

**Recenzja rozprawy doktorskiej mgra mgra inż. Damiana Frączczaka  
pt. „Wykrywanie źródeł fałszywych wiadomości w sieciach społecznościowych”**

Recenzowana rozprawa powstała w Szkole Doktorskiej Wojskowej Akademii Technicznej w Warszawie, pod kierunkiem prof. Ryszarda Antkiewicza i przy współpracy dra inż. Mariusza Chmielewskiego. Tekst główny rozprawy ma objętość ok. 180 stron i jest podzielony na 9 rozdziałów.

Autor rozprawy stawia tezę, przytoczoną w całości na str. 177 rozprawy, że „możliwa jest realizacja efektywnej i skutecznej metody wykrywania źródeł fałszywych informacji w sieciach o dowolnej strukturze, przy założeniu możliwości istnienia wielu źródeł informacji, bez możliwości uzyskania całego grafu propagacji informacji oraz braku konieczności arbitralnego ustalenia liczby źródeł informacji”. Teza ta dotyczy ważnego i aktualnego obszaru badań naukowych obejmującego tematykę procesów rozprzestrzeniania się informacji w sieciach społecznościowych. Szczególnie propagacja informacji nieprawdziwych, niejednokrotnie wprowadzanych celowo do sieci i mogących potencjalnie powodować daleko idące skutki, budzi zainteresowanie badaczy w wielu ośrodkach naukowych na świecie. Warto zaznaczyć, że zagadnienia analizowane w rozprawie mogą w szerszym ujęciu służyć także do badania zjawisk np. rozchodzenia się zagrożeń epidemicznych albo informacji marketingowych. Tak więc, praktyczna użyteczność tematyki badań nie budzi wątpliwości u recenzenta.

Teza rozprawy sformułowana jest jasno, choć zdaniem recenzenta nieco myląco. Doktorant przeprowadził w rozprawie obszerną analizę prac innych autorów dotyczących tej tematyki (spis literatury liczy 156 pozycji), w tym również prac podejmujących tematykę podobną do tej, jaką zajmował się autor rozprawy. Przegląd niektórych aspektów tych technik podano w m.in. w rozdziale 5.1 rozprawy, gdzie wypisano kilkanaście wariantów rozwiązań dotyczących identyfikacji wielu źródeł informacji wprowadzonej do sieci. Tak więc, zdaniem recenzenta, bardziej trafne byłoby takie sformułowanie tezy rozprawy, w którym nacisk położono by na poprawę niektórych parametrów procesu wyszukiwania źródeł informacji w sieci (służy temu m.in. zaproponowana w rozdziale 5.4 metoda BLOCD identyfikacji ognisk propagacji), aniżeli na samą tylko możliwość realizacji tego procesu.

Pierwsze trzy rozdziały rozprawy (około 70 stron) zawierają przegląd literatury. Jest on wykonany poprawnie, bibliografia jest obszerna, aktualna i adekwatna do tematu rozprawy. Praca zyskałaby na czytelności gdyby ta część rozprawy była bardziej zwięzła, gdyż zdaniem recenzenta niektóre fragmenty mają wręcz naturę podręcznikową (np. początek rozdziału 2). Nie ulega jednak wątpliwości, że doktorant ma dużą wiedzę i swobodnie porusza się po tematyce rozprawy.

Rozdział 4 zawiera opis autorskiej platformy symulacyjnej, która posłużyła do wykonania badań. Została ona zaprojektowana i zrealizowana tak, że możliwe jest symulowanie zarówno sieci rzeczywistych, pobieranych do układu symulacyjnego z zewnętrznej bazy danych jak i sieci sztucznie zdefiniowanych na potrzeby badań opisywanych w rozprawie. Obie te możliwości wykorzystano w rozprawie (łącznie przeanalizowano 10 sieci, zebranych w tabeli 11). W środowisku symulacyjnym zrealizowano kilkanaście znanych z literatury rozwiązań algorytmów identyfikacji ognisk fałszywych

wiadomości, co pozwoliło porównać rozwiązanie zaproponowane w rozprawie z propozycjami innych autorów. Zaprojektowanie, wykonanie i przetestowanie takiego środowiska symulacyjnego to istotny i wartościowy składnik badań opisanych w rozprawie.

W rozdziale 5, najobszerniejszym w całej rozprawie, umieszczono propozycję autorskiego algorytmu (tzw. algorytm BLOCD) detekcji ognisk propagacji oraz wyniki symulacji porównujących skuteczność działania tego algorytmu z innymi algorytmami opisywanymi w literaturze. Ta część rozprawy nie budzi zastrzeżeń recenzenta, zarówno zakres wykonanych prac jak i metodyka wyglądają solidnie. Parametry symulacji dobrano w sposób podobny do tego, w jaki symulacje przeprowadzano w pracach innych autorów. W rozprawie dokonano szczegółowej analizy uzyskanych wyników, po czym sformułowano wnioski. Zaproponowany przez autora rozprawy algorytm BLOCD detekcji ognisk propagacji dobrze wypada na tle innych algorytmów znanych z literatury. Widać także wysiłek autora aby stworzyć syntezę wyników. Nadal jednak, zdaniem recenzenta, większa selektywność przy wyborze wyników jakie umieszczono w całości w rozprawie, poprawiłaby jej czytelność (dotyczy to np. tabel 22-26 lub rysunków 71-87).

Kolejny szósty rozdział zawiera dalszy ciąg oryginalnych wyników pracy badawczej wykonanej przez autora rozprawy. Tym razem dotyczy procesów odtwarzania ścieżek propagacji informacji. Także i w tym rozdziale autorski algorytm BLOCD porównywany jest do innych algorytmów cytowanych w literaturze. Uzyskane wyniki potwierdzają wnioski z rozdziału poprzedniego o tym, że w wielu sytuacjach użycie algorytmu BLOCD daje wyniki lepsze niż pozostałe algorytmy. W tym rozdziale pojawiają się też stwierdzenia o dłuższym czasie potrzebnym na identyfikację źródeł propagacji w przypadku stosowania algorytmu BLOCD, w porównaniu z innymi algorytmami. To typowa sytuacja często występująca w pracach symulacyjnych: z jednej strony kolejne algorytmy są coraz bardziej skomplikowane, a z drugiej strony korzyści wynikające z ich stosowania widoczne są tylko niekiedy, w niektórych sytuacjach, np. tylko przy niektórych kombinacjach założeń do symulacji. W skrajnych przypadkach może to w ogóle przekreślić praktyczny sens używania konkretnego algorytmu. Te aspekty, dotyczące „kosztów” stosowania lepszego algorytmu, mogłyby zdaniem recenzenta być obszerniej omówione w rozprawie. Podobnie, autor rozprawy w kilku miejscach rozprawy podejmuje zagadnienie złożoności obliczeniowej proponowanych przez siebie rozwiązań ale w odczuciu recenzenta często w tej analizie brakuje zakończenia, podsumowującego stwierdzenia, np. porównania ze złożonością obliczeniową innych algorytmów (dotyczy to m.in. podrozdziału 5.5, w całości dotyczącego kwestii złożoności obliczeniowej, ale także kilku innych miejsc w rozprawie, np. podrozdziału 6.2).

Trzecią część swoich badań autor rozprawy opisuje w rozdziale 7, który zawiera propozycję autorskiej metody selekcji źródeł wiadomości w oparciu o specyficznie dobierany próg decyzyjny. Na początku rozdziału zdefiniowano sposób dobierania wartości progu decyzyjnego, a następnie przedstawiono obszerne fragmenty wyników liczbowych symulacji. Uzyskane przez doktoranta wyniki wskazują na to, że sformułowana przez niego modyfikacja algorytmu w niektórych przypadkach poprawia jego działanie. Ciekawą analizę zawiera dalszy ciąg rozdziału 7 poświęcony uczeniu zespołowemu, tj. technice polegającej na łączeniu wyników predykcji kilku modeli w celu uzyskania dokładniejszego wyniku. Przypomina to – jak podaje autor rozprawy – podejście wykorzystujące „mądrość tłumu”. W tej części pracy autor rozprawy sięga ponownie do uzyskanych przez siebie wcześniej wyników i używa równocześnie wyników kilku algorytmów. Okazuje się jednak, że w przypadkach analizowanych w rozprawie uczenie zespołowe nie przynosi istotnych korzyści.

Na końcu rozprawy, w rozdziale 8, opisano ciekawy i atrakcyjnie przedstawiony przykład propagacji w rzeczywistej sieci społecznościowej fałszywej informacji dotyczącej rzekomego zarażenia się papieża Franciszka wirusem COVID w 2020 roku. Dokonana przez autora rozprawy analiza tego przypadku oraz starannie wykonane rysunki poglądowe podnoszą wartość pracy.

Recenzent rozprawy dostrzegł także słabsze jej elementy. O niektórych wspomniano już we wcześniejszej części recenzji. Bez wątplenia redakcja i edycja pracy mogłaby być lepsza. Z jednej strony, czytanie rozprawy jest żmudne. Z oczywistych powodów autor rozprawy używa bardzo wielu akronimów, np. stanowiących nazwy poszczególnych algorytmów lub parametrów, ale choć w rozprawie umieszczono spis oznaczeń (rozdz. 10) to nie zamieszczono spisu akronimów. Stanowi to dodatkowe wyzwanie przy czytaniu rozprawy. W tekście sporo jest niedociągnięć edytorskich (np. pominięcia lub powtórzenia słów), zdarzają się nawet niekiedy błędy ortograficzne (łączne lub osobne pisanie słów). Kontrowersje u recenzenta budziło całkowite zdanie się przez autora rozprawy na automatyczną numerację tabel dostępną w zastosowanym przez niego edytorze tekstu, co skutkowało zdaniami typu (str. 117) „(wyniki) umieszczono w *Tabela 20*” zamiast „umieszczono w Tabeli 20” lub „umieszczono w Tab. 20”.

Co do merytorycznych aspektów pracy, niedosyt u recenzenta budził zupełny brak w rozprawie porównań bezpośrednich z wynikami badań przeprowadzonych przez innych autorów. Rozwiązania proponowane przez doktoranta obszernie porównywano z algorytmami stosowanymi przez innych autorów, ale porównania dokonywano poprzez zasymulowanie innych algorytmów we własnym środowisku symulacyjnym. Czy wspomniane wcześniej porównania bezpośrednie z wynikami badań innych autorów były niemożliwe? Zbyt żmudne do uzyskania? Brak było kompletnych danych dotyczących warunków symulacji? Nawet jeżeli tak było, to zdaniem recenzenta wartościowe byłoby choćby zasygnalizowanie tej kwestii przez autora rozprawy.

Podsumowując, w opinii recenzenta w rozprawie zostały opisane wartościowe i potencjalnie użyteczne badania naukowe. Uzyskano oryginalne wyniki i poprawnie je przeanalizowano. Osiągnięcia doktoranta zostały trafnie zebrane w 7 punktach w podsumowaniu rozprawy (str. 177, a także rys. 100 na str. 168). Zdaniem recenzenta opiniowana rozprawa stanowi istotny wkład do literatury przedmiotu i spełnia warunki stawiane rozprawom doktorskim.