

Warszawa, dn. 07.02.2022

Prof. dr hab. Ryszard Kutner

Wydział Fizyki Uniwersytet Warszawski

Recenzja rozprawy doktorskiej mgr. Łukasza G. Gajewskiego pt.:
”Reverse engineering multi-layered structures in complex networks”

Uwagi ogólne

Od razu na wstępie pragnę podkreślić, że przedstawiona mi do oceny obszerna i na czasie rozprawa mgr. Łukasza Gajewskiego jest nadzwyczaj ważna dla szybko-rozwijającej się tematyki odnajdywania ukrytych struktur/podsieci/warstw w sieciach złożonych - jest bardzo ciekawa, inspirująca, posiadająca duży potencjał. Mam nadzieję, że moja poniższa recenzja w pełni to wykaże.

Na merytoryczną zawartość rozprawy doktorskiej mgr. Łukasza Gajewskiego składają się trzy bardzo interesujące prace opublikowane w Physical Review E (wiadomość o tej trzeciej wpłynęła przed tygodniem) oraz wybrane elementy trzech innych publikacji wydrukowanych w bardzo dobrych/poczytnych czasopismach międzynarodowych.

Tematyka badania ukrytych struktur topologicznych w układach należących do różnych obszarów nauki/wiedzy mających ogromny, żeby nie powiedzieć zasadniczy wpływ na układ a w tym jego dynamikę, jest zagadnieniem niemal tak starym jak intelektualna aktywność człowieka. Natomiast odnajdywanie ukrytych warstw w sieciach złożonych jest prawdziwym wyzwaniem współczesnej nauki o sieciach złożonych. Tego typu kierunek badawczy jest obecnie bardzo płodny i bardzo modny w teoriach/modelach wielowarstwowych sieci złożonych.

Znajomość wspomnianych powyżej struktur jest konieczna w analizie dynamiki na sieciach, gdzie kluczową rolę odgrywają ścieżki propagacji a w tym ukryte ścieżki propagacji. Ta aktywność – jej natura, może być realizowana np. poprzez rozprzestrzenianie się wirusów (w przypadku epidemii) czy fałszywych/toksycznych informacji (fake news-ów). Ponieważ charakter tej aktywności może być bardzo zróżnicowany więc autor rozprawy zdecydował się na wykorzystanie dwóch (można powiedzieć) kanonicznych podejść: prostego modelu epidemiologicznego typu SI (*Susceptible-Infected*, podatny-zarażony) oraz modelu grafów kwantowych opisującego propagację o charakterze falowym. Te dwa, istotnie różne ale stosunkowo proste koncepcyjnie podejścia, dobrze ilustrują istotę idei zawartych w pracy. Jest to szczególnie ważne w sytuacji, gdy pozyskanie odpowiednich

danych emirycznych jest trudne chociaż nie niemożliwe (jak to ma miejsce w przypadku niniejszej pracy).

Jest dla mnie oczywiste, że modelowanie powyżej wspomnianych układów uzyskuje dodatkowe ogromne możliwości jeżeli oprócz tych dostarczanych przez matematykę (a w tym matematykę stosowaną) oraz IT, wykorzystuje się narzędzia pozyskiwane od nauk fizycznych. Dzięki temu możliwe się staje wiarygodne modelowanie niesłychanie ważnych problemów oraz interpretowanie uzyskanych wyników dotyczących np. takich tematów jak komunikacja społeczna, dynamika opinii, innowacji i rozprzestrzeniania się epidemii, jak też (ogólnie rzecz ujmując) niestacjonarny proces ewolucji społeczeństw. W tym kontekście należy zwrócić uwagę na „rzucające się w oczy” zjawiska polaryzacji społecznych oraz towarzyszące im istnienie/powstawanie baniek społecznych (tzw. komór echa, ang. *echo chambers*).

Ambitnym, zasadniczym celem pracy była propozycja rozwiązania wspomnianych powyżej problemów poprzez zaproponowanie nowatorskich metod identyfikacji/lokalizacji oraz rekonstrukcji ukrytych warstw/struktur w sieciach wielowarstwowych. Metody te bazują głównie na precyzyjnym wyznaczeniu prawdopodobieństw występowania kaskad na dowolnym grafie – tutaj w modelu SI. Podejście od strony łącznego prawdopodobieństwa wystąpienia serii kaskad pozwoliło autorowi zrealizować postawiony cel z dużą dokładnością. Pragnę podkreślić, że właśnie ta duża dokładność jest unikalną wartością pracy gdyż pozwala po prostu uniknąć błędnych rozwiązań. Może warto zaznaczyć w tym miejscu, że zwrot obecny w tytule pracy ‘... Inżynieria Odwrotna ...’ (‘... Reverse Engineering ...’) oznacza, że ukryte struktury odnajdowane są pośrednio, poprzez obserwowanie dynamiki kaskad a nie wprost przez jakiś rodzaj analizy strukturalnej. Należy podkreślić, że analizy były prowadzone przez autora zarówno dla sieci syntetycznych jak i rzeczywistych.

Ponadto, jak wspomniałem, celem odpowiedzi na pytanie: czy istnieją, w jakiej liczbie, gdzie są zlokalizowane i jak wyglądają ukryte struktury sieci wielowarstwowej w danej sieci złożonej, autor eksploatował podejście także od strony grafów kwantowych (ta nazwa nie jest dla mnie jasna – wolałbym nazwę typu analizy falowej, ale rozumiem, że zadomowiła się już niestety w literaturze przedmiotu). Zaproponowana koncepcja jest ciekawa, gdyż opiera się tylko na propagacji fali na pojedynczej warstwie sieci. Oczywiście, na tą propagację mają wpływ sprzężenia pomiędzy nawet odległymi wierzchołkami sieci. Autor zauważył, co widzę jako rozstrzygające, że widmo częstotliwości fal próbujących warstwę ma dodatkowe maksima, których liczba jest zależna od liczby dodatkowych/ukrytych warstw biorących udział w propagacji tych fal. Autor wykazał także,

że dla odpowiednio długich czasów obserwacji kaskad można w pełni odtworzyć widmo znormalizowanej macierzy opisującej cały graf. Podejście to autor słusznie potraktował jako rodzaj metody uczenia maszynowego zmodyfikowanego dla przypadku sieci wielowarstwowych. Uważam podejście od strony grafów kwantowych za wyrafinowane i bardzo skuteczne.

Trzeba zaznaczyć, że ważnym elementem motywującym powstanie pracy był fakt, że problematyka pracy ma bezpośrednie przełożenie na procesy komunikacji zachodzące w społeczeństwach a stąd na dynamikę opinii mającą kluczowy wpływ na postawy społeczne a więc dotykająca nas wszystkich bezpośrednio.

Trzeba wspomnieć także o sprytnie wykorzystanym przez autora pracy dwuwarstwowemu modelowi sieci złożonej dla której względnie łatwo można przećwiczyć dynamikę opinii w obecności ukrytej warstwy. Autor przedyskutował wiele różnych scenariuszy analizując je w ramach wygodnej teorii pola średniego oraz (szkicowo) w ramach użytecznej teorii bifurkacji.

Uwagi szczegółowe

Przedstawiona mi do zaopiniowania rozprawa doktorska składa się z pięciu rozdziałów poprzedzonych obszernym Streszczeniem (w języku polskim i angielskim) oraz zamknięta rozdz. pt.: "Conclusions". Ponadto, praca jest zaopatrzona w Bibliografię, na którą składa się 190 pozycji w tym 9, których współautorem jest autor rozprawy.

W rozdz. 1. Introduction autor jasno wytyczył ściśle ze sobą powiązane cele pracy, podkreślając ich wielkie znaczenie zarówno teoretyczne jak i aplikacyjne, z czym należy się zgodzić. W tym rozdziale autor zrozumiale przedstawił podstawy teoretyczne rozprawy. I tak, kluczowy wzór (1.1) z pomocniczym wzorem (1.2) tworzą teoretyczne podstawy algorytmu do badania kaskad na sieci SI (definiując właśnie 'Reverse Engineering' - inżynierii odwrotnej). W tym kontekście niezręcznie wydaje mi się bardzo długie przedostatnie zdanie na str. 24. Po pierwsze, p. winno tam występować τ_i^c a nie τ_i jak jest teraz. Po drugie, fragment (cytuje): '... given the activation times of its neighbours...' powinien być albo wzięty w nawias albo pomiędzy przecinki bądź przesunięty wyżej do pierwszego zdania tuż pod wzorem (1.2), bo inaczej nie wiadomo do czego odnosi się kluczowy zwrot 'is equal'. Te wzory są tym bardziej ważne, gdyż tworzą podstawę do dalszego rozszerzania na topologie wielowarstwowe oraz model niezależnych kaskad (ang. *Independent Cascade model*).

W rozdz. 1. Introduction autor omówił starannie także model grafów kwantowych, jako model grafów klasycznych z nałożonymi dodatkowymi warunkami, z których najważniejszymi są: 1)

naturalny warunek jakim jest metryczność sieci oraz 2) hamiltonowski charakter sieci, przy czym eksploatowane jest w pracy *de facto* podejście od strony drutów kwantowych. Ponadto, autor ogranicza się do takich grafów, które nie są skierowane, nie są ważone ani nie są grafami dwuczęściowymi (ang. bi-partite graphs). Jednak, należy wziąć pod uwagę, że autor dokonuje w trakcie obliczeń transformacji wyjściowej sieci do pomocniczej sieci skierowanej

W rozprawie wykorzystywana jest ta cecha grafów kwantowych, która pozwala na wprowadzenie równania falowego rozprzestrzeniania się (szeroko-rozumianej) informacji/funkcji – w pracy są to równania (1.2) i (1.3). To wprowadzenie jest jednak w tym miejscu niejasne – staje się jasne dopiero dalej, gdzie autor dopowiada co trzeba. Jest to rodzaj achronologii wyjaśnień. Jednak, lepiej byłoby dla zwykłego czytelnika (ponieważ praca jest już w publicznym obiegu) aby wszystkie wyjaśnienia zostały wprowadzone wcześniej, tam gdzie wprowadzone zostały równania falowe. Mianowicie, chodzi o pytania typu: w jaki (dokładnie rzecz biorąc) sposób

jednowymiarowa zmienna/współrzędna przestrzenna x_e jest przypisywana danej krawędzi grafu? Czy jest to zmienna ciągła a przypisanie jest jednoznaczne? Dalej to się wyjaśnia, że równanie falowe działa na odcinku $[0,1]$, przy czym na każdym takim odcinku niezależnie a jedynie na brzegach łączących krawędzie ma miejsce odpowiednie zszycie czyli wprowadzone są odpowiednio zsynchronizowane (np. fazowo) warunki brzegowe. Jeszcze uwaga dotycząca kolizji oznaczeń: raz zmienna u oznacza węzeł grafu a raz rozwiązanie równania falowego. Ponadto, raz dana krawędź oznaczona jest jako e a raz jako f - czy f należy traktować jako krawędź wchodzącą do warunku początkowego a e to krawędź aktualną? Dalej to jest wyjaśnione ale już tutaj przydałoby się to. Warto byłoby wprowadzić Dodatek, w którym zamieszczone byłoby drobiazgowo wyprowadzenie ogólnego rozwiązania falowego (1.7) – (1.10). Podkreślmy, że rozwiązanie falowe (na amplitudę fali) ma szczególnie prostą postać na węźle. Pouczające, zwłaszcza dla czytelnika spoza środowiska ekonofizyków, może być przykładowe rozwiązanie zamieszczone na końcu rozdz. 1, gdzie wyjściowa macierz sąsiedztwa ma prostą postać. W tym pouczający jest zwłaszcza rysunek 1.3, pokazujący wielkość amplitudy fal na krawędziach diagramu. W związku z tym mam pytanie: jak uzgadniane były warunki brzegowe na wspólnych węzłach sąsiadujących z sobą krawędzi? Bo wydaje się, że obecne są skoki amplitudy przy przechodzeniu od krawędzi do krawędzi; chyba, że to są tylko wielkości amplitud w środku krawędzi?

Dodajmy, że omówione powyżej dwie główne metody są wspomagane przez pomocnicze, takie jak analiza głównych składowych (ang. *Principal Components Analysis*), klasyfikację typu K-NN, analizę fourierowską oraz analizę falkową. Dzięki takiej wielostronności podejścia uzyskane wyniki są bardziej wiarygodne.

Z tego co powiedziałem powyżej wynika, że więcej mówiący tytuł rozdz. 1 powinien być raczej „Theoretical introduction” a nie po prostu „Introduction”.

W rozdz. 2 pt.: „Detecting hidden layers from spreading dynamics” obserwujemy `Reverse Engineering’ w bezpośrednim działaniu na sieciach typu SI. Przypomnijmy, dobrze sformułowane pytanie badawcze jakie stawia autor – jest ono następujące: czy mając możliwość obserwacji kaskady lub kaskad na danej sieci złożonej jesteśmy w stanie odnaleźć ukryte kanały/ścieżki komunikacji. Autor wykorzystuje tutaj wzory (2.1) i (2.2) – repliki wzorów (1.1) i (1.2). W związku z tym autor stawia kluczowe pytanie: jak jest wygodniej, używać więcej kaskad ale przy krótszym czasie obserwacji czy może odwrotnie: mniej kaskad ale przy dłuższym czasie obserwacji? A może występuje tutaj jakieś optimum rozumiane jako maksymalna wiarygodność? Wyniki na rysunku 2.1 (uzyskane dla grafu BA) dają odpowiedź na te pytania a także na pytanie o istnienie ukrytych rozgałęzionych ścieżek. Te wiodące wyniki są wspomagane bardziej szczegółowymi zamieszczonymi na rysunkach 2.2-2.5 wspierającymi wiarygodność metody detekcji ukrytych warstw. W tym kontekście rysunek 2.4 wydaje mi się szczególnie istotny, gdyż pokazuje wprost ułamek rozbieżnych logarytmów prawdopodobieństw (czyli znikających prawdopodobieństw) co pozwala precyzyjnie namierzać ukryte węzły i warstwy.

O ile (w powyżej opisanym) podrozdz. 2.1 autor pokazał jak określić, czy ukryta warstwa istnieje w danej sieci złożonej, czy też nie istnieje, o tyle w kolejnym podrozdz. 2.2 autor przechodzi do tematu inżynierii odwrotnej użytej celem detekcji brakujących/ukrytych połączeń w złożonej sieci SI o zadanej topologii warstwy widocznej oraz o zadanym współczynniku szybkości zarażania dla warstwy ukrytej. Opisuje staranne procedurę/algorytm, opierając się na zaproponowanym przez siebie heurystycznym wzorze (2.9). Brakuje mi tutaj przynajmniej krótkiego wyjaśnienia na jakiej podstawie autor przyjął taką właśnie potęgową postać wiarygodności? Wielce wymowne, drobiazgowo wyniki dla zadanych poziomów ufności 0.5 i 0.95 (α -CSS) autor przedstawił w tabelach 2.1 i 2.2 oraz na rysunkach 2.6-2.10. Podoba mi się proste, zrozumiałe intuicyjnie równanie (2.13) modelu zerowego/referencyjnego łączące liczbę krawędzi (linków) koniecznych do wzięcia pod uwagę aby z zadaną ufnością odczytać liczbę ukrytych krawędzi. Jest to pożyteczny punkt odniesienia. Natomiast nie jest jasne jak autor wprowadził/wyznaczył $\beta_{observed}$? Najważniejsze wyniki tego podrozdziału dotyczą porównań współczynników szybkości zarażania i czułości metody dla różnych poziomów ufności. Są to;

- 1) porównanie wyników dla sieci kwadratowej oraz bezskalowej BA, wykazujące bardzo różne własności; może dla lepszego porównania z siecią kwadratową właściwszym byłoby użycie sieci BA z $m=4$ a nie z $m=3$?

2) Wykorzystanie danych empirycznych - tutaj dotyczących sieci osób zatrudnionych w Department of Computer Science at Aarhus University – patrz rys. 2.11.

Za bardzo ważne uważam także zastosowanie metodologii do innych modeli procesów rozprzestrzeniania się epidemii np. modelu niezależnych kaskad (ang. *Independent Cascade*, IC). Bezpośrednie porównanie zostało przedstawione na rys. 2.15.

Krótko mówiąc w podrozdz. 2.1 oraz 2.2 autor drobiazgowo testował swoją metodologię wskazując na interesujące kierunki kontynuowania badań w najbliższej przyszłości z których badanie sieci ewoluujących wydaje się szczególnie intrygujące i na czasie.

Rozdz. 3 jest w całości poświęcony grafom/sieciom kwantowym. Grafy kwantowe były i są szeroko-stosowane – zwłaszcza w ostatnim czasie, ale nie w kontekście poszukiwania ukrytych węzłów, połączeń i warstw. Ważny jest tutaj fakt, że propagacja informacji na grafach kwantowych ma skończoną prędkość w przeciwieństwie do tej omawianej przez autora wcześniej. Na tej drodze autor potrafił rozróżnić topologie jedno- i wielowarstwowe oraz poprowadzić analogiczne rozważania jak dla sieci SI.

Dokładniej rzecz biorąc, autor wykorzystał gaussowską paczkę falową (ang. *Gaussian Wave Packet Signature*, WPS). Wyniki, uzyskane poprzez techniki uczenia maszynowego (K-NN, PCA), autor przedstawił na rysunkach 3.3 i 3.4 dla sieci BA, projekcji na sieć 2D z PCA oraz sieci ER, porównując ich warianty pięciowarstwowe z jednowarstwowymi. Wiele wymowny jest w tym kontekście rys. 3.5, gdzie nawet sieć pięciowarstwowa jest identyfikowana (w najgorszym przypadku) z dokładnością blisko 75% .

Warto także zwrócić uwagę na podrozdz. 3.3, który zawiera wspomagające, ciekawe podejście od strony transformacji Fouriera. Na tej drodze budowane jest widmo mocy sygnału. Istotą metody jest obserwacja, że widmo mocy jest czułe na liczbę warstw sieci. Świetnie to ilustrują rysunki 3.6 – 3.8 (dla sieci syntetycznych i rzeczywistych). Liczba pików kwadratu amplitudy sygnału oraz widma mocy (w funkcji częstotliwości) jest równa liczbie warstw. Metoda ta wyróżnia się prostotą (nie jest potrzebne tutaj uczenie maszynowe) i bardzo dobrą zdolnością rozdzielczą.

Ważny podrozdz. 3.4 pokazuje (na bazie sieci BA oraz sieci rzeczywistej) , że dysponując wystarczająco długim szeregiem obserwacji można skompletować wszystkie wartości własne macierzy sąsiedztwa a stąd wyznaczyć wszystkie ukryte warstwy (kluczowym jest tutaj wynik przedstawiony na rysunku 3.9 znakomicie rekonstruujący pełne widmo macierzy sąsiedztwa). Z

tego punktu widzenia cel pracy jawi się jako odpowiedź na proste pytanie: jak stworzyć metodę, która przy znacznie krótszym szeregu obserwacji (od tego w powyższej metodzie) dałaby analogiczny wynik?

Piękny wywód analityczny pokazujący związek pomiędzy wartościami własnymi Hamiltonianu (bezpośrednio związanego z macierzą sąsiedztwa) a pikami widma mocy autor zamieścił w podrozdz. 3.4. Podejście to autor przedstawił na przykładach grafów: jedno-klikowego (jednowarstwowego, monoplexu) oraz dwu-klikowego (dwuwarstwowego, duplexu). Wspomniany powyżej związek dobrze ilustruje wynik zamieszczony na rysunku 3.11.

W podsumowaniu tego nadzwyczaj ważnego rozdziału należy podkreślić, że autor poszukiwał ukrytych warstw (ich struktur topologicznych) w ramach użytecznego modelowania pakietów falowych na grafach kwantowych, tzn. wykorzystując dynamikę falową na dwóch drogach: 1) uczenia maszynowego oraz 2) propagowania laplasjanowskich pakietów falowych na krawędziach sieciowych (zwykle o dużym pośrednictwie).

W rozdz. 4 autor odpowiada na pytanie o implikacje wprowadzenia do układu dodatkowej warstwy wykorzystując i rozwijając metody istniejące dotychczas, przy czym punktem wyjścia jest tutaj znane równanie (4.1), opisujące dynamikę opinii poszczególnych agentów w oparciu o ich interakcje. To bogactwo interakcji jest zawarte w macierzy sąsiedztwa. Można tak dobrać tę macierz tak aby odtworzyć np. efekt klastrowania społecznego (baniek społecznych, ang. *echo chambers*). A także odtworzyć możliwość istnienia trzech dobrze znanych stanów trwałych obserwowanych w świecie rzeczywistym. Celem tego rozdziału jest rozszerzenie modelu zdefiniowanego równaniem (4.1) na sytuacje: 1) modelu sieci dwuwarstwowej sprowadzającej się do sieci mono-warstwowej w obecności biasu (tkwi to w addytywnym składniku B w uogólnionym równaniu dynamiki (4.6)) oraz 2) modelu dwuwarstwowego uwzględniającego sprzężenie międzywarstwowe.

Ad. 1) Podejście ma charakter analityczno-numerycznego. Autor wprowadza dwie skale czasowe z których jedna, opisująca zmiany topologii, jest szybka w porównaniu z drugą, opisującą rozprzestrzenianie się opinii. Pozwoliło to autorowi skorzystać z przybliżenia pola średniego formułując efektywne równanie (4.7). Jednak, wydaje mi się, że w równaniu tym zabrakło czynnika α wykorzystywanego przecież przez autora w dalszej części pracy (a w tym zwłaszcza w obliczeniach numerycznych). Wyniki uzyskane przez autora są nadzwyczaj interesujące –

przedstawił je na bardzo bogatym w informacje, pokazującym bardzo ciekawe diagramy fazowe rysunku 4.1.

Ad. 2) Autor dokonał dalszych uogólnień dla modelu dwuwarstwowego, uwzględniając sprzężenie pomiędzy warstwami (patrz równanie (4.11) a stąd przetransformowane równanie (4.12)). Szereg bardzo ciekawych wyników autor przedstawił na rysunkach 4.2-4.4 dokonując porównania rozwiązania numerycznego ze średnio-półowym.

W kolejnych podrozdziałach 4.5 i 4.6 autor omówił: 1) sytuację asymetrycznego sprzężenia pomiędzy warstwami oraz 2) antysymetrycznego sprzężenia pomiędzy warstwami. Rysunki 4.5 oraz 4.6 pokazują wpływ tego efektu na zmianę wyglądu diagramów fazowych. Dynamiczne wykresy na rysunkach 4.7 oraz 4.3 pozwoliły autorowi dokonać ciekawych porównań sytuacji asymetrycznego i symetrycznego sprzężenia międzywarstwowego.

Podsumowanie

Resumując muszę powiedzieć że praca jest świetna – jej merytoryczna zawartość robi nadzwyczaj pozytywne wrażenie jeżeli patrzymy na nią z poziomu zaawansowanego – jest ona bogata w koncepcje i wyniki. Drobne uwagi i niejasności jakie podniosłem w mojej recenzji nie zmieniają mojej bardzo wysokiej oceny pracy.

Uwaga na marginesie, praca wymagałaby rozbudowania o dodatkowe wyjaśnienia, gdy patrzymy na nią z poziomu podręcznika lub z poziomu popularnonaukowego – mam tu na myśli np. czytelnika spoza środowiska zajmującego się sieciami złożonymi. Gorąco zachęcam autora do popularyzacji tej wartościowej, nowatorskiej i inspirującej pracy.

Uważam, że praca z nawiązką spełnia wszystkie wymagania stawiane przez Ustawę przed tego typu rozprawami. Z pełnym przekonaniem rekomenduję rozprawę do dalszych etapów postępowania w procedurze o uzyskanie stopnia doktora nauk fizycznych..

Jednocześnie uważam, że praca zasługuje w pełni na wyróżnienie (podkreślam, że jej elementy ukazały się już w czterech publikacjach w Phys. Rev. E) – zatem, zgłaszam wniosek o jej wyróżnienie. Mam nadzieję, że moja opinia w pełni taki wniosek uzasadnia.

R. Kutner