

Wrocław, 31 maja 2019

Prof. dr hab. Katarzyna Sznajd-Weron
Katedra Fizyki Teoretycznej
Wydział Podstawowych Problemów Techniki
Politechnika Wrocławska

Recenzja Rozprawy Doktorskiej mgr. Mateusza Wilińskiego

„Przemiany fazowe w empirycznych korelacyjnych sieciach złożonych”

Wstęp

Obszar badań układów złożonych jest w swej naturze wysoce interdyscyplinarny i leży na styku wielu, często pozornie bardzo odległych dziedzin, począwszy od nauk ścisłych przez przyrodnicze, aż po nauki społeczne i humanistyczne. Przedstawiona do oceny rozprawa wpisuje się w ten nurt, dając wkład zarówno do nauki o sieciach złożonych (network science), jak i analizy sygnałów, ekonofizyki oraz teorii przemian fazowych. Pisząc, że „daje wkład” mam na myśli fakt, że autor rozprawy zaproponował nowe metody analizy układów złożonych, w tym nowy estymator korelacji oraz własną metodę filtracji danych, którą oczywiście porównał z kilkoma standardowymi podejściami. Ponadto zaproponował również nowy, interesujący model ewoluującej sieci złożonej. Wyniki zostały opublikowane w cyklu czterech współautorskich prac w uznanych czasopismach z listy filadelfijskiej, takich jak: Physica A, The European Physical Journal B, Journal of Statistical Mechanics oraz Physical Review E. W trzech z tych prac mgr. Wiliński jest pierwszym autorem.

Opis i ocena rozprawy

Rozprawa składa się z pięciu rozdziałów i właściwie wszystkie, poza rozdziałem pierwszym, który jest tradycyjnie wprowadzeniem do pracy, zawierają opisy autorskich metod i wyników, uzyskanych przez doktoranta dzięki wprowadzeniu tych nowych metod. We wprowadzeniu, autor opisuje cel pracy, który dotyczy: (1) analizy układów o połączeniach niejawnym, co moim zdaniem jest wyjątkowo interesującym zagadnieniem, (2) modeli ewoluujących sieci złożonych oraz (3) zastosowania zaproponowanych metod do danych empirycznych. Realizacja celu pierwszego wymagała od autora nie tylko wprowadzenia nowej metody pomiaru podobieństwa procesów, które są próbkowane w nieregularnych odstępach czasowych, ale również nowej metody filtracji danych. Trochę zabrakło mi w podrozdziale dotyczącym celu pracy sformułowania jednego nadrzędnego celu, co z pewnością mogłoby być zrobione. Jest to jednak niewielki zarzut, ponieważ z przedstawionego we wprowadzeniu opisu, wynika

jasno, że rozprawa nie składa się po prostu z trzech niezależnych zadań, ale że są one ze sobą ściśle powiązane. **Niemniej jednak jestem ciekawa, jak sformułowałby autor nadrzędny cel pracy.** Kolejne rozdziały rozprawy odpowiadają poszczególnym celom wymienionym we wprowadzeniu.

W rozdziale drugim autor opisuje estymatory korelacji szeregów czasowych oraz metody filtracji grafów ważonych. Jeśli chodzi o wyznaczanie korelacji, to autora interesują szczególnie takie przypadki, w których dane próbkowane były w nieregularnych odstępach czasu, czego konsekwencją może być brak synchronizacji badanych sygnałów. Autor zaczyna od opisu powszechnie używanego estymatora współczynnika korelacji liniowej Pearsona. Następnie przechodzi do metody liczenia estymatora kowariancji na podstawie transformaty Fouriera szeregów czasowych. Zakładając, że badane szeregi czasowe można przedstawić w postaci funkcji schodkowych, wyprowadza wzór na kowariancję ich skoków. Ten rachunek jest wprowadzeniem do opisu kolejnego estymatora, który może być wykorzystany w sytuacji, gdy analizujemy zależności między szeregami czasowymi, które nie tylko są próbkowane nieregularnie, ale również ze sobą niezsynchronizowane (tzn. przesunięte względem siebie w czasie). W końcu wyprowadza analityczne postaci przesuniętych kowariancji dla trzech typów procesów: (1) będących sumą Delt Diraca, (2) mających postać schodkową (tzn. będących sumą funkcji Heaviside'a) oraz (3) funkcji ciągłych, odcinkami liniowych. Na szczególne podkreślenie zasługuje fakt, że wszystkie wyprowadzenia, przedstawione w tej części rozprawy, są autorskie.

W dalszej części rozdziału autor opisuje zastosowanie przedstawionych wcześniej metod do sygnałów syntetycznych, co jest dobrą praktyką ponieważ pozwala na przetestowanie metody. Samo generowanie sygnałów syntetycznych jest nietrywialnym problemem, ponieważ generowane sztuczne dane muszą wykazywać te same problemy, które potem będą obecne w danych empirycznych. Autor proponuje tu własny, wydajny algorytm, którego opis znajduje się w dodatku A. W rozdziale tym autor wskazuje na problem wyboru odpowiedniej skali czasu, pokazuje też, że faktycznie metoda bazująca na transformacjach Fourierowskich radzi sobie lepiej z nieregularnościami procesu, niż klasyczna metoda estymacji korelacji Pearsona.

W kolejnych podrozdziałach autor opisuje metody filtracji grafów ważonych, w tym klasyczną metodę minimalnego drzewa rozpinającego (MST od ang. *minimum spanning tree*). Ponieważ MST eliminuje z grafu pętle więc zmienia również wiele niezwykle istotnych miar sieciowych, w szczególności współczynnik gronowania. Dlatego autor przedstawia kolejną metodę filtracji, nazwaną filtrującym maksymalnym grafem planarnym (PMFG od ang. *planary maximally filtered graph*), zaproponowaną w roku 2005 przez Tumminello i innych, która jest uogólnieniem MST, polegającym na zamianie warunku braku pętli, na warunek zakazujący tworzenia grafu nieplanarnego. Jako główną wadę metody PMFG autor podaje wysoki stopień komplikacji metody. Tak na marginesie, to jestem ciekawa dlaczego autor używa terminu „metodologia” zamiast

po prostu „metoda”? W celu wyemitowania wad wcześniejszych metod filtracji, autor proponuje własną metodę, której kluczowym elementem jest rozkład stopni wierzchołków. Tu nasuwa mi się kolejna uwaga językowa. Autor napisał „Najpowszechniej analizowaną topologiczną cechą sieci jest rozkład jej stopni”, co jest dość niefortunnym sformułowaniem, ponieważ nie chodzi o rozkład stopni sieci, a rozkład stopni wierzchołków, czyli spodziewałam bym się tu raczej zdania: „Najpowszechniej analizowaną topologiczną cechą sieci jest rozkład stopni jej wierzchołków”.

Na kolejnych stronach rozdziału drugiego, autor opisuje zastosowanie wszystkich trzech przedstawionych metod, w tym autorskiej o roboczej nazwie DEG oraz dodatkowo prostej (można powiedzieć referencyjnej) metody progowania polegającej na usunięciu wszystkich połączeń o wartości mniejszej lub większej niż określony próg, do analizy danych syntetycznych. Tworzona jest macierz zależności o elementach d_{kl} , które, jeśli dobrze rozumiem, opisują zależność między wierzchołkami k i l i które są zadane formułą (2.4). Tak przynajmniej wynika z opisu na stronie 32. Jeśli faktycznie d_{kl} opisuje zależność pomiędzy wierzchołkami k i l to czy wówczas nie jest to po prostu waga, a macierz zależności to po prostu macierz sąsiedztwa? Czy autor celowo nie używa tu terminu „wierzchołek” tylko „element”, ponieważ nie chce jeszcze w tym miejscu opisywać danych w sposób sieciowy? Byłabym wdzięczna za wyjaśnienie. W definicji d_{kl} kluczową rolę odgrywa parametr λ , który określa wkład rzeczywistej siły zależności w stosunku do szumu. **Jak pisze autor, pierwszym celem filtrowania było pozostawienie $N - 1$ najważniejszych zależności. Co to znaczy najważniejszych? Czy nie oznacza to po prostu, że celem jest pozostawienie $N - 1$ krawędzi o największych wagach? Domyślam się, że nie chodzi tylko o to, ponieważ w przeciwnym wypadku idealna byłaby prosta metoda progowania.**

Autor pokazuje, że w przypadku gdy wartość parametru λ jest duża, czyli szum mały, wówczas wszystkie metody dają podobne rezultaty, jeśli mierzyć liczbę połączeń wewnątrz grup. Dla małych wartości λ wszystkie metody działają mniej skutecznie, ale ilościowo najlepiej wypada prosta metoda progowania, a zaraz po niej DEG. Dalsze analizy pokazują, że prosta metoda progowania nie nadaje się jednak do filtracji, ponieważ grafy uzyskane poprzez progowanie zawierały jedynie połączenia wewnątrz najsilniej połączonej grupy. Pozostałe metody są wolne od tej wady, jednak MST nie pozwala na uzyskiwanie grafów o rozmiarach większych niż $N - 1$, co w wielu zastosowaniach może być wadą. Natomiast metoda PMFG, o czym już wspomniano, jest znacznie trudniejsza numerycznie od DEG. Wydawałoby się więc, że metoda DEG jest idealna i powinna zastąpić wcześniejsze metody filtracji i w związku z tym mam pytanie. **Czy potrafi autor podać jakąś wielkość, która pokazywałaby różnice pomiędzy PMFG oraz DEG i w przypadku jakich zastosowań miałyby to znaczenie?**

Kolejny rozdział dotyczy przemian fazowych na sieciach złożonych, czyli tematyce najbliższej moim zainteresowaniom naukowym. Przedstawione tu zostały metody, które (zgodnie ze słowami autora) mają „ułatwić wykrywanie strukturalnych przemian

fazowych na podstawie danych empirycznych opisujących dynamikę układu". We wstępie do tego rozdziału (czyli w podrozdziale 3.2) autor przybliży czytelnikowi podstawowe pojęcia z teorii przemian fazowych, takie jak parametr porządku, wykładniki krytyczne i klasy uniwersalności. Następnie przechodzi do opisu modelu Isinga i jego rozszerzenia czyli modelu Potts'a. **Zabrakło mi tu uzupełnienia informacji o wymiarze układu w miejscu, w którym autor pisze „Co ciekawe, dla $q \geq 3$ w modelu Potts'a obserwuje się przemianę fazową pierwszego rodzaju.” Jest to bowiem prawda dla układu trzywymiarowego, ale już dla dwuwymiarowego przemianę nieciągłą obserwuje się dopiero dla $q > 4$.** Nie zmienia to jednak faktu, że fragment poświęcony wstępowi do teorii przemian fazowych jest opisany dobrze, podane zostały w sposób zwięzły i klarowny wszystkie informacje, które są potrzebne do zrozumienia dalszej części rozprawy.

Kolejne podrozdziały poświęcone są modnemu w ostatnich latach zagadnieniu, jakim są topologiczne przemiany fazowe, obserwowane na przykład w dwuwymiarowym modelu XY. Po wstępie, dotyczącym właśnie tego modelu, autor przechodzi do opisu układów, w których przemianie fazowej podlega topologia grafu. **Tu zabrakło mi trochę wyjaśnienia, dlaczego w przypadku np. grafu Wattsa-Strogatza i Barabasięgo-Alberta możemy mówić o różnych topologiach, a nie po prostu strukturach.** Zacytowałabym tu klasyczną pozycję Whitneya z 1933 „A Set of Topological Invariants For Graphs” i wyjaśniła jak definiuje się fakt, że dla grafy są topologicznie równoważne i jaka wielkość może być niezmiennikiem topologicznym. Wydaje mi się to o tyle istotne, że klasycznie w topologii proces deformacji powinien być ciągły, zabronione jest rozrywanie i zlepianie różnych części, a grafy to struktury dyskretne. Autor napisał wprawdzie, że „Sieci są natomiast obiektami dyskretnymi co paradoksalnie powoduje, że bardzo rzadko mamy do czynienia z dwiema, które są topologicznie równoważne w sensie ścisłym”. Wydaje mi się, że to jednak jest zdecydowanie zbyt lakoniczne stwierdzenie, biorąc pod uwagę jakie emocje wywołuje w ostatnim okresie słowo „topologiczne”.

W dalszej części autor opisuje przykłady przemian fazowych związanych ze strukturą grafu oraz ich związki z przemianami fazowymi, znanymi z tradycyjnej fizyki statystycznej, takimi próg perkolacji czy kondensacja Bosego-Einsteina. Nazwa tego podrozdziału to „Modele sieciowe z topologicznymi przemianami fazowymi.” i stąd moje pytanie. **W jakim sensie są to przemiany topologiczne? Czy można wobec tego powiedzieć, że zwykła przemiana zachodząca w modelu perkolacji jest też przemianą topologiczną? Jeśli nie, to czym różni się ona wobec tego od przemiany w grafie losowym, związanej z rozmiarem największego komponentu?** Tu jeszcze mała uwaga to opisu na stronie 48. Autor odwołuje się do Rysunku 2.1, ale powinien odwołać się do Rysunku 3.8, na którym widnieje diagram fazowy dla grafu losowego.

Na kolejnych stronach autor przechodzi do opisu idei sieci koewoluującej czyli takiej, w której dynamiki wierzchołków i połączeń są ze sobą w pewien sposób powiązane. Podane tu zostają przykłady sieci bazujących na modelu kulturowym Axelroda oraz na modelu Isinga. W podrozdziale 3.5 doktorant wykorzystuje formalizm hamiltonowski do

opisu do koewoluującego modelu Isinga, co jest jego autorskim pomysłem polegającym na tym, że Hamiltonian zawiera w sposób jawny zarówno stany wierzchołków jak i ich stopnie. Pomysł jest prosty, ale nie tylko skutkuje niezwykle interesującymi wynikami, ale również ma bardzo ciekawą interpretację z punktu widzenia nauk społecznych. **Pytanie tylko czy układy społeczne są faktycznie równowagowe? Chętnie poznałabym tu zdanie autora. Jak mógłby obronić zarzut stosowania metod równowagowej fizyki statystycznej do układów społecznych?** Na kolejnych stronach autor analizuje model w ramach symulacji Monte Carlo oraz analitycznie w sposób przybliżony dla wybranych wartości parametrów.

W ostatnim podrozdziale rozdziału trzeciego autor zdecydował się opisać samoorganizującą się krytyczność (SOC od ang. *self-organized criticality*) i ta decyzja nieco mnie zdziwiła, ponieważ ta część sprawia wrażenie oderwanej od reszty rozdziału, co zaburza zdecydowanie czytanie rozprawy. Tym bardziej, że autor opisuje tu klasyczny model pryzm piasku oraz płonącego lasu bez nawiązania do sieci złożonych. Taka urwana narracja pojawiła się już wcześniej w rozdziale drugim, kiedy to autor przechodził od opisu metod wyznaczania korelacji w szeregach nieregularnych do filtrowania grafów, ale w rozdziale trzecim zaburzenie narracji jest w moim odczuciu bardziej rażące.

Ostatni rozdział poświęcony jest danym empirycznym. W tym rozdziale autor wykorzystał wcześniej zaproponowane metody. Najpierw rozważa zmiany strukturalne na rynkach finansowych, a następnie także zmiany w ludzkim mózgu. Szczególnie interesujące wydaje mi się to drugie zastosowanie. Autor analizuje strukturę korelacji w mózgu na podstawie sygnału pochodzącego z badania EEG. W szczególności, bada on wpływ padaczki na strukturę sieciową sygnału EEG, wykorzystując dane pochodzące z Europejskiej Bazy Danych Epilepsji. Na podstawie przeprowadzonej analizy autor stwierdza dwa dominujące zachowania w przypadku ataków typu złożonego (18 obserwacji): (1) znaczny wzrost maksymalnego stopnia lub (2) nagły spadek maksymalnego stopnia. Wiem, że autor jest fizykiem, a nie neurologiem, ale czy udałoby się jakoś skomentować ten wynik? Oczywiście liczba przypadków przebadanych przez autora jest zbyt mała, żeby móc ocenić zaproponowaną przez niego metodę, ale jest to niewątpliwie interesujący kierunek badań.

Podsumowanie

Rozprawa jest bardzo obszerna i mam tu na myśli nie tylko liczbę stron, ale przede wszystkim liczbę uzyskanych przez autora wyników. W dwóch miejscach nie do końca zrozumiała jest dla mnie struktura logiczna rozprawy, ale generalnie rozprawa przygotowana jest bardzo starannie, zarówno pod względem językowym, jak i edytorskim. Autor jasno formułuje swoje cele, wyjaśnia przyczyny wprowadzania nowych metod i podkreśla wyniki, które uznaje za istotne. Wszystko to wskazuje na dojrzałość autora. Wyniki uzyskane przez autora są cenne z punktu widzenia nauki o układach złożonych. Wydaje się, że użyteczna może być zarówno zaproponowana przez niego, stosunkowo prosta metoda filtracji grafów, jak i metoda estymacji korelacji, która

nadaje się do analizy większości danych empirycznych, w tym tych o nieregularnym kroku czasowym. Ponadto, zaproponowany przez niego hamiltonowski model Isinga może być inspirujący dla innych badaczy, czego szczerze autorowi życzę.

Podsumowując, uważam że rozprawa doktorska mgr. Mateusza Wilińskiego spełnia wszelkie wymogi stawiane przez Ustawę tego typu rozprawom. Wnoszę zatem o dopuszczenie jej do kolejnych etapów przewodu doktorskiego. Jednocześnie, biorąc pod uwagę wkład jaki wnosi rozprawa w rozwój fizyki układów złożonych, wnioskuję o jej wyróżnienie.

Krzysztof Weron