

Dr hab. Tadeusz Pankowski
Instytut Automatyki i Inżynierii Informatycznej
Politechnika Poznańska
Pl. M. Skłodowskiej-Curie 5
60-965 Poznań

Recenzja rozprawy doktorskiej pt.:
Sieci bezskalowe – bezpieczeństwo i zastosowania
Autor: mgr Marek Gałązka

1. Uwagi wstępne i układ pracy

Opiniowana rozprawa ma 60 stron i składa się ze wstępu, czterech rozdziałów oraz wykazu literatury obejmującego 70 pozycji, gdzie w dwóch przypadkach doktorant jest jedynym autorem, a w jednym współautorem. Do pracy dołączono płytę CD zawierającą dwa programy opracowane na potrzeby pracy.

We wstępie doktorant stwierdza, że jednym z naturalnych kierunków badań informatyki teoretycznej jest badanie sieci rzeczywistych za pomocą modeli grafowych. Zauważa, że ważnym przypadkiem sieci obserwowanych w praktyce są tzw. *sieci bezskalowe* charakteryzujące się potęgowym rozkładem stopni wierzchołków. Stwierdza, że wiele sieci obserwowanych w świecie rzeczywistym ma cechy bezskalowości. Wskazuje też na ich specyfikę w porównaniu z klasycznymi grafami losowymi. Jako cel rozprawy wskazuje analizę sieci bezskalowych obejmującą problemy bezpieczeństwa w tych sieciach (rozumianych jako odporność na losowe uszkodzenia oraz wrażliwość rozumianą jako podatność na złośliwe ataki) oraz wybrane zastosowania sieci bezskalowych.

W rozdziale 1 autor omawia problematykę sieci bezskalowych z uwzględnieniem historii rozwoju teorii tych sieci w szczególności z uwzględnieniem wpływu, jaki na ten rozwój wywierały sieci obserwowane w świecie rzeczywistym, w tym zwłaszcza sieci społeczne. Opisuje także cechy, które odróżniają modele oparte na sieciach bezstanowych od modeli klasycznych, a mianowicie: (a) potęgowy rozkład stopnia wierzchołków, (b) wysoki współczynnik skupienia i (c) tak zwany "efekt małego świata". Następnie przytacza sposób generowania grafu bezskalowego oparty na metodzie LCD zaproponowanej przez Bollobasa i Riordana.

Rozdział 2 poświęcony jest problematyce bezpieczeństwa sieci bezskalowych. Stawia pytanie "Jak odporne są sieci bezskalowe?". W poszukiwaniu odpowiedzi na to pytanie autor opracował oprogramowanie zamieszczone na płycie CD. Program generuje graf bezskalowy, a następnie przeprowadza symulacje ilustrujące efekt losowego usuwania wierzchołków i krawędzi. Następnie formułuje i dowodzi szereg lematów dotyczących liczby wierzchołków izolowanych powstających w grafie bezskalowym w wyniku procesów usuwania z niego wierzchołków i krawędzi. W rozdziale tym formułuje też i udowadnia twierdzenie dotyczące zależności między stopniami a wysokościami wierzchołków w grafach bezskalowych.

W rozdziale 3 koncentruje się na metodach wyznaczania parametru γ , który jest podstawową charakterystyką rozkładu potęgowego. Rozkład ten stwierdza, że prawdopodobieństwo, iż losowo wybrany wierzchołek ma stopień k , jest w przybliżeniu równe k do potęgi $(-\gamma)$. Zwraca uwagę na problemy poprawnego wyznaczenia tego parametru w sieciach

empirycznych. W podrozdziale 3.3 proponuje nową metodą, która bazuje na znanych metodach, ale eliminuje pewne ich wady. Zastosowanie proponowanej metody ilustruje opracowanym przez siebie przykładem sieci rozgrywek sportowych w ramach pucharu UEFA.

W rozdziale 4 autor przedstawia zastosowanie grafów bezskalowych do wybranych problemów występujących w świecie rzeczywistym, a konkretnie w ekonomii. Koncentruje się przy tym na prognozowaniu cen akcji na rynkach finansowych. Wprowadza w tym celu model rynku akcji posiadający właściwości grafów bezskalowych, gdzie krawędzie grafu odzwierciedlają korelacje krzyżowe między notowaniami akcji na warszawskiej Giełdzie Papierów Wartościowych. Proponuje w tym celu ważne grafy bezskalowe stanowiące ważony model rynku akcji, gdzie waga przypisana krawędzi grafu odpowiada współczynnikowi korelacji krzyżowej między spółkami powiązanymi rozpatrywaną krawędzią.

2. Ocena ogólna: Uwagi pozytywne

Przedstawiona rozprawa spełnia postawione przed nią zadania: analizuje sieci bezskalowe, bada ich formalne właściwości w tym ich bezpieczeństwo rozumiane jako odporność na zachodzące w nich zmiany, a także pokazuje nietrywialne zastosowania sieci bezskalowych w modelowaniu wybranych zjawisk i sytuacji w świecie rzeczywistym.

Badania autora nawiązują do bogatej literatury przedmiotu odnoszącej się zarówno do klasycznej teorii grafów, jak i do teorii i praktyki sieci bezskalowych, których intensywny rozwój datuje się od około 20 lat i związany jest zwłaszcza z rozwojem internetu. Uważam, że rozprawa stanowi istotny wkład w rozwój zarówno teorii sieci bezskalowych, jak i ich zastosowań. Kompozycja rozprawy jest zwarta i logiczna. Jasno rozdzielono część dotyczącą podstaw teoretycznych sieci bezskalowych i część dotyczącą ich zastosowań.

Ważne wyniki teoretyczne zawarto w rozdziale 2. Przytoczone w tym rozdziale lematy, twierdzenia i ich dowody można uznać za istotne osiągnięcie autora. Tym bardziej, że rezultaty te zostały opublikowane i są cytowane. Lematy dotyczą problemów bezpieczeństwa rozumianego jako zdolność grafu do zachowania spójności przy losowym usuwaniu z niego wierzchołków i/lub krawędzi oraz oszacowania wartości oczekiwanej liczby wierzchołków izolowanych. Ważnym problemem przy modelowaniu procesów rzeczywistych za pomocą modeli grafowych jest badanie zależności między budową sieci (jej strukturą) a dynamikę zachodzących w niej zjawisk. Problem ten doktorant bada na przykładzie szybkości rozprzestrzeniania się plotki w sieci rozległej, co jest formalizowane jako badanie zależności między stopniami wierzchołków a ich wysokościami. Podsumowanie tych rozważań zawiera twierdzeniem 2.12 oraz wniosek 2.13.

Wyniki, ważne z praktycznego punktu widzenia, zawarto w rozdziale 3. Główny problem tego rozdziału to metoda wyznaczenia parametrów rozkładu potęgowego związanego z siecią bezskalową. W podrozdziale 3.1 autor analizuje metody graficzne i wskazuje ich wady. Pokazuje w szczególności metody wykorzystywane przy badaniu struktury sieci WWW. Okazuje się, że rozkłady stopnia wejścia i wyjścia wierzchołków sieci WWW spełniają wymagania rozkładu potęgowego, gdzie wartość γ jest równa odpowiednio 2,1 i 2,7. W celu weryfikacji graficznej metody wyznaczania parametru γ rozkładu potęgowego w sieciach empirycznych, doktorant opracował program komputerowy, za pomocą którego przeprowadził szereg eksperymentów. Celem tych eksperymentów było wykazanie niewiarygodności tzw. metody graficznej. Metody określone jako "wiarygodne testowanie rozkładu potęgowego" omawiane są w podrozdziale 3.2. Autor przedstawia znane metody i wykorzystuje je do zaproponowania własnej procedury do analizy rozkładu potęgowego dla sieci rzeczywistych.

Autor omawia metody Goldsteina i Newmana oraz omawia ich wady. Za główną wadę uważa fakt, że metody te można stosować przy wartości γ większej niż jeden ($\gamma > 1$),

co wynika z właściwości stosowanych w nich funkcji dzeta Riemanna. Następnie, w podrozdziale 3.3, autor przedstawia własną tzw. empiryczną metodę wyznaczenia parametrów rozkładu potęgowego, tzn. parametrów k_{min} i γ , oraz ich testowania metodą Kołmogorowa-Smirnowa. Jako p -wartość w tym testowaniu przyjmuje $p = 0,1$. Metodę tę autor ilustruje badaniem sieci rozgrywek piłkarskich UEFA. Wykazuje, że badana sieć ma rozkład potęgowy z parametrami k_{min} i γ równymi odpowiednio 5 i 2,58.

Podstawowe wyniki dotyczące zastosowań sieci bezskalowych zawarte są w rozdziale 4, gdzie analizowane są wybrane sieci bezskalowe w ekonomii. Autor koncentruje się na rynku kapitałowym, a w szczególności na znajdowaniu przykładów występowania rozkładu potęgowego w zjawiskach ekonomicznych. Jako przykład obiera prognozowanie zmian cen akcji poszczególnych spółek oraz indeksów notowanych na giełdzie. W tym celu wprowadza model ważonego grafu rynku akcji (WSG). Badania przeprowadził na danych z Warszawskiej Giełdy papierów Wartościowych za okres styczeń – grudzień 2007 dla akcji 252 spółek. W tym celu zdefiniował współczynnik korelacji krzyżowej pomiędzy różnymi spółkami z uwzględnieniem różnych czynników wpływających na wahania kursów akcji (np. inflacja, stopy procentowe, kursy walutowe). Zastosował model ważonej sieci bezskalowej, gdzie wagi przypisane są krawędziom, a waga wierzchołka jest sumą wag krawędzi incydentnych z tym wierzchołkiem. Pokazuje, że rozkład wag wierzchołków jest wówczas zgodny z pewnym rozkładem potęgowym. Autor bada graf pełny, gdzie wierzchołkami są spółki notowane na GPW, a krawędzie między spółkami opisują wpływ tych spółek na siebie nawzajem za pomocą korelacji dziennych zwrotów notowań na zamknięcie dnia. Następnie konstruuje minimalne drzewo rozpinające oparte na metryce określającej odległość między wierzchołkami zdefiniowaną na podstawie współczynnika korelacji krzyżowej. Doktorant uzasadnia, że rozkład prawdopodobieństwa sił wpływu danej spółki na inne jest zgodny z rozkładem potęgowym o wykładniku potęgowym równym 2,7.

3. Ocena ogólna: Uwagi krytyczne

W pracy nie znalazłem istotnych błędów merytorycznych. Moje zastrzeżenia wynikają głównie z dwóch powodów: po pierwsze, zbyt skrótowo potraktowano niektóre kwestie przez co powstają niejasności interpretacyjne, a po drugie – rozważanie prowadzone są w sposób niejednolity – są fragmenty bardzo precyzyjne, ale są również fragmenty potraktowane bardzo pobieżnie.

Co właściwie ma wykazać eksperyment opisany w podrozdziale 2.1? Autor pisze, że opracował oprogramowanie do generowania grafu bezskalowego. Wykorzystał je też do badania odporności grafu na usuwanie wierzchołków. Następnie stwierdza (str. 11): *"Przeprowadzone symulacje pokazały, że losowe usuwanie wierzchołków będzie dotyczyć tych o małym stopniu, ponieważ jest ich znacznie więcej niż hubów."* Nie bardzo rozumiem po co przeprowadzono symulacje? Czyż nie jest oczywiste, że przy losowym wyborze elementów częściej będziemy wybierać elementy ze zbioru bardziej licznego niż ze zbioru mniej licznego? Czy to trzeba potwierdzać eksperymentalnie? Dalej stwierdza *"eliminacja takich wierzchołków [o małym stopniu] nie rozrywa sieci w sposób znaczący"*. A jaki stopień "rozerwania" jest "znaczący"? Dalej mówi o "rozłączaniu systemu" przez usuwanie *"paru kluczowych hubów"*. Mamy tu do czynienia z potocznymi sformułowaniami, które w konfrontacji z innymi, bardzo sformalizowanymi fragmentami pracy, wprowadzają dysonans. Po przeprowadzeniu eksperymentów symulacyjnych polegających na usuwaniu wierzchołków i krawędzi, autor stawia hipotezę: *"Sieć bezskalowa stanie się niespójna w czasie, kiedy któryś z wierzchołków zostanie odizolowany"*. Stwierdzenie to brzmi jednak jak definicja braku spójności – dlaczego więc nazywać to hipotezą? Jaki to ma związek z przeprowadzonymi doświadczeniami symulacyjnymi? Podobnie niejasno brzmi kolejne zdanie tej hipotezy: *"Izolowane wierzchołki mają*

jedne z najmniejszych stopni w pierwotnym grafie, co jest całkiem oczywiste". Hipoteza powinna być nietrywialnym stwierdzeniem, które wymaga przeprowadzenia jakiegoś uzasadnienia (dowodu) w sposób formalny lub eksperymentalny. W tym przypadku trudno mi znaleźć uzasadnienie dla sensowności formułowania przytoczonej hipotezy. Na rysunkach 2.1, 2.2 i 2.3 pokazano wyniki symulacji. Jednak rysunki te nie są właściwie opisane. Jak należy interpretować osie na tych rysunkach? W lemacie 2.1 nie podano co oznacza lambda (λ). Wyjaśnienie pojawia się dopiero w dowodzie. I dlaczego ożywa się λ , zamiast ciągle w pracy używanego parametru gamma (γ)?

Celem eksperymentów opisanych w podrozdziale 3.1 było wykazanie niewiarygodności metody graficznej. Jednak zarówno opis przeprowadzonych eksperymentów, jak i sposób opisu ich wyników nie jest do końca przekonujący. W szczególności opis wyników prezentowanych w tabeli 3.1. budzi niedosyt – chodzi głównie o traktowanie odchylenia standardowego jako miary wiarygodności metody.

W podrozdziale 3.2.3, autor stwierdza, że ograniczeniami metod Goldsteina i Newmana jest przyjęcie założenia, że $\gamma > 1$. W jaki sposób autor unika tego ograniczenia, bo przecież w podrozdziale 3.3 stwierdza, że do wyznaczenia parametrów k_{min} i γ , używa metody opisanej w podrozdziale 3.2.2, a więc metody Newmana?

Uwagi redakcyjne:

1. Ważne w pracy pojęcie "bezpieczeństwo" nie zostało właściwie zdefiniowane. Pojawiają się co prawda w różnych miejscach pracy wyjaśnienia tego terminu, ale niedosyt pozostaje.
2. Niezręczne sformułowanie (str. 34) "*istnieje możliwość hipotezę o istnieniu rozkładu potęgowego wykluczyć*". Niepoprawność gramatyczna. Ale także logiczna – raczej możemy odrzucić hipotezę, że badana zmienna losowa ma rozkład potęgowy a nie hipotezę "*o istnieniu rozkładu potęgowego*".
3. Zamiast "*niech jest*", należy używać "*niech będzie*" (str. 35).
4. Brak w pracy rozdziału zawierającego podsumowanie i wnioski.

4. Konkluzja

Powyższe uwagi krytyczne nie mają większego znaczenia dla mojej pozytywnej oceny rozprawy doktorskiej pana mgra Marka Gałązki. Uwzględniając wagę i aktualność tematyki poruszanej w rozprawie, a także oryginalność prezentowanych rozważań i uzyskanych wyników, stwierdzam co następuje:

1. Problem naukowy rozprawy sytuuje się w obszarze badań informatyki teoretycznej i jej zastosowań. W pracy przedstawiono analizę sieci bezskalowych głównie z punktu widzenia ich bezpieczeństwa (rozumianego jako odporność na losowe uszkodzenia i podatność na złośliwe ataki). W części teoretycznej pracy sformułowano i udowodniono szereg lematów i twierdzeń. W części praktycznej rozprawy pokazano zastosowanie teorii sieci bezskalowych do modelowania systemów rzeczywistych w zakresie zastosowań ekonomicznych.
2. Rozprawa stanowi oryginalne rozwiązanie problemu naukowego oraz wykazuje właściwy poziom wiedzy teoretycznej, a także świadczy o umiejętności samodzielnego prowadzenia pracy naukowej przez doktoranta.
3. Przedstawiona rozprawa doktorska spełnia ustawowe warunki ubiegania się o stopień doktora w myśl *Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i*

tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz. U. Nr 65, poz. 595, Art. 13, ust. 1), w tym przypadku doktora nauk matematycznych w zakresie informatyki.

4. Wnoszę o dopuszczenie rozprawy doktorskiej pana mgr Marka Gałązki do publicznej obrony.

A handwritten signature in blue ink, appearing to read 'Gałązki', is written in a cursive style.