątpienia zostały zebrane wartościowe doświadczenia: zarówno w kraju jak i zagranicą, oraz zarówno w środowisku akademickim jak i poza nim. Odzwierciedleniem tego jest szerokie spektrum współpracowników naukowych. Jeśli chodzi o zarówno formalne jak i zwyczajowe wymagania dotyczących mobilności, w mojej ocenie Habilitant spełnia je z naddatkiem.

Ponadto, Habilitant jest w tym momencie kierownikiem grantu SONATA, przyznanego przez NCN, w którym zajmuje się głównym nurtem swoich badań.

Uwagę również zwraca szeroko zakrojona działalność dydaktyczna Habilitanta: imponująca lista prowadzonych kursów i zajęć, oraz opieka nad dużą liczbą prac dyplomowych (4 magisterskie, 6 licencjackich). Na razie Habilitant nie opiekował się jeszcze doktorantami, ale biorąc pod uwagę jego profil, stanie się to niechybnie.

W ogólności, działalność akademicka Habilitanta jest absolutnie bez zarzutu: jest aktywnym członkiem społeczności akademickiej i rozwija swoją działalność naukową zarówno poprzez różne współprace, w tym zagraniczne, jak i poprzez prowadzenie grantów.

Podsumowanie. W mojej ocenie osiągnięcia naukowe dr inż. Krzysztofa Turowskiego przedstawione we wniosku habilitacyjnym spełniają zarówno formalne jak i zwyczajowe wymagania uzasadniające przyznanie stopnia doktora habilitowanego, choć muszę przyznać, że konkluzja ta wymagała pewnego zastanowienia. W dorobku Habilitanta brak jest prac wybijających się, zaś badane zagadnienia są często wąskie, niszowe, a ich widoczność w środowisku jest niska. Z drugiej strony, prace Habilitanta reprezentują dobry poziom matematyczny, są bardzo zwarte tematycznie i kompleksowo opracowują badane zagadnienia. Dzięki temu, przedstawione wyniki stanowią solidny i indywidualny wkład w dziedzinę badań. Tym samym, **rekomenduję nadanie dr inż. Krzysztofowi Turowskiemu stopnia doktora habilitowanego**.

Michał Pilipczuk