

20 sierpnia 2021

Dr hab. Paweł Sobkowicz
Centrum Doskonałości NOMATEN
Narodowe Centrum Badań Jądrowych
Ul A. Sołtana 7, 04-500 Otwock Świerk

Recenzja

rozprawy doktorskiej mgr inż. Joanny Linczuk, pt. "Koewolucja dynamiki stanów i połączeń na sieciach złożonych", przedstawionej Radzie Naukowej Dyscypliny Nauki Fizyczne Politechniki Warszawskiej.

1. Wstęp

Recenzowana praca dotyczy dynamiki sieci koewoluujących w teoretycznym modelu głosującego (ang. *voter model*) oraz w zastosowaniu do opisu danych dotyczących dwóch sieci komunikacji społecznej o bardzo zróżnicowanym charakterze: sieci kontaktów w społeczności studentów uczelni wyższej oraz statystycznych własności komentarzy na platformie Twitter.

Rozprawa ma 85 stron i podzielona jest na cztery rozdziały, z czego wyniki uzyskane przez autorkę przedstawione są w dwóch rozdziałach (drugim i trzecim) dedykowanych odpowiednio modelowi teoretycznemu i analizie danych empirycznych. Bibliografia rozprawy liczy 94 pozycje z czego cztery to publikacje autorki (dwie opublikowane, jedna wysłana do publikacji i jedna w przygotowaniu). Obie opublikowane wyniki ukazały się w wysoko ocenianym czasopiśmie *Physical Review E*. W jednej z nich autorka jest pierwszym a w drugiej drugim autorem.

Celem rozprawy było zrozumienie mechanizmów koewolucji w modelach sieci społecznych oraz identyfikacja procesów wpływających na koewolucję takich sieci w oparciu o dane rzeczywiste.

2. Najważniejsze elementy i osiągnięcia rozprawy

- **Pierwszy rozdział** rozprawy zawiera wprowadza czytelnika w zagadnienia i metody stosowane w dalszej części pracy. Wprowadzenie to obejmuje szeroki zakres tematów: informacje o sieciach, w tym w szczególności definicje podstawowych pojęć używanych w analizie sieci; wprowadzenie do modeli agentowych i celów i sposobów ich stosowanie, ze szczególnym uwzględnieniem modelu głosującego, stosowanego w analizie związków społecznych; zdefiniowanie procesu koewolucji sieci, łączącego zmiany stanów węzłów/połączeń sieci ze zmianami topologii połączeń; wybrane sposoby uproszczeń pozwalających na porównanie wyników modeli agentowych z obliczeniami analitycznymi (np. metoda pola średniego czy przybliżenia par); oraz wprowadzenie do teorii równowagi Heidera, będącej psychologiczną podstawą modeli teoretycznych i agentowych. Rozdział, mimo koniecznej skrótowości (omawiane zagadnienia są bardzo rozbudowane), jest napisany klarownie i wprowadza czytelnika w podstawowe zagadnienia. Autorce nie udało się jednak uniknąć stwierdzeń niespójnych. Omawiając zjawiska koewolucji, autorka podkreśla znaczenie podobieństwa skali czasowej procesów zmiany stanów sieci i zmian jej topologii, jednak jako przykład podaje powstawanie nowych dróg jako rezultatu występowanie korków – procesy o bardzo różnych skalach czasowych.
- **Rozdział drugi** rozprawy dotyczy teoretycznego opisu koewolucji sieci w modelu głosującego i podzielony jest na trzy części: wprowadzenie do modelu, liniowego modelu głosującego i nieliniowego modelu głosującego ze zmienionym mechanizmem przełączania powiązań między węzłami sieci. Podrozdziały zredagowane zostały w spójny sposób i obejmują definicje modeli, opis analityczny, symulacje numeryczne oraz dyskusje wyników uzyskanych w ramach obu modeli. W modelu liniowym wykazane zostało powiązanie między magnetyzacją węzłów sieci i połączeń w przybliżeniu średniego pola. Model

analizy przedstawiony jest bardzo szczegółowo i przekonująco (w ramach przyjętych przybliżeń), jednak pewien niedosyt pozostawia brak dyskusji roli skalowania czasu symulacji dla sieci o różnych ilościach węzłów (przykładowo w kontekście skalowania przedstawionego na rysunku 2.3). Autorka wykazała istnienie statystycznej stałej ruchu, będącej liniową kombinacją magnetyzacji węzłów i magnetyzacji krawędzi. Podrozdział poświęcony nieliniowemu modelowi głosującego i koewolucji sieci w tym modelu jest bardzo interesujący, w szczególności w zakresie analizy stabilności sieci i przejść fazowych między stanami symetrycznymi i antysymetrycznymi, a także dyskusji różnic między wynikami analitycznymi a wynikami z symulacji Monte Carlo. W tym ostatnim przypadku można zwrócić uwagę na brak dyskusji dotyczącej zależności efektów fluktuacji od liczby węzłów sieci (symulacje przeprowadzone zostały dla 10000 węzłów).

- **Rozdział trzeci** poświęcony jest analizie zjawisk koewolucji dla autentycznych sieci społecznych – w tym próbie odkrycia i wyjaśnienia mechanizmów powodujących zmiany topologii sieci. Jako przedmiot analiz wybrane zostały dwie bardzo różne sieci społeczne. Pierwszy przeanalizowany zbiór danych opisuje osobiste kontakty grupy studentów uniwersytetu Notre Dame w Stanach Zjednoczonych, badane w ramach eksperymentu NetSense, przeprowadzonego w latach 2011-2013. Dane te mają unikatowy charakter, gdyż opisują zarówno sieć kontaktów jak też długoterminowe, obejmujące 6 semestrów, zmiany opinii poszczególnych studentów na osiem istotnych tematów. Mimo stosunkowo małej próbki (108 studentów), ten długoterminowy charakter badania eksperymentalnego czyni jego wyniki bardzo interesującymi z punktu widzenia analizy dynamiki sieci społecznej. W ramach przeprowadzonej analizy, autorka skoncentrowała się na wpływie zrównoważenia bądź jego braku w oddziaływaniach między trójkami studentów (triad), posługując się pojęciami teorii równowagi Haidera. Autorka zaproponowała cztery modele zachowań agentów tworzące macierz oparta na dwóch charakterystykach: założenia braku oddziaływań lub występowania tendencji to połączeń między agentami o podobnych poglądach oraz rozważania opinii na jedynie jeden temat lub na zważonej kombinacji poglądów na wszystkie osiem tematów. W analizie statycznego zachowania systemu (własności

uśrednionych po całym okresie eksperymentu) autorka stwierdziła brak dowodów na występowanie oddziaływań Heiderowskich dla modeli opartych o pojedynczy temat (modele A1, E1), podczas gdy oddziaływania takie mogą być obecne (dla pewnych wartości parametrów modelu) w modelach uwzględniających kombinacje tematów (modele A8, E8). Jest to ciekawa obserwacja, jednak należy zwrócić uwagę, że model nie uwzględnia korelacji między opiniami danego studenta na różne tematy (prawdopodobnie to samo założenie występuje też w modelu E8, nie jest to jednak jasno stwierdzone przez autorkę). Takie założenie braku systematycznych korelacji jest dość ryzykowne. W eksperymencie NetSense pytano o opinie na temat: aborcji, kary śmierci, eutanazji, homoseksualizmu, małżeństw jedнопłciowych, użycia marihuany, poglądów politycznych i seksu przedmałżeńskiego. W znacznym zakresie opinie te w społeczeństwie amerykańskim są mocno skorelowane ze sobą, tworząc systemy wartości (liberalny i konserwatywny) czy definiując poglądy polityczne (Demokraci/Republikanie). Jeśli przyjąć założenie silnych korelacji, wyniki uzyskane dla modeli jednotematycznych mogą lepiej odpowiadać rzeczywistości, pomniejszając znaczenie wniosku o występowaniu oddziaływań Heiderowskich. W części podrozdziału poświęconej dynamice sieci (koewolucji poglądów i kontaktów) autorka porównała prawdopodobieństwa zmiany stanu triad (w modelach wielotematycznych). Porównując wyniki z danych empirycznych i modeli, autorka wnioskuje o mieszanych (Heiderowskim i przypadkowym) charakterze dynamiki. Jednak brak skali błędów na wykresie 3.6 zdecydowanie utrudnia określenie czy wnioski te są statystycznie uzasadnione. Dodatkową słabością podrozdziału poświęconego występowaniu (lub nie) równowagi Heidera w środowisku studentów jest brak dyskusji alternatywnych wyjaśnień obserwowanych cech statystycznych. Autorka, w dyskusji wyników, przywołuje możliwe niedoskonałości wyznaczenia parametrów modelu, to jednak nie zastępuje, w opinii recenzenta, dyskusji alternatywnych do mechanizmu Heiderowskiego procesów determinujących zmiany opinii i sieci kontaktów.

- Drugi podrozdział części analizy danych empirycznych poświęcony jest danym zebranym z Twittera w dość krótkim okresie października-listopada 2011. W kontekście różnicy między skalami czasowymi eksperymentu NetSense i danych

Twittera zwrócić należy uwagę na znacząco szybszą dynamikę zachowań w sieciach społecznych, zatem nawet tak krótki okres objęty analizowanymi danymi może być wystarczający do zaobserwowania własności dynamicznych. Autorka starała się odkryć mechanizmy rządzące obserwowaniem lub zaniechaniem obserwowania pomiędzy użytkownikami Twittera. Podstawowym kryterium podobieństwa między użytkownikami była odległość między ich zainteresowaniami (określonymi na podstawie słów używanych w tweetach, analizowanych metoda TF-IDF). W tym miejscu zwrócić należy uwagę na fakt, że dyskusje obejmujące ten sam temat (np. konkretne wydarzenie lub osobę) mogą być prowadzone także między osobami o przeciwstawnych poglądach, zatem podobieństwa wektorów używanych słów nie koniecznie prowadzą do wniosku o podobieństwie poglądów. Wpływ kontrowersji i kłótni na interakcje w przestrzeni sieci społecznych i komunikacji elektronicznej był wcześniej analizowany, jednak rozprawa nie zawiera odnośników do tych wyników, ani dyskusji przyjętych założeń. Uwagę zwraca stwierdzenie na stronie 69: *„Użytkownicy zostają otoczeni przez innych takich samych jak oni i widzą coraz więcej postów, które mogą ich zainteresować. Ponieważ interesujące tweety będą częściej przesyłane dalej, ta zbieżność zainteresowań między sąsiadami prowadzi do łatwiejszego rozpowszechniania informacji i tworzenia ścieżek, po których informacje mogą się łatwo rozprzestrzeniać.”* Takie wyjaśnienie powstawania grup o gęstej sieci powiązań (luźno powiązanych z pozostałymi częściami sieci) wymaga czegoś więcej niż prosta analiza podobieństw wektorów słów uzyskanych z metody TF-IDF. Analiza statystycznych własności danych wskazała jednak na istnienie bardzo interesujących regularności w prawdopodobieństwach warunkowych retweetów, zerwania obserwacji i kontynuowania (rys 3.10) są bardzo interesujące. Uzyskane w modelu wyniki dotyczące dynamiki odległości między agentami (rys 3.11) są jakościowo zbliżone do obserwacji empirycznych (rys 3.7). Drobnym brakiem w tym porównaniu jest użycie różnych skal czasowych (przeliczenie ze skali dni na liczbę tweetów na rys 3.7 mogłoby być dokonane przez określenie średniej liczby tweetów w ciągu dnia). W dalszej części podrozdziału autorka analizowała wpływ parametrów modelu na jakość odtworzenia zachowań rzeczywistych i otwarcie wskazała na słabości modelu – ta otwartość zasługuje na

uznanie, gdyż wskazuje na potencjalne dalsze kierunki rozwoju badań. Podrozdział zawiera także krótką analizę zachowań na sieciach syntetycznych (Erdős-Rényiego (ER) i Barabási-Albert (BA)) i wskazanie uniwersalności zjawiska zmian odległości między użytkownikami w czasie, podobne do zaobserwowanych w danych rzeczywistych.

- Rozdział czwarty zawiera klarowne podsumowanie wyników, niestety w podsumowaniu tym wskazano jedynie dokonania i pozytywne rezultaty, pomijając dyskusję ograniczeń przyjętych założeń i uzyskanych wyników.

3. Najciekawsze osiągnięcia rozprawy

Rozprawa oparta jest o prace badawcze łączące bardzo zróżnicowane podejścia. Rozdział drugi, skoncentrowany na analizie teoretycznej modelu głosującego jest bogaty zarówno w wyniki analityczne jak też symulacje komputerowe. Zidentyfikowanie w modelu nieliniowym dwóch dynamicznie aktywnych faz: dobrze znanej symetrycznej fazy oraz niesymetrycznej fazy, która pojawia się w efekcie spontanicznego złamania symetrii, oraz opis przejścia fazowego między tymi stanami uważam za jedno z najważniejszych osiągnięć pracy.

Stanowiące drugą część pracy modele opisujące dane empiryczne wymagały zupełnie innego podejścia i – mimo wskazanych powyżej słabości – zasługują na uznanie, w szczególności zważywszy na fakt, że modele agentowe ze swojej natury mogą opisać zaledwie fragment skomplikowanych zjawisk psychologicznych i społecznych. Nowe i interesujące analizy danych empirycznych są niezbędne dla modyfikacji i rozwoju modeli agentowych.

4. Wnioski końcowe

W mojej opinii rozprawa mgr inż. Joanny Linczuk zawiera zróżnicowane i interesujące opracowania w dziedzinie układów złożonych. Wśród oryginalnych i wartościowych rezultatów, należy podkreślić wyniki dotyczące dynamiki przejść fazowych w modelu głosującego oraz analizę zmian czasowych podobieństw między użytkownikami Twittera. Część wyników uzyskanych przez autora rozprawy została przedstawiona w

dwóch publikacjach, dla pozostałych publikacje są złożone do recenzji lub w przygotowaniu.

Praca w zupełności spełnia warunki stawianym rozprawom doktorskim. Stawiam wniosek o dopuszczenie pracy do publicznej obrony.



Dr hab. Paweł Sobkowicz