

Warszawa, 11.05.2017

Dr hab. Adam Wierzbicki
Profesor Polsko-Japońskiej Wyższej Szkoły
Technik Komputerowych

OPINIA
O ROZPRAWIE DOKTORSKIEJ

MGR INŻ. BOGDANA GLIWY

“ANALIZA DYNAMIKI GRUP W MEDIACH
SPOŁECZNOŚCIOWYCH”

1. Jakie zagadnienie naukowe jest rozpatrzone w pracy i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Rozprawa dotyczy problemu modelowania grup za pomocą sieci społecznych. Sieć społeczna jest traktowana jako zadany model całej społeczności. Natomiast w tej sieci mogą istnieć grupy uczestników sieci. Grupy te definiowane są w różny sposób, zależny jednak na ogół od struktury wewnętrznej (łącz pomiędzy uczestnikami grupy) oraz łącz od uczestników grupy do pozostałych uczestników społeczności. Jednym z problemów informatycznych jest znajdowanie grup w zadanej sieci.

Sieć nie jest statyczna, lecz temporalna (zmienna w czasie). Grupy zatem mogą także ulegać zmianom. Drugim problemem rozpatrywanym w rozprawie jest wykrywanie zmian grup różnego typu (podział, połączenie, odejście lub przybycie nowych uczestników grupy).

Problemy badawcze poruszane w rozprawie to:

- Określanie miar do porównywania metod znajdowania grup oraz wykrywania zmian grup
- Opracowanie nowych metod wykrywania oraz predykcji zmian grup
- Opracowanie nowych metod wizualizacji zmian grup
- Opracowanie nowych metod wykrywania tematyki grup w sieciach, których uczestnicy wymieniają między sobą informacje

Rozprawa ma charakter empiryczny oraz przedstawia nowe miary służące do oceny metod wykrywania grup oraz wykrywania zmian grup. Rozprawa przedstawia także nowe metody wykrywania zmian grup, mające zastosowanie w

marketingu Internetowym. Identyfikacja grup oraz analiza ich tematyki pozwala wyłaniać grupy, które mogłyby być zainteresowane kampanią marketingową. Natomiast predykcja zmian w grupie umożliwia stwierdzenie, czy grupa jest trwała, co ma znaczenie przy podejmowaniu decyzji, czy kierować do grupy przekaz marketingowy.

Zagadnienia naukowe rozpatrywane w pracy zostały jasno sformułowane przez autora. Teza rozprawy jednakże sformułowana jest zbyt ogólnie: „*Observation of the characteristics of groups forming in social media allows for the systematic analysis of content spread through them, modeling group dynamics and predicting their future behavior.*”. Rozprawa nie zawiera opisu badań poświęconych propagacji bądź rozprzestrzenianiu się informacji w sieciach społecznych (co jest bardzo popularną tematyką badań), jak sugeruje sformułowanie tezy. Niejasność ta wskazuje na ograniczony sens tradycyjnego formułowania ogólnej „tezy” doktoratu. Na szczęście doktorant przedstawił także szczegółowe zestawienie wyników osiągniętych w rozprawie.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczącej o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Rozprawa zawiera bibliografię o 236 pozycjach, która skupia się przede wszystkim na literaturze światowej z dziedziny. Stan wiedzy omawiany jest w oddzielnym rozdziale (Rozdział 3: „Selected contemporary methods related to Group Analysis in Social Networks”). Jest to dobry przegląd literatury, który dobrze organizuje aktualną wiedzę. Jednak brak jest sformułowania wniosków z przeglądu literatury, które jasno motywowałyby prowadzenie dalszych badań. Brak ten jest częściowo skompensowany w dalszych rozdziałach, choć i tam uzasadnienie prowadzonych badań jest zdawkowe.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Autor rozpoczął pracę od przeglądu sposobów oceny metod wykrywania grup oraz zmian grup. Następnie opracował własne metody, porównał je z istniejącymi i uzasadnił ich sens. Wprowadził własną miarę oceny wykrywania grup o nazwie CMatrix. Wprowadził także zestaw miar do porównywania metod wykrywania zmian grup o nazwie GEM. Zaproponował także rozszerzenie powszechnie stosowanej metody generowania przykładowych sieci w celu testowania metod znajdowania grup – rozszerzenie to nosi nazwę eLFR, oraz do testowania metod wykrywania zmian grup (GevBen). Wprowadzenie tych metod oznacza stworzenie solidnych podstaw do empirycznej analizy proponowanych metod algorytmicznych wykrywania grup i wykrywania zmian grup. Dlatego uznaję, że autor użył właściwych metod badawczych.

Dzięki solidnym podstawom, autor rozprawy przekonująco uzasadnia uzyskanie lepszych wyników dzięki zaproponowanej przez siebie nowej metodzie wykrywania zmian grup, SGCI.

Jednakże nie dotyczy to zaproponowanego narzędzia do wizualizacji zmian grup, GEVi. Proponowane narzędzie nie zostało poddane analizie użyteczności, ani ocenie interfejsu użytkownika.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Najważniejszym oryginalnym osiągnięciem autora jest opracowanie nowej metody wykrywania zmian grup, SGCI, oraz metody predykcji zmian grup. Metoda SGCI została porównana z innymi popularnymi metodami – Asur, GED, Greene oraz Takaffoli. Porównanie zostało wykonane na syntetycznych zbiorach generowanych przez eLFR benchmark (w takim przypadku znamy oczekiwany rezultat określany jako „ground truth” i możemy określić, która metoda najlepiej modeluje dynamikę grup) oraz na rzeczywistych zbiorach z dziedziny blogosfery (Salon24, Huffington Post). Do porównania rezultatów różnych metod użyte zostały miary GEM. Przeprowadzone eksperymenty na sztucznych zbiorach pokazały bardzo dużą dokładność metody SGCI w stosunku do pozostałych metod. Metoda SGCI słabiej sobie radzi na zbiorach syntetycznych generowanych przez eLFR benchmark w sytuacji, gdy prawdopodobieństwo migracji osób jest duże lub jest duży szum (w takiej sytuacji ma gorsze wyniki niż wersja metody Greene bez odległych przejść tj. zdarzeń na niesąsiednich slotach czasowych). Pod względem wydajności jest także gorsza od metody Greene, natomiast lepsza od pozostałych testowanych metod.

Predykcja złączeń grup jest badana po raz pierwszy w literaturze światowej. Eksperymenty zostały przeprowadzone na 5 zbiorach rzeczywistych. Wyniki pokazały, że rozpatrywane miary w tym zagadnieniu wykazywały własności predykcyjne.

W zagadnieniu predykcji zdarzeń grupy zostały zaproponowane nowe metryki. Ich użyteczność została zbadana przy użyciu selekcji cech w zagadnieniu predykcji, analizy cech pod kątem AUC (pole powierzchni pod krzywą ROC) oraz korelacji cech. Eksperymenty zostały przeprowadzone na 5 rzeczywistych zbiorach.

Kolejnym opisanym w rozprawie osiągnięciem jest opracowanie metody znajdowania tematyki grup (GTA). Jednakże metoda ta posługuje się standardową metodą Latent Dirichlet Allocation (LDA), aby wykryć tematy dyskusji w całej sieci (w sposób półautomatyczny, bowiem proponowane tematy muszą być zweryfikowane przez człowieka), a następnie przypisuje wykryte tematy do wykrytych grup. Nie jest to w moim odczuciu znaczna innowacja, aczkolwiek metoda GTA została opublikowana w literaturze światowej.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienie uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Rozprawa jest dobrze zredagowana. Praca została napisana w języku angielskim. Autor w rozprawie wielokrotnie posługuje się pojęciem „metric”, określając w ten sposób wprowadzone przez siebie miary jakości metody wykrywania grup oraz zmian grup. Jednakże w rozprawie nie uzasadniono, że wprowadzone miary spełniają aksjomaty metryki (nierówność trójkąta). Lepiej byłoby posłużyć się terminem „measure”.

6. Jakie są słabe strony rozprawy i jej główne wady?

Najsłabszym elementem rozprawy jest narzędzie GEVi do wizualizacji zmian grup. Narzędzie to zostało opracowane z pominięciem podstawowych zasad projektowania interfejsu użytkownika. Nie zostały opisane w pracy żadne metody projektowania tego narzędzia, jego specyfikacja lub założenia.

Narzędzie wizualizacyjne GEVi jest mało intuicyjne. Korzystające z niego, trudno jest zrozumieć, w jaki sposób określona grupa zmienia się w czasie, ponieważ grupy zmieniają identyfikatory w różnych momentach czasowych, a zmiana członkostwa grup jest trudna do zrozumienia. Jest to szczególnie dziwne, ponieważ celem GEVi jest właśnie wizualizacja zmian grup w czasie. Autor nie przeprowadził analizy użyteczności dla proponowanego narzędzia.

GEVi nie pokazuje także wewnętrznej spójności grup i siły łącz pomiędzy nimi, co jest funkcją jednego z konkurencyjnych narzędzi (SnailVis) opisanego w rozprawie. Taka funkcjonalność jest szczególnie ważna w sieciach społecznych, w których pojęcie grupy nie jest precyzyjne, bowiem ten sam węzeł może należeć do wielu grup. Wizualizacja zmian siły łącz wewnątrz i pomiędzy grupami w czasie byłaby także przydatna dla pokazania, jak grupa staje się bardziej lub mniej spójna. GEVi w ogóle ignoruje fakt, że grupy istnieją w kontekście łączącej je sieci społecznej.

Narzędzie GEVi pozwala także na wizualizację „kontekstów”, które zdefiniowane są w pracy jako zadane z góry kategoryzacje grup (na przykład, według sentymentu lub tematu). Kategoryzacja nie jest binarna, ale ciągła, czyli grupa może posiadać określoną przez kontekst kategorię w jakimś stopniu. GEVi umożliwia wizualizację kontekstów grup, jest to jednak znowu niemal kompletnie nieczytelne. Przy użyciu narzędzia nie daje się na przykład wybrać grup o określonych kategoriach z wykresu czasowego, ani też porównać ze sobą intensywności posiadania określonych cech przez różne grupy (do czego można by na przykład użyć wykresów typu radio).

Sądzę, że GEVi zostało opracowane przede wszystkim jako narzędzie dla naukowców pracujących w dziedzinie, nie zaś dla „zwykłych użytkowników” (na przykład osób zajmujących się marketingiem Internetowym). Jeśli miało służyć takim odbiorcom, GEVi musi zostać znacznie poprawione. Opisane tu braki nie oznaczają jednak, że GEVi nie może służyć jako prototyp, inspirujący dla przyszłych rozwiązań.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Przydatność rozprawy dla nauk technicznych oceniam wysoko. Zaproponowane metody oceny i porównania algorytmów wykrywania grup oraz wykrywania zmian grup mogą być użyteczne dla opracowania lepszych algorytmów w przyszłości. Opracowana metoda wykrywania i predykcji zmian w grupach istotnie poprawia aktualnie znane rozwiązania.

WNIOSKI

Podsumowując stwierdzam, że **rozprawa spełnia wymagania stawiające rozprawom doktorskim przez obowiązujące przepisy. Uważam, że przedstawiona rozprawa spełnia warunki określone przez Ustawę z dnia 14 marca 2003 (wraz z późniejszymi zmianami) o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. 2003 Nr 65 poz. 595), dlatego też wnoszę o dopuszczenie mgr inż. Bogdana Gliwy do dalszych etapów przewodu doktorskiego.**



Sekretariat WIEIT

06. 06. 2017 *jm*

data wpływu