Gdańsk 02.02.2022

dr hab. Julian Szymański, prof. PG
Politechnika Gdańska
Wydział Elektroniki, Telekomunikacji i Informatyki
Katedra Architektury Systemów Komputerowych
ul. Narutowicza 11/13, 80-233 Gdańsk

RECENZJA

osiągnięcia naukowego oraz istotnej aktywności naukowej

dr inż. Radosława Michalskiego

w związku z postępowaniem o nadanie stopnia doktora habilitowanego nauk technicznych

Niniejsza recenzja przygotowana została na zlecenie przewodniczącego Rady Dyscypliny Naukowej Informatyki Technicznej i Telekomunikacji Politechniki Wrocławskiej (pismo z dnia 14.12.2021). Recenzja zawiera ocenę osiągnięcia naukowego, istotnej aktywności naukowej i dydaktycznej Kandydata.

1. Ocena osiągnięcia naukowego

Jako osiągnięcie naukowe zatytułowane „Analiza sieci społecznych: próbkowanie, klasyfikacja i dyfuzja" Habilitant podał następujący cykl 14 artykułów, wymieniam wraz z jego procentowym udziałem w ich powstanie:

[H1 - 40%] Jarosław Jankowski, Radosław Michalski, Piotr Bródka, Przemysław Kazienko, and Sonja Utz. Knowledge acquisition from social platforms based on network distributions fitting. Computers in Human Behavior, 51:685–693, Elsevier, 2015

[H2 - 10%] Jarosław Jankowski, Piotr Bródka, Przemysław Kazienko, Bolesław K. Szymański, Radosław Michalski, and Tomasz Kajdanowicz. Balancing speed and coverage by sequential seeding in complex networks. Scientific reports, 7(1):1–11, Nature Publishing Group, 2017.

[H3 - 33%] Jarosław Jankowski, Radosław Michalski, and Piotr Bródka. A multilayer network dataset of interaction and influence spreading in a virtual world. Scientific data, 4(1):1–9, Nature Publishing Group, 2017.

WPŁYNĘŁO
08-02-2022
RDN ITiT/24 /2022

[H4 - 35%] Jarosław Jankowski, Bolesław K. Szymański, Przemysław Kazienko, Radosław Michalski, and Piotr Bródka. Probing limits of information spread with sequential seeding. Scientific reports, 8(1):1–9, Nature Publishing Group, 2018.

[H5 - 25%] Marcin Kulisiewicz, Przemysław Kazienko, Bolesław K. Szymański, and Radosław Michalski. Entropy measures of human communication dynamics. Scientific reports, 8(1):1–8, Nature Publishing Group, 2018.

[H6 - 50%] Michał Weskida and Radosław Michalski. Finding influentials in social networks using evolutionary algorithm. Journal of Computational Science, 31:77–85, Elsevier, 2019.

[H7 - 75%] Mateusz Nurek and Radosław Michalski. Combining machine learning and social network analysis to reveal the organizational structures. Applied Sciences, 10(5):1699, MDPI, 2020.

[H8 - 40%] Radosław Michalski, Daria Dziubałtowska, and Piotr Macek. Revealing the character of nodes in a blockchain with supervised learning. IEEE Access, 8:109639–109647, IEEE, 2020.

[H9 - 65%] Radosław Michalski, Jarosław Jankowski, and Piotr Bródka. Effective influence spreading in temporal networks with sequential seeding. IEEE Access, 8:151208–151218, IEEE, 2020.

[H10 - 25%] Piotr Bródka, Jarosław Jankowski, and Radosław Michalski. Sequential seeding in multilayer networks. Chaos: An Interdisciplinary Journal of Nonlinear Science, 31(3):033130, AIP Publishing, 2021.

[H11 - 60%] Radosław Michalski, Bolesław K. Szymański, Przemysław Kazienko, Christian Lebiere, Omar Lizardo, and Marcin Kulisiewicz. Social networks through the prism of cognition. Complexity, 2021:4963903, Wiley/Hindawi, 2021.

[H12 - 30%] Patryk Pazura, Kamil Bortko, Jarosław Jankowski, and Radosław Michalski. A dynamic vote-rank based approach for effective sequential initialization of information spreading processes within complex networks. In ICCS 2020: International Conference on Computational Science, vol. 12137 of Lecture Notes in Computer Science, pp. 638–651. Springer, Cham, 2020.

[H13 - 100%] Radosław Michalski. Analysing and predicting the adoption of anonymous transactions in cryptocurrencies. In BIS 2020: International Conference on Business Information Systems, vol. 394 of Lecture Notes in Business Information Processing, pp. 132–144. Springer, Cham, 2020.

[H14 - 65%] Radosław Michalski, Jarosław Jankowski, and Patryk Pazura. Entropy-based measure for influence maximization in temporal networks. In ICCS 2020: International Conference on Computational Science, vol. 12140 of Lecture Notes in Computer Science, pp. 277–290. Springer, Cham, 2020.

W przedstawionym cyklu wątpliwości budzi podział udziałów w pracach np.: H7 – habilitant jest drugim autorem a przypisuje sobie 75% udziału w powstaniu publikacji. Poruszanie się po oświadczeniach współautorów jest utrudnione w sytuacji, gdy biorą oni udział w więcej niż jednej publikacji. Wynika to ze sposobu przedstawienia udziałów w publikacjach, co zrobione zostało dla każdego autora a nie dla każdej z publikacji oraz faktu, że w oświadczeniach wykorzystano numeracje ciągłą a nie tak jak miało to miejsce w indeksowaniu przedstawionym w cyklu [H1-H14].

Przedstawiony cykl jest logicznie podzielony na dwie grupy związane z rodzajem analizy wykonywanej w sieciach społecznościowych: wnioskowaniem i dyfuzją. W obrębie każdego z nich wyszczególniono osobne zagadnienia będące przedmiotem badań, których wyniki opisano w artykułach [H1-H14], uporządkowanych po dacie publikacji.

Poniżej omówię i skomentuję główny wkład Habilitanta w cyklu publikacji wymienionych jako osiągnięcie naukowe.

Artykuł H1 przedstawia autorski algorytm mający na celu poprawę pozyskiwania informacji w sieciach w przypadku brakującej wiedzy. Opracowane podejście MSCS (*Multistage Survey Control System*) wybiera węzły do próbkowania, tak by optymalizować czas do uzyskania reprezentatywnej próbki sieci. W autoreferacie zabrakło jednak opisu zaproponowanego podejścia, co jest ewidentnym niedopatrzeniem Habilitanta, w szczególności, że deklaruje on autorstwo opracowania założeń metody będących głównym jego wkładem w powstanie tej publikacji. Podejście to składające się z 5 kroków opisane zostało w artykule H1, w sekcji 3.

W artykule tym, Autor stara się odpowiedzieć na pytanie - „czy możliwe jest pozyskanie informacji o nieznanych atrybutach węzłów?". Jednak nie daje na nie satysfakcjonującej odpowiedzi, pisząc jedynie, że wnioski zawarte w tej pracy dają częściową odpowiedź na to pytanie. Nie zostało doprecyzowane tu czy jest ona twierdząca, pozostawiając w domyśle również ocenę jakości proponowanego podejścia.

Kolejna grupa artykułów [H7, H8, H13] kontynuuje temat odkrywania informacji o nieznanych etykietach węzłów skupiając się na dwóch rodzajach sieci społecznych: sieciach organizacyjnych i sieciach bazujących na łańcuchu bloków.

Zaznaczyć tu należy mylące używanie przez Habilitanta pojęć opisujących węzły, raz odnosi się on do atrybutów węzłów, w innych miejscach wskazuje na poszukiwanie nieznanych

etykiet. Jest to szczególnie niefortunne ze względu na używanie metod klasyfikacyjnych, gdzie typowo termin atrybuty wiązany jest z cechami obiektów a etykiety z ich klasami.

Artykuł H7 opisuje zastosowanie klasyfikacji kolektywnej do odkrywania nieznanych atrybutów węzłów. Podobnie jak wcześniej Habilitant przypisując sobie opracowanie metody klasyfikacji kolektywnej zupełnie nie opisał jej w autoreferacie, zaznaczając jedynie, że zyskała ona przewagę nad innymi przetestowanymi metodami uczenia: drzewami decyzyjnymi, lasami losowymi, siecią neuronową i SVM. Nie podano tu również jak zaproponowana metoda ma radzić sobie z przyjęciem założenia o niezależności zmiennych co Habilitant wskazał jako ograniczenie typowych metod klasyfikacyjnych.

Kolejne dwie prace H8 i H13 rozwijają tematykę „możliwości zastosowania uczenia nadzorowanego do wnioskowania o etykietach węzłów". Tu znowu widać błąd terminologiczny: metody uczenia nadzorowanego realizując zadanie klasyfikacji dokonują przypisania obiektów do wybranego podzbioru klas a nie realizują procesu wnioskowania.

Artykuły te skupiają się na analizie sieci Blockchain, w obrębie których dokonują klasyfikacji węzłów (w autoreferacie nie podano istotnych parametrów tego zadania w szczególności nie podano dokładnie do jakiej liczby klas następuje klasyfikacja – informacje te trzeba wydobywać z publikacji). Wskazane prace opisują badania nad poszukiwaniem metod klasyfikacji, które będą najbardziej użyteczne do analizy danych, które reprezentowane są, jak pisze Autor, z użyciem „wybranych cech" (nie zostało opisane jakimi cechami reprezentowano węzły) jak i analizą wpływu długości historii transakcji oraz wykorzystania tzw. osadzeń. Ostatnia z metod jest obecnie modnym podejściem i użycie jej do analizy jest prawdopodobnie dobrym wyborem, nie wskazano jednak na jego istotną przewagę nad innymi metodami.

W pracy H13 przeanalizowano przebieg adopcji transakcji anonimowych w kryptowalutach przez członków tej sieci oraz podjęta została próba oceny uwarunkowana jej czynnikami zewnętrznymi wynikającymi z rozwoju ekosystemu danej kryptowaluty. Dokonano analizy wpływu rozwoju takiego ekosystemu na popularyzację transakcji opartych na bazie dowodów z wiedzą zerową uniemożliwiających wykonywanie i autoryzowanie transakcji, których nie da się prześledzić. W pracy analizowano również możliwość prognozy z użyciem metod klasyfikacji czy wskazany użytkownik będzie usiłował przenieść środki do puli środków pozwalających na wykonywanie transakcji anonimowych. Dokonana analiza reprezentacji węzła wraz z metodami klasyfikacji dała możliwości przewidzenia zaangażowania adresów w transakcje z pulą anonimową.

Powyższe prace starają się rozwiązać problem powszechnie znany w uczeniu maszynowym tzw. *missing data inputation,* dla którego jest stosowane wiele podejść. Zaznaczyć należy, że w pracach przedstawionych w cyklu, nie znalazłem porównania do tej popularnej grupy metod.

Kolejnym tematem, którym zajął się Habilitant w ramach przedstawionego cyklu jest sposób pomiarów dynamiki interakcji pomiędzy węzłami. Wychodząc od wskazania problemu utraty informacji przy zapisie w sieciach statycznych jako wagi na krawędziach grafu, Autor by zachowywać kolejności interakcji pomiędzy węzłami opisywał je w rozszerzonej formie jako sekwencje kontaktów.

Podejście to wykorzystano w pracy H5, gdzie do kwantyfikacji dynamiki kontaktów węzłów w sieciach temporalnych, bazując na mierze entropii zaproponowane zostały trzy miary: pierwszego rzędu – entropię węzłową (w trzech wariantach), drugiego rzędu – entropię krawędziową, trzeciego rzędu – entropię następstwa oraz znormalizowanej miary.

Miary te umożliwiają określenie dynamiki i wymienność kontaktów pomiędzy węzłami w sieciach. Ocena wykonana została zarówno na rzeczywistych jak i na sztucznie generowanych danych. Podobnie jak w opisie wcześniejszych prac, esencja zaproponowanych podejść (wyniki ich oceny) zaznaczona w podziale prac jako „współudział w opracowaniu koncepcji" (a będąca kluczowym wkładem w powstanie artykułu) nie została jasno zawarta w autoreferacie a ich rezultatów trzeba szukać w załączonych publikacjach co sprawia wrażenie lakonicznego opisu indywidualnego wkładu Habilitanta.

Druga część cyklu związana z dyfuzją w sieciach podzielona jest na dwie części: w pierwszej Habilitant zajmuje się tematyką selekcji węzłów początkowych, w drugiej - metodami sekwencyjnej aktywacji węzłów.

Zadanie selekcji węzłów początkowych habilitant oparł na dwóch podejściach: algorytmach ewolucyjnych H6 i wyborze opartym o miary istotności informacyjnej wykorzystującej entropię H14.

Artykuł H6 opisuje algorytm genetyczny do wyboru węzłów początkowych. W autoreferacie nie opisano najważniejszych elementów algorytmu, co jest ponownie istotnym brakiem w sytuacji przedstawiania własnych osiągnięć Habilitanta. Opisane w samym artykule kolejne kroki algorytmu genetycznego od inicjalizacji do mutacji obrazują już lepiej zaproponowaną metodę, zaznaczyć należy, że najistotniejszy element z punktu widzenia procesu optymalizacji – funkcja celu nie jest wyraźnie zdefiniowana i trzeba się jej doszukiwać w sekcji 3.3 Social influence model and influence maximization.

Wnioski płynące z eksperymentów potwierdzają użyteczność zaproponowanej metody – jej dobrą zbieżność do rozwiązania a przez to akceptowalny koszt obliczeń. Drugą bardzo interesującą obserwacją jest możliwość transferu parametrów algorytmu pomiędzy sieciami.

W pracy zabrakło mi uzasadnienia dlaczego akurat wybrane zostało podejście ewolucyjne do selekcji węzłów. Wiadomo, iż metoda ta jest rodzajem przeszukiwania na ślepo z heurystyką poprawiającą czas odnalezienia rozwiązania, jednakże jej zastosowanie ma sens, gdy inne algorytmy optymalizacyjne nie są możliwe. Brak dyskusji wyboru algorytmu optymalizacyjnego i wybór od razu podejścia genetycznego, uważam za istotny mankament tej pracy.

W artykule H14 wykorzystano miary bazujące na miarach entropii opisanych w H5 na podstawie, których zaproponowano podejście bazujące na wymienności sąsiedztwa w sieciach temporalnych. Pomimo braku determinizmu metody (co również powinno być zaznaczone w przypadku podejścia ewolucyjnego) ustalona liczba powtórzeń (opisana w pracy H9) dająca satysfakcjonujące przybliżenie, potwierdza reprezentatywność zaproponowanego podejścia.

Pewien niedosyt budzi brak przeprowadzenia porównania zaproponowanych metod, co powoduje, że podejścia te są zawieszone niejako w próżni – z jednej strony Habilitant pisze, że obie „wzbogacają zestaw metod selekcji węzłów" i są one akceptowalne obliczeniowo, z drugiej strony brak porównania i określenia cech każdej z nich nie pozwala na jednoznaczne stwierdzenie, która z nich jest lepsza lub też do jakiej klasy problemów warto je stosować.

Ostatnia część cyklu skupia się na metodach sekwencyjnej aktywacji węzłów. W pięciu pracach związanych z tą częścią przebadano trzy typy sieci: statyczne, temporarne i wielowarstwowe. Do istotnych rozwinięć metody dyfuzji zaliczyć należy podejście, w którym na początku nie są aktywowane wszystkie wybrane węzły a aktywacja ta odbywa się wraz z procesem dyfuzji co poprawia całkowity efekt dyfuzji.

W pracy H2 zaproponowano metodę polegającą na równoważeniu prędkości aktywacji i pokrycia sieci. Wprowadzenie możliwości balansowania tych dwóch parametrów jest istotne z punktu widzenia rezultatu końcowego jaki został założony. Przeanalizowanych zostało pięć scenariuszy na 15 sieciach, w znaczącej większości przypadków (85%) zaproponowane podejście okazało się lepsze niż jednoczesna aktywacja wszystkich węzłów z zestawu węzłów początkowych. Rozwinięcie badań w tym obszarze o sieci skierowane opisano w pracy H4. Istotnym wynikiem zaprezentowanym tutaj jest analityczny dowód przewagi zaproponowanego podejścia nad aktywacjami pełnymi.

Ważnym elementem pracy jest również usystematyzowanie podejścia do porównania rozpatrywanych metod dyfuzji dla niedeterministycznego modelu, dzięki czemu możliwe jest

wykonanie lepszej analizy porównawczej metod. W ramach pracy wskazano górną granicę możliwości aktywacji w najlepszym koszyku węzłów początkowych i określono jak ma się do tego opracowana metoda.

Dodatkowo na siedmiu rzeczywistych sieciach eksperymentalnie wykazano, że sekwencyjna aktywacja węzłów ze stopniem węzła jako metodą selekcji jest skuteczniejsza od bardziej złożonej obliczeniowo metody zachłannej z jednoczesną aktywacją.

Wyznaczanie rankingów do wyboru węzłów początkowych do sekwencyjnej aktywacji węzłów w sieciach statycznych było przedmiotem pracy H12. Podejście to rozwija metodę stosowaną w poprzednich dwóch pracach, gdzie selekcji węzłów dokonywano jednokrotnie przed rozpoczęciem procesu aktywacji. W pracy przedstawiono wyniki wpływu zmian wyznaczania heurystyki vote-rank (zrealizowanej w jej wersji dynamicznej) badając sposób wyznaczenia miary przed aktywacją kolejnych węzłów oraz zależność pomiędzy częstotliwością przeliczania miary a efektem końcowym. Zaproponowane podeście dało poprawę wyników do 10% w porównaniu do sytuacji, gdy metoda ta nie była stosowana.

W H9 opisano metodę sekwencyjnej aktywacji węzłów dla sieci temporalnych w wariancie okien czasowych dla modelu dyfuzji *independent cascade* IC. Porównano tu metodę jednoczesnej aktywacji wszystkich węzłów i sekwencyjnej aktywacji dla heurystyki wykorzystującej stopień węzła. Uzyskane wyniki na dwóch zbiorach danych zawierających informacje o korespondencji mailowej pomiędzy pracownikami firmy i kontaktami pomiędzy ludźmi pokazały, że w zdecydowanej większości przypadków metoda sekwencyjnej aktywacji jest tak samo dobra lub lepsza niż metoda jednoczesnej aktywacji węzłów. Eksperymenty na modelach losowych potwierdziły tą obserwację. Istotne jest również udostępnienie kodu eksperymentów jak i danych przez Code Ocean repository. Zaznaczyć należy, że opis w autoreferacie jest bardzo pobieżny i bez odnoszenia się do publikacji nie jest możliwe ocenienie istotności opisanych tam wyników.

Praca H10 proponuje adaptację podejścia sekwencyjnej aktywacji węzłów w sieciach wielowarstwowych. Przedstawiono tu wyniki uzyskane dla 10 różnych typów sieci zarówno rzeczywistych jak i sztucznie generowanych, z modelem dyfuzji IC. Habilitant przypisuje sobie udział „w opracowaniu koncepcji dla sieci wielowarstwowych i planu eksperymentów, weryfikacja wyników eksperymentów" jednak w autoreferacie o samej metodzie mającej być jego indywidualnym wkładem nie jest nic napisane a główny opis jest związany z wnioskami płynącymi z pracy. Przypuszczam, że wyciągnięcie wniosków i dyskusja uzyskanych wyników jest podciągnięta pod wkład w pracę opisany jako „weryfikacja wyników eksperymentów", jednak jest to kolejny przykład mało precyzyjnego opisu podziału wkładu autorów w powstanie prac.

Przedstawiony cykl wiąże tematyka analizy danych w sieciach społecznościowych, co w dobie intensywnego rozwoju tego typu sieci w internecie jest tematyką aktualną i istotną.

W skład cyklu wchodzi 11 artykułów indeksowanych na liście JCR oraz 3 wysoko punktowane artykuły konferencyjne indeksowane w WoS. Przedstawione w cyklu publikacje w większości są wieloautorskie, jedynie jedna konferencyjna praca jest w pełni indywidualnym dziełem Habilitanta.

Sposób, w jaki zostało przedstawione dość obszerne osiągnięcie naukowe utrudnia ocenę indywidualnego wkładu Habilitanta. W szczególności wynika to z faktu, że większość zawartych w cyklu publikacji ma wielu autorów a w autoreferacie nie jest jasno wskazane co jest indywidualnym wkładem Habilitanta, opisany jest tam ogólnie zawartość całego artykułu. Biorąc dodatkowo pod uwagę liczbę wysoko punktowanych artykułów przedstawionych w cyklu, w mojej ocenie osiągnięcie spełnia wymagania stawiane pracom habilitacyjnym.

Zaznaczyć należy, że pomimo kilku uwag zgłoszonych w recenzji - cykl prezentuje wartościowy wkład w rozwój nauki. W mojej opinii spełnia on wymogi określone w ustawie „Prawo o szkolnictwie wyższym i nauce" i oceniam go pozytywnie.

2. Ocena istotnej aktywności naukowej

Poniżej podsumuję i omówię główne pozostałe naukowe, organizacyjne i dydaktyczne aktywności Habilitanta.

Sposób przedstawienia we wniosku pozostałych prac Habilitanta zdecydowanie utrudnia ocenę całościowego dorobku – w wykazie artykułów cytowanych w autoreferacie przemieszane są publikacje Habilitanta i prace stanowiące tło bibliograficzne, co np. można było rozwiązać dobierając metodę indeksowania lub rozdzielając te dwie grupy artykułów.

W bazie WoS zaindeksowanych jest 35 artykułów, których Habilitant jest autorem. 16 z nich stanowią publikacje w renomowanych czasopismach posiadających IF. W mojej opinii liczba publikacji, jak i liczba ich cytowań jest w pełni satysfakcjonująca dla kandydata do stopnia doktora habilitowanego.

H-index Habilitanta w/g WoS wynosi 8, co mieści się w średniej dla osób wnioskujących o stopień doktora habilitowanego.

Kandydat był wielokrotnie (23 razy) nagradzany za swoją działalność naukową.

Habilitant prowadził działalność dydaktyczną i badawczą na dwóch uczelniach, realizując zajęcia w Wyższej Szkole Informatyki i Zarządzania „Copernicus" we Wrocławiu oraz na

Politechnice Wrocławskiej, na Wydziale Informatyki i Zarządzania. We wniosku Habilitant w sekcji 3 „Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych lub artystycznych" wykazał również zatrudnienie w Hutmen S.A. we Wrocławiu, jak mniemam, nie jest to jednostka naukowa lub artystyczna a firma komercyjna.

Wymienione w tym punkcie miejsca zatrudnienia spełniają formalnie zdefiniowany w ustawie wymóg „wykazuje się istotną aktywnością naukową albo artystyczną realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej" Co więcej Habilitant odbył szereg staży naukowych w szczególności w ośrodkach zagranicznych.

Habilitant brał również udział w 10 projektach badawczych w jednym z nich jako kierownik, jak również zaznaczyć należy jego udział w 6 projektach B+R, gdzie w dwóch pełnił rolę kierownika.

W swojej działalności dydaktycznej Kandydat wymienia opracowanie nowych kierunków studiów oraz prowadzenie szeregu kursów, jak również opiekę naukową nad 19 pracami inżynierskimi i 18 magisterskimi. Sprawuje również opiekę nad grupą naukową BERG – Blockchain Exploration Research Group.

Habilitant był również promotorem pomocniczym obronionego z wyróżnieniem w 2020 roku doktoratu, a obecnie jest promotorem pomocniczym w kolejnym przewodzie. Obrazuje to jego dobre przygotowanie do samodzielnego prowadzenia młodych pracowników naukowych.

Przedstawiona przez Kandydata pozostała aktywność naukowa i dydaktyczna, w mojej opinii, w pełni spełnia wymagania stawiane w stosownej ustawie i oceniam ją pozytywnie.

3. Wniosek końcowy

Biorąc pod uwagę przedstawiony cykl publikacji pt. „Analiza sieci społecznych: próbkowanie, klasyfikacja i dyfuzja " oraz pozostałe aktywności naukowe Kandydata uważam, że spełniają one kryteria określone w stosownej ustawie do nadania stopnia doktora habilitowanego nauk inżynieryjno – technicznych w dyscyplinie informatyka techniczna i telekomunikacja i oceniam je pozytywnie. W związku z tym, wnoszę o dopuszczenie Kandydata do dalszych etapów procedury habilitacyjnej.

dr hab. inż. Julian Szymański, prof. PG