

dr hab. inż. Krzysztof Malarz, prof. AGH
Akademia Górniczo-Hutnicza w Krakowie
Wydział Fizyki i Informatyki Stosowanej
Katedra Informatyki Stosowanej i Fizyki Komputerowej
Zespół Układów Złożonych

Kraków, 21 lutego 2022

Recenzja rozprawy doktorskiej mgr. inż. Łukasza G. Gajewskiego

1 Informacje wstępne

Przedstawiona do recenzji rozprawa doktorska mgr. inż. Łukasza G. Gajewskiego zatytułowana „Reverse engineering multi-layered structures in complex networks” powstała na Wydziale Fizyki Politechniki Warszawskiej pod opieką prof. dr. hab. inż. Janusza Hołysta przy wsparciu dr. inż. Juliana Sienkiewicza jako promotora pomocniczego. Badania były współfinansowane z kilku grantów, w tym z europejskiego grantu „Horizon 2020” czy grantu badawczego Narodowego Centrum Nauki. Wyniki stanowiące przedmiot rozprawy zostały opublikowane w trzech artykułach w prestiżowym czasopiśmie *Physical Review E*:

- Ł. G. Gajewski, J. Chołoniewski, M. Wilinski, „Detecting hidden layers from spreading dynamics on complex networks”, *Physical Review E* **104**(2), 024309 (2021);
- Ł. G. Gajewski, J. Sienkiewicz, J. A. Hołyst, „Discovering hidden layers in quantum graphs”, *Physical Review E* **104**(3), 034311 (2021);
- Ł. G. Gajewski, J. Sienkiewicz, J. A. Hołyst, „Transitions between polarisation and radicalisation in a temporal bi-layer echo chambers model”, *Physical Review E* **105**(2), 024125 (2022).

W pracach tych kandydat do stopnia doktora jest pierwszym autorem co może wskazywać na jego wiodącą rolę w przygotowaniu tych publikacji.

Pracę poświęcono wykrywaniu i rekonstrukcji topologii warstw ukrytych w sieciach wielowarstwowych z wykorzystaniem technik inżynierii wstecznej. Przedstawiono w niej również znaczenia tych technik oraz samych sieci wielowarstwowych. Do rekonstrukcji ukrytych warstw w różnorodnych sieciach (tak rzeczywistych jak i wyhodowanych *in silico*) wykorzystano procesy propagacji na nich epidemii zakażeń, pakietów falowych oraz wiadomości (opinii).

Praca napisana jest w języku angielskim, liczy sto szesnaście stron i została strukturalnie podzielona na pięć rozdziałów. Ich układ jest typowy dla tego typu prac: ze wstępem, trzema rozdziałami poświęconymi tematyce trzech wyżej wskazanych artykułów oraz podsumowaniem. Pracę kończy blisko dwustu elementów spisu bibliografii.

2 Ocena formalna

Zgodnie z artykułem 187, punkt 1, ustawy „Prawo o szkolnictwie wyższym i nauce” (Dz.U.2020.85) rozprawa doktorska powinna prezentować ogólną wiedzę teoretyczną kandydata do stopnia w dyscyplinie, w której stopień ma być nadany oraz jego umiejętność do samodzielnego prowadzenia pracy naukowej. Bez wątplenia przedstawiona praca spełnia oba te warunki.

Przedmiotem tej rozprawy doktorskiej jest (znów ustawowo dozwolone w artykule 187, punkcie 2) oryginalne rozwiązanie problemu naukowego postawionego jasno na piętnastej stronie pracy. Na podkreślenie zasługuje fakt oparcia całości pracy na trzech pracach opublikowanych w renomowanym czasopiśmie *Physical Review E*, których mgr inż. Łukasz Gajewski jest pierwszym współautorem.

Zgodnie z artykułem 187, punkt 3, „rozprawę doktorską może stanowić praca pisemna, w tym monografia naukowa, zbiór opublikowanych i powiązanych tematycznie artykułów naukowych, praca projektowa, konstrukcyjna, technologiczna, wdrożeniowa lub artystyczna, a także samodzielna i wyodrębniona część pracy zbiorowej.” Przedstawiona dysertacja ma formę pracy pisemnej — opartej na zbiorze opublikowanych i powiązanych tematycznie artykułów naukowych — wydanej w formie książkowej.

Ciążący na Autorze obowiązek z artykułu 187, punkt 4, przygotowania streszczenia pracy w języku polskim został dopełniony.

3 Ocena merytoryczna

W rozdziale pierwszym, po wstępnym przedstawieniu celów pracy i zaprezentowaniu ich na tle historycznym, kolejno przedyskutowano przykłady, sposób reprezentacji i cechy charakterystyczne sieci złożonych (w tym klasycznych grafów przypadkowych, sieci Watta–Strogratza, rosnących sieci bezskalowych i sieci wielowarstwowych). Następny punkt poświęcono modelowaniu rozprzestrzenienia się choroby zakaźnej (albo ogólniej mówiąc informacji) na przykładzie klasycznego szufladkowego modelu agentowego SI („Susceptible–Infected”). Kolejny punkt dotyczy grafów kwantowych to jest grafów z nałożoną na nie metryką oraz hamiltonianem. Wreszcie punkt 1.4 opisuje trzy wykorzystywane w pracy techniki: jedną z technik analiz czynnikowych (analizę głównych składowych), jedną z technik badania sygnałów w domenie częstotliwościowej

(szybką transformację Fouriera) oraz technikę do badań czasowo-częstotliwościowych (analizę falkową).

W rozdziale drugim Autor stawia i próbuje odpowiedzieć na dwa pytania: *i*) czy obserwując rozprzestrzeniające się w sieci złożonej kaskady informacji można ustalić, czy istnieją kanały komunikacji, które są przed nami ukryte a jeśli tak, *ii*) to czy jest szansa na ich ujawnienie i rekonstrukcję?

Do tego celu Autor wykorzystał model kompartmentowy propagacji zakażenia z modyfikacją literaturowych wzorów na prawdopodobieństwa propagacji zakażenia (sygnału) w sieciach. Proces sukcesywnego przekazywania informacji między węzłami określono mianem „kaskady” co prawdopodobnie ma podkreślić dynamikę procesu propagacji informacji. Tu chciałbym się na chwile zatrzymać nad zdaniem z trzydziestej szóstej strony pracy: „As seen from Eq. (2.1), adding more cascades reduces the likelihood”. Pomijając, że formalnie czytelnik nie zgadnie „likelihood” czego Autor miał na myśli, to w połączeniu z kolejnym zdaniem (którego znaczenie czytelnik ma obiecać „zobaczyć później” i groźba ta jest realizowana w okolicach pięćdziesiątej strony pracy) chętnie usłyszałbym rozwinięcia tezy o jednoczesnym dość przeciwnym charakterze mnożenia liczby kaskad na jakość uzyskiwanych wyników.

Na stronach 43–44 przedstawiono algorytm wykrywania ukrytych wiązań przy założeniu znajomości przynajmniej intensywności propagacji informacji w warstwie ukrytej. Z jego wykorzystaniem pokazano, że na skuteczne odnajdywanie połączeń w ukrytych warstwach sieci ma gęstość połączeń zarówno w ukrytej jak i tej jawnie obserwowanej warstwie sieci wielowarstwowej — no i to dziwić nie może. Ale w szczególności pokazano, że im lepiej połączona jest warstwa ukryta, tym trudniej jest zrekonstruować połączenia w niej istniejące. Ten wniosek dla odmiany wydaje się sprzeczny z intuicją: można się raczej było spodziewać, że trafienie (nawet przypadkowo) w niewidoczne wiązanie warstwy ukrytej jest coraz łatwiejsze ze wzrostem usieciowienia warstwy stając się pewnym dla topologii warstwy ukrytej odpowiadającej grafowi zupełnemu. Czy mogę liczyć na wskazanie błędu w moim rozumowaniu przytoczonym w poprzednim zdaniu?

Druga konkluzja z tego rozdziału — mówiąca, że wraz ze wzrostem liczby połączeń na widocznej warstwie jesteśmy w stanie zidentyfikować średnio mniej ukrytych krawędzi — też do intuicyjnych nie należy. Chciałbym usłyszeć obronę tych dwóch tez przy pierwszej nadarzącej się ku temu okazji.

W rozdziale trzecim Autor skupił się na analizie propagacji pakietu falowego na grafie kwantowym. Kolejno Autor omawia pierwotne ukształtowanie i propagację pakietu falowego, przyjęte warunki brzegowe, metodologię Aziza et al. pozwalającą na rozróżnianie między różnymi rodzajami grafów, nowatorski sposób wykrywania ukrytych warstw w sieciach (bazujący na analizie fourierowskiej sygnału powsta-

łego z sumowania amplitud pakietów w dyskretnych chwilach czasowych), sposób rekonstrukcji wartości własnych macierzy połączeń.

Rozdział kończą rozważania dotyczące propagacji pakietu na grafie zupełnym oraz sprzężonej parze takich grafów. Rozumiem atrakcyjność tych rozważań w kontekście dostępnych rozwiązań analitycznych ale absolutnie nie potrafię ich powiązać ze stanowiącym cel pracy wykrywaniem i rekonstrukcją topologii warstw ukrytych w sieciach wielowarstwowych. Co może być ukrytego w topologii grafu w pełni połączonego?

W rozdziale czwartym Autor analizuje możliwe implikacje wynikające z wprowadzenia dodatkowej ukrytej warstwy do układu z dynamiką opinii społecznej modelowanej układem równań różniczkowych opisujących intensywność (ciągłej) opinii „za” lub „przeciw” czemuś. Układ opisany takimi równaniami dla pojedynczej warstwy dąży do jednego z trzech stanów stacjonarnych: polaryzacji, radykalizacji bądź neutralizacji opinii.

Wśród przedstawionych tez szczególnie atrakcyjna z punktu widzenia socjofizyki (a może nawet inżynierii społecznej) jest ta pokazująca niszczycielską (czy jak to woli sprawczą) rolę propagandy w zmianie średniej opinii w układzie oraz gwałtowność tej zmiany.

Bezpośrednio przed równaniem (4.6) wprowadzona zostaje wielkość X , którą wspiera (bądź przeciwstawia się jej) zewnętrzny trend (o charakterze zewnętrznego pola magnetycznego w układach magnetycznych a propagandy w układach społecznych) B . Czym jest ten X pozostanie tajemnicą do końca pracy.

Pracę kończy bardzo zwarte podsumowanie i spis literatury.

Ponieważ zawartość merytoryczna pracy została już oceniona przez recenzentów czasopisma, w którym opublikowano przedstawiony w niej materiał, nie ma potrzeby ponownego dogłębnego i szczegółowego powtórnego analizowania poprawności przedstawionych wywodów. Z obowiązku recenzenta zwrócę jednak uwagę na kilka wątpliwości, które nasunęły mi się w czasie lektury pracy.

Na stronie 25 spotykamy typową definicję grafu: „Let $G = (V, E)$ be a graph” — ale próżno czytelnik poszukuje tam informacji, o tym, że V i E są odpowiednio zbiorami węzłów i wiązań między nimi. O ile oczywiście prawidłowo odgadłem zamysł Autora.

W równaniu (1.5) v jest określone jako należące do grafu G czyli do uporządkowanej pary dwóch zbiorów (węzłów i krawędzi). To czym właściwie jest v skoro warunek brzegowy w równaniu (1.5) definiowany jest dla każdego v ? Zakładając, że są to wierzchołki to po czym odbywa się sumowanie w tym równaniu? Wygląda, że po e — czyli krawędziach, których jednym z końców jest v . Aż boję się zapytać czym jest operator gradientu operujący na funkcji $u(e, x_{e,v})$, której pierwszy argument

jest krawędzią a drugi prawdopodobnie jakąś modyfikacją współrzędnej wzdłuż tej krawędzi. Piszę „prawdopodobnie”, bo symbol $x_{e,v}$ pozostaje do tego momentu lektury pracy niezdefiniowanym (sam symbol x_e został wcześniej zarezerwowany na oznaczenie współrzędnej na krawędzi e). Teraz przechodzimy do równania (1.6), w którym kształtowany jest gaussowski pakiet falowy mający propagować się po grafie. Być może oczywiste jest znaczenie symboli a i μ wpływających na kształt (szerokość) i początkowe położenie tego pakietu na krawędzi e — no ale nie jest rolą czytelnika pracy odgadywanie intencji Autora.

Kolejna uwaga dotyczy również zbioru Ω , do którego należenie bądź nienależenie jest dyskutowane i wskazywane w indeksach sumacyjnych kilku równań. Czym jest ten zbiór w punkcie poświęconym grafom kwantowym niestety nie napisano.

Następnie, Autor anonsuje, że będziemy poszukiwać wielkości C_π . By się zorientować czym jest C_π musimy przebrnąć przez kolejne cztery strony. Ale nurtująca nas ciekawość zostanie zaspokojona tylko częściowo, gdyż finalnie Autor zdecyduje się na zmianę oznaczenia tej wielkości i w równaniu (1.17) odkryjemy, że poszukiwaną wielkością jest wektor własny \mathbf{C}_π pewnej macierzy stowarzyszony z jej wartością własną $\lambda = -1$. Osobiście, zamieniłbym tu kolejność akcentów: zaczynając od tego czego i po co szukamy a dopiero potem opisując jak to można osiągnąć.

Na stronie pięćdziesiątej siódmej czytamy, że prędkość rozprzestrzeniania się pakietów falowych na grafach kwantowych jest skończona w przeciwieństwie do propagacji „informacji” w modelu epidemiologicznym. Być może mogą sobie wyobrazić intencje Autora: ale dla pewności wolałbym usłyszeć rozwinięcie tej tezy. Bo czyż prędkość propagacji zakażeń w kaskadzie zarażeń nie jest bądź przynajmniej nie może być również skończona? Wydaje się, że zarówno w stochastycznym modelu agentowym jak i w rzeczywistości front zakażenia przesuwa się po sieci ze skończoną prędkością. Na sieci wyhodowanej *in silico* propagacja ta taktowana jest dyskretnie przejściami od „pacjenta zero” przez stopniowo zarażanych agentów z jego kolejnych stref koordynacyjnych. A o konsekwencjach (komforcie) dla systemu opieki zdrowotnej wynikających ze skończonego tempa rozprzestrzeniania się epidemii w świecie realnym nie muszę chyba nikogo w drugim roku pandemii przekonywać.

4 Ocena redakcyjna

Praca napisana jest dość poprawną angielszczyzną (na ile recenzent potrafi to ocenić) ale przy uważniejszej redakcji udałoby się z pewnością uniknąć co najmniej kilku niezręczności. By nie być gołosłownym pozwolę sobie jawnie wskazać tylko niektóre z nich zaczynając od tych, które udało mi się znaleźć na jednej tylko (dwudziestej) stronie pracy:

- „Each element A_{ij} represents whether there is a connection $A_{ij} > 0$ or not $A_{ij} = 0$ ”. Pewnie byłoby lepiej, gdyby czytelnik wiedział, że i oraz j numerują

węzły i niezerowa wartość A_{ij} rzutuje na istnienie wiązania pomiędzy węzłami o etykietach i i j .

- Dalej Autor posługuje się pojęciem „node degree”, które zdefiniowane jest („that is the number of connections a node has”) pół strony dalej.
- Definicja sieci kwadratowej jako „a square lattice — each vertex is connected to four other vertices (up, down, left, right)” także pozostawia wiele do życzenia: od braku osadzenia jej na płaszczyźnie z węzłami np. w punktach o całkowitych współrzędnych zaczynając.
- Zdanie „This results is that the degree (...) distribution is binomial” napisałbym raczej albo jako „This results in binomial degree distribution (...)” albo jako „The result of such approach is that the degree (...) distribution is binomial”.

W końcu zdania po równaniu (2.2) „where $\partial_j i$ represents the neighbourhood of node i in layer j , V the set of all nodes in the graph G , τ_i an activation time of node i , t time and t_{max} the final observation time” konieczne wydaje się albo uporczywe powtarzanie czasowników (np. „represents” albo „is” albo „stands for”) po symbolach V , τ_i , t , t_{max} albo wstawienie po nich myślnika, albo przeformułowanie zdania do postaci „where $\partial_j i$, V , τ_i , t , t_{max} represent the neighbourhood of node i in layer j , the set of all nodes in the graph G , an activation time of node i , time and the final observation time, respectively”.

Udało mi się też znaleźć pomyłki redakcyjne w odwołaniach do numerowanych elementów w pracy, np. na stronie 27 czytamy „this value is responsible for the constant term $1/E$ in Eq. (1.8)” a równanie (1.8) takiego stałego składnika nie zawiera — ma je równanie wcześniejsze: (1.7).

Osobiście odstąpiłbym od pisania kursywą tekstowych indeksów dolnych typu „max” i „crit”.

Zdaje się też, że w ocenie Autora znak dywizu (łącznika) może swobodnie pełnić rolę znaku (pół)pauzy (myślnika) — zwłaszcza do wyodrębniania wtrąceń. Recenzent nie podziela tej oceny.

Przygotowując spis literatury do pracy, prawdopodobnie kierowałbym się większą konsekwencją w uwzględnianiu bądź pomijaniu elementów opisu bibliograficznego (np. przy podawaniu dat publikacji — ograniczając się do wskazania tylko roku, pomijaniu numerów ISSN, pisaniu wszystkich członów nazw czasopism oraz wydawnictw wielką literą a pozostawienie majuskuły tylko przy pierwszym wyrazie tytułu artykułu, itp).

Te (wskazywane z obowiązku recenzenta) uwagi językowo/redakcyjne nie umniejszają jednak wysokiej oceny wartości merytorycznej pracy.

5 Konkluzje

Praca doktorska mgr. inż. Łukasza G. Gajewskiego spełnia formalne i zwyczajowe wymogi dla tego typu prac. W konkluzji jeszcze raz chciałbym podkreślić oparcie całej dysertacji na artykułach opublikowanych w renomowanym czasopiśmie *Physical Review E*. Treść kolejnych rozdziałów odpowiada zawartości tych prac. Szkoda więc, że Autor nie zdecydował się na pracę doktorską w formie zbioru opublikowanych i powiązanych tematycznie artykułów naukowych, której mógł towarzyszyć zdecydowanie krótszy (np. dwudziestostronicowy) przewodnik po nich. Pewnie od przedstawienia rozprawy w powyższej formie powstrzymała Autora jedynie uciążliwa konieczność zbierania oświadczeń od współautorów o ich procentowym udziale w przygotowaniu tych prac.

Reasumując, z pełnym przekonaniem rekomenduję Komisji Doktorskiej dopuszczenie mgr. inż. Łukasza G. Gajewskiego do dalszych czynności w prowadzonym przewodzie doktorskim, w tym do publicznej obrony.

