

Recenzja rozprawy doktorskiej **mgra Andrzeja Jarynowskiego** pod tytułem:
*“Exploring the dynamics and the complexity of human behavior using
nonlinear physics methods”*.

Przedstawiana poniżej ocena pracy mgra Andrzeja Jarynowskiego ubiegającego się o nadanie stopnia doktora nauk fizycznych na Wydziale Fizyki, Astronomii i Informatyki Stosowanej, Uniwersytetu Jagiellońskiego w Krakowie, jest opinią dotyczącą rozprawy powtórnie przez niego przygotowanej. Promotorem rozprawy jest dr hab. Andrzej Grabowski, prof. CIOP-PIB. W swojej poprzedniej recenzji wnosiłam o ponowne przygotowanie rozprawy, tym razem spisanej w formie samodzielnej rozprawy, ponieważ uważałam dorobek publikacyjny mgra Jarynowskiego, choć doniosły liczbowo, to jednak jakościowo za słaby, by ubiegać się na jego podstawie o stopień doktora. Przedstawiona obecnie do oceny praca nadal jest zbiorem artykułów Autora.

Mgr Jarynowski przedstawia do oceny cykl siedmiu prac, zebranych w dwa tematy: *Modelowanie epidemiologii* oraz *Sieci i modelowanie danych*. W porównaniu do poprzednio prezentowanych artykułów, Autor pominął szereg prac, w tym te, które nie zostały jeszcze opublikowane (był to wniosek obu Recenzentów), natomiast włączył inne swoje prace, uprzednio niezgłoszone do oceny. Poszerzył także omówienie prezentowanych wyników. W skrócie mogę stwierdzić, że obecnie przedstawione do oceny artykuły zdecydowanie lepiej niż poprzednie prezentują warsztat naukowy Doktoranta, co było głównym niedostatkami poprzednio rozpatrywanej rozprawy doktorskiej. To szersze omówienie problematyki i wszystkich wyników Doktoranta (strony 7-75 załączonej do recenzji dokumentacji) uważać będę za rozprawę.

Mam do tej rozprawy wiele zastrzeżeń przede wszystkim formalnych, ale też i merytorycznych. Główne zastrzeżenia są związane z niewłaściwym, a w wielu wypadkach błędnym operowaniem informacją naukową. W tekście naukowym używa się stwierdzeń ogólnych – zwłaszcza w części wprowadzającej, ale takie ogólne stwierdzenia muszą być poparte cytowaniami. Brak odnośnika literaturowego czyni je bowiem ogólnikowymi.

Doktorant na swojej stronie udostępnia wszelkie swoje opracowania: o mniej lub bardziej naukowym charakterze. Zapewne tak też stanie się i z recenzowaną rozprawą. Dlatego też moja ocena musi być wyjątkowo szczegółowa.

Na stronach od 7 do 28 rozprawy, Pan Jarynowski przedstawia **ogólne wprowadzenie do tematyki**, którą rozwija w dalszej części a także prezentuje stosowane przez siebie metody. Przedstawione tu treści są ujęte hasłowo. Nie pozwalają czytelnikowi ani na poznanie definicji pojęć kluczowych dla rozważań, ani nauczenia się stosowanych przez Autora metod. Dużą

część stanowi rozbudowany opis zagadnień, które, na różnych etapach swojego rozwoju naukowego, Autor rozważał.

Uwagi szczegółowe:

- Strona 7: Za podstawowe dla teorii sieciowych i badań własnych uznał Pan prace Lazer, et al. 2009 i Barabasi, et al. 1999. Obie prace są źle opisane w bibliografii. Proszę o wskazanie poprawnych opisów.
Dlaczego powołuje się Pan na pracę Green, et al. 2009, jako na pracę wprowadzającą/wyjaśniającą zjawisko emergencji? Nie znalazłam tej pozycji w Pana bibliografii. Jest tam natomiast książka Greena i Bossomaiera z 2000, która jednakże na rynku wydawniczym funkcjonuje jako "T.R.J. Bossomaier and D.G. Green, Editors. Complex Systems. Cambridge University Press, 2000". Proszę o wyjaśnienie.
Dlaczego powołuje się Pan na nierecenzowane opracowanie Jarzynowski, 2007 jako podstawowy przewodnik, wprowadzenie do zjawisk krytycznych? Wspomniana praca to zlepek haseł i rysunków poprzeklejanych z prac różnych badaczy, przy czym żadna z tych prac nie jest w opracowaniu Jarzynowski, 2007 opisana. Praca ta bowiem nie ma bibliografii.
Proszę też o wyjaśnienie co to jest „Metropolis algorithm” oraz podanie literatury, z której Pan swoją wiedzę na ten temat czerpał.
- Strona 8: Proszę o uzupełnienie literaturowe do przedstawianej przez Pana tezy, iż Polska socjologia będąc w stanie kryzysu nie była w stanie przewidzieć kluczowej roli ruchu „Solidarność” w transformacji systemu.
- Strona 10 i dalsze: Dwukrotnie powołuje się Pan na pracę Jarzynowski, 2016, jednakże w spisie literatury takowej nie znajdują. Proszę o wyjaśnienie i uzupełnienie.
Dlaczego cytując pracę Brouwers, 2009, a dalej opisując ją w bibliografii ogranicza się Pan do jednego autora? Praca ta została przygotowana przez pięciu autorów.
Czy może Pan uzupełnić swoje opracowanie o podanie literatury, na podstawie której przygotował Pan opis metod stosowanych w badaniach nad układami złożonymi czyli metodę symulacji modelowania agentowego, metod obliczeniowych pochodzących z układów dynamicznych, miar stosowanych w układach sieciowych czy pojęć fizyki statystycznej? Proszę o uzupełnienie do rysunków 1 i 3 o ich źródła.
- Strona 13: Proszę o uzupełnienie literatury o wskazanie pozycji Anderson, 1972 dla „more is different”. Proszę o sprecyzowanie odniesienia do Mantenga, 1998.
- Strona 14: Proszę o wyjaśnienie pojęcia 'negative likelihood' oraz wszystkich niezdefiniowanych wielkości używanych przez Pana w formułach tej strony. Proszę o pełne objaśnienie metody, o której pisze Pan 'Bayesian statistics as Markov Chain- MC based estimation algorithms'. Proszę o wyjaśnienie dla cytowanej pozycji Jarzynowski, Morris, 2011.
Pozycja Makowiec, et al., 2006 jest niewłaściwie opisana w bibliografii.
- Strona 15: Strona ta prezentuje kompletnie nieczytelne rysunki. Ponadto w rozprawie nie ma do nich odniesienia. Czemu zatem ma służyć prezentowana tu zawartość?
- Strona 16: Proszę o wyjaśnienie krytycznego stanowiska społeczności klinicznej odnośnie stosowania zaproponowanych przez Pana analiz.
Proszę o rozwinięcie pojęcia 'Digital epidemiology' - techniki którą, jak Pan pisze w niniejszej rozprawie Pan rozwinął (a w mojej opinii rozwija) czyli o umiejscowienie Pana badań w epidemiologii.
Proszę o wyjaśnienie i zacytowanie źródła dla pojęcia 'minority game'. Jaki cel ma rozdział 'Game theory (the prisoner's dilemma)' dla tej rozprawy?

- Strona 17 i 18: Rozważania przedstawione na tych stronach są w większości ogólne, niepoparte naukowymi argumentami. Proszę o przedstawienie źródeł, na podstawie których Pan formułuje tu opinie.
- Strona 19: Proszę o wyjaśnienie, dlaczego i jak Rysunek 6 lewy powinien reprezentować diagram fazowy wody. Proszę też o wyjaśnienie, jakie treści ilustruje Rysunek 6 prawy, przedrukowany z pracy Mossong et al., 2008, niewłaściwie cytowanej w podpisie do Rysunku i nieujętej w zestawie literatury. Proszę o uzupełnienie opisu i bibliografii.
- Strona 20: Użyte jest cytowanie do pracy nieumieszczonej w spisie literatury: Jarynowski, ..., Rosiński, 2017. Błędnie jest cytowana praca Bujok, Fronczak, Fronczak, 2014. Proszę o uzupełnienie i wyjaśnienie.
- Strona 21: Zaprezentowane rysunki są słabo objaśnione i brakuje źródła skąd pochodzą. Proszę o wyjaśnienie, jak informacje prezentowane tu związane są z tezami doktoratu?
- Strona 22: Mimo, że modele Malthusa i Verhulsta to podstawowe modele dynamiki populacji, jednakże powinny zostać opisane bibliograficznie. Proszę o uzupełnienie.
- Strona 23: Treści ogólnikowe a cytowany manifest (Berman, Jarynowski, et al., 2016) jest niedostępny ani komercyjnie ani w dostępnie otwartym. Proszę o uzupełnienie literaturowe.
- Strony 24-28: Treści tu ujęte miały zapewne przedstawiać metody stosowane przez autora w badaniach. Jednakże zebrane informacje są typu instrukcji obsługi. Są zamieszczone ekrany używanych aplikacji ze wskazaniem klawiszy, które trzeba nacisnąć, aby uzyskać określony efekt, zamiast spodziewanej krytycznej oceny możliwości i ograniczeń oprogramowania. Brak jest wyjaśnień, w jaki sposób budowane są matematyczne zależności i jakimi metodami są one rozwiązywane w wykorzystywanych przez doktoranta aplikacjach.

W pracy (1) wymienionej na liście III rozprawy, znajduje się opis matematyczny z elementami zastosowanych metod numerycznych modelu współzawodnictwa gatunków symulowanego przez Vensim. Wskazane byłoby przedstawienie analogicznego opisu dla modeli ujętych w pracach przedstawianych jako rozprawa doktorska, do których zastosowano pakiet Vensim.

Na część pierwszą rozprawy „Modelowanie epidemiologiczne” składa się wprowadzenie do problemu modelowania epidemii wraz z opisem wykorzystania podstawowych modeli w badaniach propagacji rzeczywistych zakażeń w społeczeństwie o złożonych interpersonalnych kontaktach (strony 29-42 przewodnika) oraz trzy prace doktoranta z kolejno Aną Serafimovic, Fredrikiem Liljeros i promotorem Andrzejem Grabowskim.

W pracy (1) „*Studying possible outcomes in a model of sexually transmitted virus (HPV) causing cervical cancer for Poland*” Autorzy zajęli się propagacją wirusa HPV 16, wirusa przenoszonego drogą płciową i uznawanego odpowiedzialnym za 56% wystąpień raka szyjki macicy. Punktem wyjścia modelu i symulacji były dane o aktywności seksualnej zebrane w społeczeństwie Finlandii w 1992 roku, przytoczone w tabelach 1 i 2 pracy (1). Jednakże dane zaprezentowane przez Autorów są inne od tych, które zostały podane jako źródło. Tabele w pracy w

Barnabas et al., 2006 kończą się na opisie własności osób w wieku 40-49, podczas gdy tabele w omawianej pracy traktują populacje do 64 roku życia (wszystkie starsze roczniki są opisane danymi 40-44-latków). Proszę Doktoranta o wyjaśnienie, dlaczego mógł dokonać takiego przeniesienia własności oraz o przedyskutowanie wpływu tej decyzji na uzyskane wyniki. Innych danych nie mogłam zweryfikować, bowiem opis ich źródła był zbyt lakoniczny - Autorzy odsyłają na główną stronę GUS.

Badanie wpływu badań przesiewowych a także profilaktycznych szczepień zostało przeprowadzone w oprogramowaniu Vensim – środowisku zaprojektowanym do hybrydowego modelowania układów o ciągłej dynamice opisywanej systemem równań różnicowych czy różniczkowych zwyczajnych oraz układów z przejściami dyskretnymi. Projekt stosowanego modelu jest przedstawiony na nieczytelnym grafie (Rysunek 1), parametry użyte w symulacjach umieszczone są w tabeli 3, natomiast źródło modelu jest dostępne na stronie Autora. Autorzy włączyli w model specyficzną dla Polski strukturę wiekową, a także rozważyli efektywną strukturę czasową przeprowadzanych badań przesiewowych. Model został zweryfikowany poprzez porównanie wyników symulacji do danych historycznych raportujących nowe wystąpienia raka szyjki macicy w Polsce. Wyniki w ogólności są zgodne z przewidywaniem, iż najlepszą strategią jest szczepienie młodych dziewcząt. Niemniej, dzięki analizom prowadzonym na danych rzeczywistych Autorzy mogli ustalić spodziewaną liczbę nowych zachorowań na raka w poszczególnych grupach wiekowych, co mogłoby posłużyć przygotowaniu naszej służby zdrowia do stawienia czoła temu wyzwaniu.

W drugiej pracy „*Contact networks and the spread of MRSA in Stockholm hospitals*” Autorzy modelują potencjalne drogi rozprzestrzeniania się wirusa MARS w sieci kontaktów zbudowanej w oparciu rejestry wizyt pacjentów w określonych szpitalach Sztokholmu. Zestaw danych jest imponujący, to zestawienie obejmujące ponad dwumilionową populację. Do posiadanych danych Autorzy zastosowali standardowy test statystyczny: regresję logistyczną, badając zależność efektu, to jest zarażenia wirusem MARS, od czasu przebywania w szpitalach i przychodniach. Niepokój mój budzi opisana w pracy konstrukcja testu statystycznego, hipoteza zerowa H_0 , i w konsekwencji interpretacja otrzymanego wyniku testu. Standardowo, test statystyczny buduje się tak, by odrzucić hipotezę H_0 na rzecz hipotezy alternatywnej. Prawo do odrzucenia daje mała (<0.05) wartość p uzyskana dla tzw. statystyki testu – wielkości, na podstawie której oceniane jest prawdopodobieństwo wystąpienia próby o zadanej charakterystyce. Wówczas duża wartość p ($p \geq 0.05$) nie pozwala na odrzucenia hipotezy H_0 na poziomie istotności 0.05. Jak te standardowe interpretacje mają się do wyników testu opisanego w Tabeli 1 omawianego artykułu? Jaką miarę statystyczną użyto do przeprowadzenia testu regresji logistycznej, a zatem co opisuje podana tu p -value? Nie podano także, jakim pakietem statystycznym posłużono się w obliczeniach.

Dla wzmocnienia czasowego efektu bezpośredniego kontaktu zaproponowano procedurę osłabiania i wzmacniania prawdopodobieństwa zakażenia z parametrem $x=0.1$. Jak wartość tego parametru została dobrana i jak zmiany w tym parametrze wpływają na dalsze szacowania?

Errata do pracy, wprowadzona w odpowiedzi na zarzut Recenzenta uzupełnia główny wynik pracy, to jest statystykę ilości zakażeń uzyskaną z modelu zakażeń propagujących się na sztucznej sieci Rysunek 4 pracy, o statystyki rzeczywistego przebiegu tej epidemii oraz wynik

uzyskany w pracy magisterskiej, gdzie zakażenia rozważano na sieci rzeczywistych kontaktów. Errata ta, rozumiem oznacza także, iż Rysunek 4 znika z omawianej pracy.

Trzecia praca przedstawiona do oceny w tej części rozprawy, to praca Pana Jarynowskiego z Promotorem, zatytułowana "*Rumor propagation in temporal contact network from Polish polls*". Bazując na statystykach otrzymanych z danych zebranych w ramach projektu POLYMOD (Mossong et al. PLoS Med. 5, e74, 2008) Autorzy zasymulowali sieć osobników i kontaktów między nimi tak, by odwzorować rzeczywiste charakterystyki kontaktów międzyludzkich. Na planszy tak skonstruowanej sieci Autorzy obserwowali rozprzestrzenianie się plotki. W szczególności zbadali zależność zasięgu plotki od atrakcyjności plotki, czyli od szybkości jej propagacji i od tempa zaniku zainteresowania nią. Symulacje pokazały, że przeciwnie niż przy standardowej metodzie SIR propagacji wirusa, dynamiczne zmiany połączeń obniżają raczej zasięg plotki, natomiast wprowadzenie ważonych połączeń w sieci wspiera propagację plotki.

Część "*I Epidemiological modeling*" rozprawy to strony 29 do 42.

Uwagi szczegółowe do części I:

- Strony 29, 30: Treści zostały dość wiernie przekopiiowane z pracy magisterskiej doktoranta, a opracowane na podstawie znanego podręcznika „Wprowadzenie do biomatematyki” J. D. Murray, pozycji niewłaściwie opisanej w bibliografii. W przedstawieniu modelu brak jest definicji parametru α . Nie opisano także, co to jest „non-fatal disease”?
- Strona 31: Rysunek 17 jest niecytowany w tekście i nieopisane jest jego źródło (praca magisterska Doktoranta?). Także Rysunek 18 powinien mieć opis skąd został skopiowany. Brakuje opisu skąd Autor zaczerpnął dane rzeczywiste odpowiadające propagacji zaproszeń do STOP-ACTA. Proszę o wyjaśnienie i uzupełnienie.
- Strona 32: W pracy Jarynowski, Grabowski, 2015, nie znalazłam badań nad istnieniem rozwiązań modelu SIR. Proszę o wskazanie literatury, w której ten problem jest podnoszony. Proszę także o uzupełnienie literaturowe dla Rysunku 19.
- Strona 33: Wzór na prawdopodobieństwo przejścia $i \rightarrow j$ jest niewłaściwy. Nie wyjaśniono zawartości Rysunku 20. Czy jest to wykres zakażeń z przykładowej symulacji? Jeśli tak, to proszę o opisanie wszystkich parametrów symulacji. A jeśli jest to wizualizacja rzeczywistych zakażeń, to proszę o opis danych i relacji, które zostały ujęte w prezentowanym grafie.
- Strona 34: Cytowanie Nyczka, 2015, nie jest ujęte w bibliografii. Proszę o przedstawienie tej pozycji.
- Strona 35: Sekcja ‘State of art in epidemiological modeling’ jest ogólnikowa. W roku 2015 grupa naukowców z najlepszych ośrodków świata, zajmujących się stosowaniem symulacji w służbie zdrowia, opublikowała raport jak powinny wyglądać symulacje epidemiologiczne (Marshall DA et al., „Applying dynamic simulation modeling methods in health care delivery research – the SIMULATE checklist: report of the INSPOR simulation modeling emerging good practices task force”, *Value in Health*, 18, 5-16, oraz Marshall DA et al., „Selecting a dynamic simulation modeling method for health care delivery research—Part 2: Report of the ISPOR dynamic simulation modeling emerging good practices task force”, *Value in Health*, 18, 147-180.) Czy mógłby

Pan przedyskutować, w jakim stopniu Pana badania spełniają wytyczne z listy SIMULATE?

- Strona 36: Autor nie podaje źródła danych, które posłużyły do rozważań prowadzonych na temat afrykańskiego pomoru świń. Proszę o uzupełnienie. Rysunek 24 nie jest cytowany w tekście, wielkości wyświetlane na rysunkach nie są zdefiniowane. Proszę o wyjaśnienia.
- Strona 37: Ponownie użyto wiele ogólnikowych stwierdzeń. Proszę o uzupełnienia literaturowe. Powołuje się Pan na artykuł (Hohle, 2017), którego nie ma w bibliografii. Rysunek 25 nie jest cytowany w tekście. Nie ma też opisu źródła prezentowanych danych. Proszę o uzupełnienie. Praca Kleczkowski, et al., 2011 jest błędnie opisana. Proszę o podanie poprawnej informacji.
- Strona 38: Jako źródło wykresu przedstawionego na Rysunku 26 podana jest praca Jarzynowski, 2010b. Nie znalazłam w tej pracy ani tego rysunku, ani też definicji „medium scenario”. Proszę o uzupełnienie.
Autor prezentuje tu także wyniki swoich symulacji rozprzestrzeniania się w Polsce (powiatach Polski) wirusa afrykańskiego pomoru świń. Model bazuje na wielkości populacji świń, odległościach geograficznych oraz wielkości pokrycia leśnego powiatu. Wyniki tego modelu zaprezentowane na Rysunku 27 są zgodne z intuicją: niskie spożycie wieprzowiny i niska populacja świń ogranicza rozprzestrzenianie się wirusa. Model jest zaprezentowany bez podania źródeł danych.
- Strony 39-40: Autor uzasadnia ważność podjęcia tematu rozprzestrzeniania się infekcji bakteryjnych i wirusowych przekazywanych drogą płciową. Informacje – tezy, podane są bez żadnych odnośników literaturowych, a w zasadzie każde zdanie sformułowane na tych stronach wymaga takiego wsparcia. Proszę o właściwe uzupełnienie przedstawionych treści. W szczególności proszę o uzasadnienie dla stwierdzenia „HPV - a sexually transmittable virus infection is not only the main but also necessary risk factor for developing cervical cancer”.
- Strona 41: Proszę o szczegółowe uzasadnienie dla treści – tezy, sformułowanych w paragrafie pierwszym. Odnośnik literaturowy przy Rysunku 28 jest źle opisany w bibliografii. Rysunek 28 nie jest opisany w tekście. Proszę o uzupełnienia.
- Strona 42: Proszę o podstawowe wskazania literaturowe dla zagadnienia modelowania sieciowe meta-populacji (meta-population network modelling). Nie podana jest literatura definiująca model wieloszpitalny czy też algorytm Google Page Rank. Rysunek 29 nie jest cytowany w treści.

Na część drugą rozprawy „Sieci i modelowanie danych” składają się prace (4) (5) i (6) tworzące cykl opisujący badania przeprowadzone przez Autora i Andrzeja Budę dotyczące rynku fonograficznego.

W pracy (4)- „*Network structure of phonographic market with characteristic similarities between artists*” Autorzy w oparciu o publicznie dostępne dane zbudowali bazy tygodniowej sprzedaży płyt trzydziestu najczęściej kupowanych muzyków w świecie. Zgromadzone informacje posłużyły Autorom do wyznaczenia współczynnika korelacji $p(i,j)$ pomiędzy sprzedażą płyt dowolnej pary muzyków. Policzyli także dla każdej pary parametry połączenia $l(i,j)$. Istota

tej ostatniej relacji nie została wyjaśniona w pracy, dlatego też proszę o wyjaśnienie, co ta relacja reprezentuje.

Przedstawione w pracy (4) na Rycinach 5 i 7 oraz w Tabelach I, II i III zestawienia dotyczą korelacji o poziomie istotności $\alpha=0.05$. Natomiast hierarchie na Rycinach 4 i 8 są dużo bogatsze niż te reprezentowane na Rycinach 5 i 7. Proszę o wyjaśnienie.

W konkluzji Autorzy stwierdzają, że nie ma istotnych korelacji pomiędzy sprzedażą płyt a opiniami ekspertów.

Ekspercka znajomość mechanizmów funkcjonujących na rynku fonograficznym w połączeniu z wiedzą fenomenologiczną, zdobytą przez Autorów przy badaniu własności serii czasowych sprzedaży płyt, posłużyły do zaproponowania modelu rynku sprzedaży płyt, **praca (5)** - „*Dynamics of popstar record sales on phonographic market – stochastic model*”. Obserwowana empirycznie antypersystość serii czasowych z tego rynku (praca Andrzeja Buda, *Physica A* 391(21) 5153–5159, 2012, praca nie jest cytowana w rozprawie) stała się punktem wyjścia do modelowania zjawiska sprzedaży płyt w oparciu o proces typu Ornsteina-Uhlenbecka powracania do średniej, uzupełniony o losowe przejścia pomiędzy stanami rynku: spokoju i pobudzenia. W modelu uwzględniono wiele specyficznych dla rynku fonograficznego aspektów, w tym eksplozję sprzedaży wszelkich płyt danego artysty przy wydaniu nowego singla, a także rolę sprzedaży płyt przy okazji świąt Bożego Narodzenia i Nowego Roku. Model jest wieloaspektowy, a co za tym idzie wieloparametrowy. Do walidacji parametrów Autorzy posłużyli się własnościami obserwowanych serii, przy czym także wykorzystali nowoczesne techniki zgłębiania danych takie jak metodę największej wiarygodności czy funkcję hazardu. Niemniej, dla pełności opisu poprawności modelu brakuje testów jego stabilności. Proszę o krótką dyskusję, które z parametrów są dla modelu krytyczne?

Praca (6) „*Diffusion paths between product life-cycles in the European phographic market*” to opracowanie statystyczne przepływu upodobań względem hitów muzycznych pomiędzy dwunastoma europejskimi krajami. Zanalizowano czasowe zależności – czas zakazania, pomiędzy pikami największej popularności dla stu najpopularniejszych singli, które pojawiły się w okresie pięćdziesięciu ostatnich lat. W oparciu o własności indywidualnych singli Autorom udało się zbudować sieć przepływu zainteresowań pomiędzy wielbicielami muzyki popularnej, Rysunek 3. Zastosowali tu ciekawy algorytm błędzenia losowego po rozważanych krajach przy założeniu określonej przez dane charakterystyki czasowej przepływów.

Uzyskane wyniki wydają się być w zgodzie z powszechną wiedzą. Rynek brytyjski dominuje. Ciekawym spostrzeżeniem tej analizy jest wyróżniona rola Francji, która okazuje się być zarówno odporna na inwazję muzyki brytyjskiej jak i podatna na jej działanie w zależności od zastosowanego algorytmu klastrowania. Algorytm Louvain detekcji podziału sieci jest algorytmem zachłannym. Proszę zatem przedyskutować niebezpieczeństwa, jakie wiążą się z wynikami uzyskanymi w oparciu o ten algorytm i jak Pan sobie radził z tymi niebezpieczeństwami. Jak z tymi niebezpieczeństwami radzi sobie algorytm VOS? Jak to się ma do wyników uzyskanych w pracy (6). Rysunek 4 pracy, na którym ta dwojaka rola Francji jest przedstawiona, niestety jest nieczytelny, i w moim odbiorze, powyższej tezy nie prezentuje.

Praca (7) „*Rola analizy sieci społecznych w odkrywaniu narracyjnej struktury fikcji literackiej*” (A. Jarynowski, S. Boland) odnosi się do sieci społecznych konstruowanych przez czytelnika w oparciu o narrację literacką. Autorzy skonfrontowali ze sobą cztery sieci budowane z relacji

pomiędzy bohaterami danego opowiadania: (I) sieć odkrywaną przez czytelnika w trakcie czytania, (II) sieć uzyskaną z oceny tych relacji po przeczytaniu całości, oraz dwie metody automatycznego tworzenia sieci: (III) dwie postaci w tym samym zdaniu, versus (IV) w tym samym akapicie. Skrupulatnie opisane zostały metody uzyskania danych do analiz. Mimo stosunkowo małej próby uczestniczącej w badaniu (36 osób) i wykorzystania zaledwie jednego opowiadania, to wyniki są ciekawe. Przede wszystkim wysoka korelacja pomiędzy sieciami (I) i (III) i dużo niższa korelacja pomiędzy (II) i (III), nie pozwala na jednoznaczne wskazanie algorytmu automatycznej analizy treści. Badane sieci nie pozwoliły także na odtworzenie punktu kulminacji opowieści, punktu naturalnego dla każdej opowieści. Zatem, a może i na szczęście, analizowane sieci nie do końca pozwalają reprezentować rzeczywistą narrację. Wyniki ujęto w tabelach 1 i 2 i na Rysunku 5 przedstawiającego kompletnie nieczytelne grafy badanych relacji.

Na część „*II Networks and data modeling*” rozprawy składają się strony od 43 do 58. Tak jak i w poprzedniej części wprowadzenie jest ogólnikowe i z błędami. Autor opisowo przedstawia tu podstawowe charakterystyki niezbędne dla analiz sieci złożonych.

Uwagi szczegółowe do części II.

- Strona 43: Opisu pozycji Kulig at. el 2013 nie ma w spisie literatury. Proszę o uzupełnienie.
- Strona 44: W definicji ścieżki jest odwołanie do Rysunku 9, dlaczego? Proszę o bibliografię dla problemu klastrowania sieci.
- Strona 45: Proszę o uzupełnienie literaturowe do omawianej przez Pana własności Facebooka. Definicja rozkładu prawdopodobieństwa stopni wierzchołków odsyła do Rysunku 10. Nie rozumiem tego odesłania, proszę o wyjaśnienie. Wzór na współczynnik klastrowania węzła korzysta z oznaczeń niezdefiniowanych wcześniej. Proszę o uzupełnienie. Ponadto formuła ta nie jest formułą ogólną, stosuje się bowiem jedynie do grafów skierowanych. Rysunek 32 ma być ilustracją tej definicji. Proszę o wyjaśnienie, jak Rysunek 32 wiąże się z formułą, którą ilustruje.
- Strona 46: Proszę o udowodnienie tezy sformułowanej w pierwszych dwóch zdaniach paragrafu ‘Dynamics and temporal aspects’. Sieci zmieniające się w czasie, inaczej zwane sieciami ewoluującymi mają bogatą bibliografię i to nie jest jedynie praca Holme et al., 2012. Są tu na przykład wszystkie prace dotyczące dyfuzji krawędzi, a więc i model małego świata Watts-Strogatz z 1998 roku. Gdzie zaliczyłby Pan prace badające zjawisko ferromagnetyzmu na sieciach ewoluujących? Proszę zatem o wyjaśnienie podstawy dla twierdzenia, iż „*integrative framework of both theoretical and empirical approaches is missing*”. Proszę też o uzupełnienie bibliograficzne dla listy o zastosowaniach sieci ewoluującej.
- Strona 47: Proszę o objaśnienie, jak czytać Rysunek 35, czemu on służy i wskazanie, gdzie jest opisany w rozprawie.
- Strona 48: Wiadomo jest, że konstrukcja minimalnego drzewa spinającego nie jest jednoznaczna. Proszę o dyskusję, jaki to wpływ ma na prezentowane przez Pana wyniki.

Czy mogę prosić o dyskusję statystycznej istotności dla prezentowanych współczynników korelacji dla poparcia pomiędzy partiami politycznymi? Zaprezentowane relacje na Rysunku 36 uważam za przypadkowe.

Na **stronach od 49 do 58** Autor prezentuje historię użycia metod sieciowych w naukach socjologicznych. Dyskutuje wzajemny wpływ i przepływ tych metod w fizyce czy matematyce i socjologii. Historia ta ujęta jest ciekawie, ale niestety tu widoczne są braki literaturowe i błędy bibliograficzne. Ponownie pojawia się pozycja Barabasi et al., 1999 nieistniejąca w spisie literatury. Nie ma w spisie literatury także opisu Watts, Strogatz, 2000, ani też opisu Jarynowski, Kułakowski, 2011. Praca Coleman, Katz, and Menzel, 1957 jest niewłaściwie cytowana. Rysunek 37 jest nieczytelny, słabo objaśniony, a w tekście nie ma do niego odwołania. Zapewne stanowi ilustracje do badań Autora nad zamieszkami w Sztokholmie, które miały miejsce w maju 2013. Niemniej ich zawartość w obecnym ujęciu jest niewyraźna.

Autor podsumowując swoje osiągnięcia, **strony 59 i 60**, stwierdza, że główny atut jego badań to zastosowanie nowoczesnych narzędzi układów złożonych do rozwiązania problemów świata rzeczywistego. Zestawia uzyskane wyniki w postaci analizy - tabeli SWOT, wyliczając mocne i słabe strony prowadzonych badań oraz ich szanse i zagrożenia. Jest tu przez Autora zgłoszone zagrożenie nadinterpretacji wyników uzyskanych często bez stosownej analizy. Zgadzam się z tą opinią. Uważam, że analizy bywają powierzchowne. W mojej ocenie dodatkowo, podnoszone przez Autora problemy świata rzeczywistego bywają postrzegane emocjonalnie. Jak wcześniej wielokrotnie dawałam przykłady, Autor niewystarczająco konfrontuje założenia i dalej wyniki swoich analiz z aktualną wiedzą na dany temat.

Strony 61-69 to zestawienie głównych wyników osiągniętych w załączonych do rozprawy pracach. Pan Jarynowski uzupełnił to zestawienie o akapity zawierające opis metod gromadzenia i przetwarzania danych. Niemniej pozostały wszystkie błędy poprzedniej wersji rozprawy, które to stanowią o szczególnej uciążliwości przy odczytaniu zsyntetyzowanych tutaj informacji.

- Rysunek 39 jest inny od schematu obliczeń zaprezentowanego w pracy (1). Proszę o wyjaśnienie, dlaczego? Proszę też o systematyczne wyjaśnienie, jakie przybliżenia zostały zastosowane w obliczeniach.
- Rysunek 40: Indicence?
- Formuła określająca zmianę stanu jednostki jest inna od tej przedstawionej w pracy. Proszę o wyjaśnienie, która formuła była stosowana w obliczeniach.
- Na stronie 64 jest odwołanie do pracy Mossong et al., 2008, której to nie ma w bibliografii. Proszę o wyjaśnienie. Nie znalazłam także opisu pracy Grabowski, Jarynowski, 2016 cytowanej na stronie 65. Proszę o wyjaśnienie pojęcia 'unitary distribution', użytego na stronie 66, a także w pracy (5).
- Na stronie 69 przedstawiono procedurę edycji danych uzyskanych od ankietowanych. Proszę o dyskusję, jaki wpływ na wyniki miały wprowadzone poprawki. Rysunek na tej stronie jest kompletnie nieczytelny. Dlatego, proszę też o wyjaśnienie, jaki był cel umieszczania tu takiego rysunku.

Bibliografia, zaprezentowana na stronach 70-74 zamyka rozprawę. Opis w poszczególnych pozycjach nadal nie jest jednolity, choć Autor deklaruje na liście wprowadzonych zmian, że ujednoczenia stylu dokonał. Przykładem jest chociażby użycie symbolu &, czy stosowanie małych i dużych liter w tytułach artykułów. Wiele indywidualnych błędnych pozycji tej listy opisałam wcześniej choć wiele zapewne mi umknęło.

Podsumowanie:

Rozprawa Pana Jarynowskiego ma charakter interdyscyplinarny. Jest to praca na granicy dziedzin: epidemiologii i układów dynamicznych o niejednorodnej strukturze, czy też socjologii i psychologii społecznej, w której interesujące badacza relacje społeczne są wizualizowane w postaci grafu sieciowego związków, by następnie być analizowane metodami wywodzącymi się z układów złożonych. W szeregu swoich badań Autor zastosował nowatorskie miary dla badanych zjawisk. Mogłyby zatem wyniki tej pracy być inaczej odbierane na przykład przez grono epidemiologów czy socjologów. Tym bardziej, że w rozprawie Pan Jarynowski postawił na opisy badanych zjawisk, czy to biologicznych, czy to społecznych, czyli na tematykę, która pozostaje poza moją- fizyka, kompetencją oceny. Mniej natomiast doktorant zajął się prezentacją i oceną technik zbierania i klasyfikacji danych, a także krytyczną oceną metod obliczeń, którymi posługują się proponowane przez niego modele. Można to próbować tłumaczyć efektem osobistego entuzjazmu Autora do prowadzonych badań oraz jego młodzieńczą niecierpliwością w dotarciu do wyniku.

Wiele treści przedstawianych przez mgra Andrzeja Jarynowskiego w jego kolejnych publikacjach jest powtarzane wielokrotnie i to bezkrytycznie. Korzystałam z tego faktu, kiedy to, na przykład, nieczytelne rysunki z niniejszej rozprawy odnajdywałam w innych pracach Autora. Zauważyłam przy tym powielanie tych samych błędów merytorycznych i redakcyjnych. Właśnie takie powielane błędne treści niszczą zaufanie do Autora, wierzę, iż przedstawione wyniki zostały uzyskane z największą starannością, przy pełnym wykorzystaniu dostępnych Autorowi środków i wiedzy. Tak też jest z redakcją niniejszej rozprawy, która w wielu miejscach jest wierną kopią poprzedniej wersji wraz z niedostatkami poprzedniczki, w tym także z ogromną ilością tych samych błędów ortograficznych i gramatycznych, o których wspominały obie wcześniejsze recenzje.

Konkluzja:

Przedstawiona mi do oceny rozprawa mgra Andrzeja Jarynowskiego „*Exploring the dynamics and the complexity of human behavior using nonlinear physics methods*” nie spełnia zwyczajowo uznanych norm oczekiwanych od rozpraw doktorskich. Rozprawa ta jednakże zadowala ustawowe wymogi stawiane rozprawom doktorskim. W oparciu o powyższe wnoszę o dopuszczenie Pana Andrzeja Jarynowskiego do dalszych etapów przewodu doktorskiego.

Joanna Maliniewicz