

Anna ZYGMUNT, Jarosław KOŹLAK, Łukasz KRUPCZAK
Akademia Górniczo-Hutnicza, Katedra Informatyki

ANALIZA GRUP W SERWISACH SPOŁECZNOŚCIOWYCH¹

Streszczenie. Artykuł dotyczy zastosowania analizy sieci społecznych do identyfikacji grup ściśle związanych blogów oraz do analizy charakterystyk tych grup. Analiza grup polega na wyliczeniu wartości miar opisujących gęstość grup oraz ich stabilność oraz na dyskusji uzyskanych wyników.

Słowa kluczowe: analiza sieci społecznych, miary centralności, grupy, blogi

ANALYSIS GROUPS IN SOCIAL MEDIA

Summary. In the paper, methods of identification groups in the blogosphere using the social network analysis approach are presented. Dynamics of groups are measured by density, stability and cohesion. Additional measure, light stability is introduced.

Keyword: social networks, centrality measures, communities, weblogs

1. Wprowadzenie

W ciągu ostatniej dekady można zaobserwować ogromne zmiany, jakie zaszły w formach aktywności w Internecie. Użytkownicy z biernych konsumentów informacji stali się również jej twórcami. Internetowe media społeczne mogą pojawiać się w różnych formach [9]:

- blogi (np. Blogspot², Blox³),
- fora (np. Yahoo!answers⁴),

¹ Prace badawcze prowadzące do opisanych wyników były częściowo finansowane z umowy grantowej nr 218086 Siódmego Programu Ramowego Unii Europejskiej

² <http://www.blogspot.com>

³ <http://www.blox.pl>

⁴ <http://answers.yahoo.com>

- współdzielenie mediów (np. Flickr⁵, YouTube⁶),
- mikroblogi (np. Twitter⁷, Blip⁸),
- portale społecznościowe (np. Facebook⁹, MySpace¹⁰, LinkedIn¹¹, Orkut¹²),
- społecznościowe oznaczanie (np. Del.icio.us¹³)
- oraz Wiki (np. Wikipedia¹⁴, Scholarpedia¹⁵).

Internetowe serwisy społecznościowe zrewolucjonizowały Internet, wielu uważa, że w niedalekiej przyszłości Internet będzie głównym lub wręcz jedynym medium informacyjnym świata.

2. Blogi i blogosfera

Wśród internetowych serwisów społecznościowych blogi odgrywają szczególną rolę w kształtowaniu opinii. Są one pewnego rodzaju dziennikiem internetowym, na których autor wyraża swoją opinię na jakiś temat lub opisuje ciekawe zdarzenie, a czytelnicy komentują te wpisy. Obecnie blogi prowadzą popularne osoby ze świata sportu, show biznesu czy polityki, wiele firm posiada własne blogi, coraz częściej blogi prowadzą naukowcy, gdzie na bieżąco prezentują swoje badania czy przemyślenia.

Typowy wpis na blogu może zawierać tekst, zdjęcia, filmy oraz linki do innych blogów bądź stron internetowych. Wpisy na blogu można kategoryzować, przypisując im tagi, czyli krótkie hasła opisujące zawartość wpisu. Bardzo ważnym elementem blogów jest możliwość dodawania komentarzy, sprzyja to rozwojowi dyskusji i pozwala w łatwy sposób odnieść się to przeczytanej notki. Dostęp do blogów jest najczęściej otwarty, czyli każdy może przeczytać wpis autora oraz komentarze innych użytkowników.

Blogi najczęściej umieszczane są na specjalnych serwisach, ułatwiających pisanie i publikowanie postów. Serwisy te mogą być przeznaczone dla blogów o dowolnej tematyce (np. Blogger¹⁶) bądź dla autorów interesujących się konkretną dziedziną, np. polityką – Salon24¹⁷. Większość serwisów do stworzenia bloga wymaga założenia konta, natomiast wy-

⁵ <http://www.flickr.com>

⁶ <http://www.youtube.com>

⁷ <http://twitter.com/>

⁸ <http://blip.pl>

⁹ <http://www.facebook.com>

¹⁰ <http://www.myspace.com>

¹¹ <http://www.linkedin.com>

¹² <http://www.orkut.com>

¹³ <http://www.delicious.com>

¹⁴ <http://pl.wikipedia.org>

¹⁵ <http://www.scholarpedia.org>

¹⁶ <https://www.blogger.com/start>

¹⁷ <http://www.salon24.pl>

stawianie komentarzy najczęściej nie wymaga logowania, chociaż istnieje możliwość podpisania komentarza np. *nickiem*.

Podstawowymi interakcjami zachodzącymi pomiędzy blogerami jest udzielanie komentarzy. Komentarz można wystawić w odniesieniu do posta bądź innego komentarza. Można także wyrazić fakt czytania innego bloga, umieszczając jego link na swoim blogu.

Ogół wszystkich blogów określa się mianem blogosfery (ang. *blogosphere*). Blogosfera charakteryzuje się bardzo dużą dynamiką, codziennie powstają tysiące nowych postów i miliony komentarzy¹⁸. Dlatego bardzo dynamiczne są też powiązania pomiędzy blogerami. Osoby komentujące dany blog często się zmieniają. Ogromna przestrzeń blogosfery stanowi źródło wielu informacji i poddawana jest różnorodnym badaniom i analizom [1].

3. Analiza sieci społecznych

Obecnie pomiędzy ludźmi na całym świecie mogą zachodzić różnego rodzaju interakcje. Są to m.in.: rozmowy telefoniczne, wymieniane maile, rozmowy za pomocą komunikatorów internetowych, komentarze na blogach. Takie interakcje mogą być modelowane za pomocą tzw. sieci powiązań. Sieć ta jest najczęściej grafem skierowanym, w którym wierzchołkami są osoby, natomiast krawędziami interakcje zachodzące pomiędzy nimi.

Dziedziną zajmującą się analizą wszelkiego rodzaju zależności występujących w takich sieciach jest analiza sieci społecznych (ang. *Social Network Analysis - SNA*).

Początki SNA sięgają XIX wieku, a obecnie – ze względu na bardzo łatwą dostępność danych sieci społecznych, m.in. z portali internetowych – metody SNA są powszechnie wykorzystywane w wielu dziedzinach. Metody analizy sieci społecznych pozwalają badać pozycję jednostek w strukturze, wyodrębniać grupy w ramach sieci czy badać globalne właściwości struktury całej sieci. SNA wykorzystywana jest w wielu dziedzinach, głównie tam, gdzie istnieją przepływy danych bądź zachodzą interakcje pomiędzy ludźmi.

Podstawowym parametrem SNA jest centralność (ang. *centrality*). Opisuje ona pozycję węzła w strukturze sieci stanowiąc miarę ważności, znaczenia i jego wpływu na inne obiekty w sieci [13]. Centralność można liczyć na wiele różnych sposobów. W zależności od stosowanego algorytmu zmienia się interpretacja uzyskanych wartości. Oprócz centralności wykorzystuje się inne miary, m.in. ranking Page'a. Miary te, jak również ich interpretacje, w kontekście analizy blogów, autorzy omówili w [15, 16]. W pracach tych przedstawiono metodę znajdowania ważnych blogów na podstawie analizy miar SNA oraz przeprowadzono analizę ewolucji zmian w czasie ważnych blogów. Analizując treści postów oraz komentarze opracowano metodę znajdowania wpływowych blogerów [16]. Wokół takich blogerów wykształcają się grupy,

¹⁸ <http://www.sifry.com/alerts/archives/000443.html>

które tworzą autorzy wzajemnie czytający i komentujący swoje blogi. Struktura takich grup może ujawnić użyteczne informacje o sieci: grupę można zastąpić metawęzłem i analizować sieć na wyższym poziomie abstrakcji, ujawniając tzw. słabe więzi (ang. *weak ties*), które zgodnie z postulatem Granovettera [5] mają ogromne znaczenie w rozwoju społeczności i odgrywają znaczącą rolę w dostarczaniu nowych informacji. Obserwowanie ewolucji grupy w czasie pozwala wyciągnąć interesujące wnioski odnośnie do kondycji grup i stopnia ich odporności na warunki zewnętrzne. Analiza grup jest przedmiotem niniejszego artykułu.

4. Definicja i właściwości grup

Trudno jest w literaturze znaleźć jednoznaczną definicję grupy [8]. Wasserman w [12] zauważa, że pojęcie grupy (podgrupy, kliki itd.) szeroko wykorzystywane jest w naukach socjologicznych bez podawania jej formalnej definicji. Grupy traktowane są jako podstawowy proces socjalnych wzajemnych oddziaływań [12]. Obecnie socjologowie coraz częściej uważają, że wyjaśnienia przyczyn różnych socjologicznych zjawisk sensowniej jest szukać w analizie zachowań grup, a nie jednostek.

Grupa jednostkom potrzebna jest do realizacji celów, których nie byłaby w stanie osiągnąć samodzielnie, jak również do określenia swojej tożsamości. Grupa istnieje tak długo, jak długo jednostki ją tworzące są zainteresowane, aby być jej członkami [3]. Ostatnio w literaturze dużo uwagi poświęca się identyfikacji oraz zrozumieniu grup i społeczności w sieciach społecznych. Zazwyczaj terminy: społeczność, klaster, moduł czy spójny podgraf używane są zamiennie w zależności od kontekstu, czynią tak np. Tang i Liu w [9]. W [3] dokonano rozróżnienia między grupą i społecznością. Grupa jest podzbiorem jednostek w populacji, z których każda identyfikowana jest jako należąca do określonej organizacji, podczas gdy społeczność, to podzbiór jednostek w sieciach społecznych, które są dużo gęściej (silniej) połączone między sobą niż z pozostałymi jednostkami sieci.

Obecnie duża część badań skoncentrowana jest na znajdowaniu grup w dużych zbiorach danych, a dużo mniej dotyczy analiz przyczyn powstawania grup oraz zrozumienia ich ewolucji. Zrozumienie takich procesów mogłoby pozwolić przewidzieć perspektywę sukcesu lub porażki utworzenia grupy np. z określonym programem politycznym, określić stopień przystosowania się organizacji np. do zmieniających się warunków środowiskowych.

Socjologowie rozróżniają grupy pierwotne (ang. *primary*) i wtórne (ang. *secondary*). Pierwsze z nich to małe grupy, w których wszyscy członkowie mają ze sobą bezpośredni kontakt (np. rodzina, zespół, przyjaciele), podczas gdy drugie są większe, mogą być rozproszone geograficznie i mieć dużo bardziej złożoną wewnętrzną strukturę. Przynależność do grup pierwszego rodzaju jest zazwyczaj automatyczna i jednostka ma niewielki wybór. Przystąpienie do grup

drugiego rodzaju w pełni zależy do jednostki, tak samo jak decyzja o jej opuszczeniu (można tu mówić o niewielkim wpływie czynników zewnętrznych). Członkowie takich grup są zazwyczaj połączeni wspólnymi zainteresowaniami lub celami. Jeżeli cel zostanie osiągnięty, grupa może trwać lub się rozpaść. Zrozumienie przyczyn, które wpływają na decyzję jednostki o dołączeniu się do grupy albo odłączeniu jest tematem wielu prac badawczych.

Znajdowanie grup w sieciach społecznych polega na identyfikacji zbioru węzłów, takich które komunikują się między sobą dużo częściej niż z węzłami spoza grupy [4,9]. Tak więc grupy można traktować jako zwarte podzbiory węzłów sieci, które są z kolei luźno połączone z węzłami spoza grupy. Jednostki częściej kontaktują się z członkami grupy niż z tymi, którzy do grupy nie należą. W praktyce, w złożonych sieciach społecznych grupy nie są odizolowane między sobą, jednostka w danym czasie może być członkiem wielu grup.

5. Metody wyodrębniania grup w sieciach społecznych

Różnorodność definicji grupy i metod ich znajdowania być może jest odzwierciedleniem braku obiektywnej, ogólnie przyjętej definicji struktury społecznościowej w rzeczywistych sieciach [10]. Metody detekcji grup opracowywane są dla konkretnych zastosowań. W [2] przedstawiono szczegółowy opis tych metod.

Zaproponowano wiele metod znajdowania spójnych grup. W [9, 11] przedstawiono interesujące podejście do ich usystematyzowania w cztery kategorie na podstawie różnych kryteriów: węzłowych, grupowych, sieciowych i hierarchicznych. Metody oparte na kryterium węzłowym wymagają, aby każdy węzeł grupy spełniał określone właściwości. Kryterium grupowe traktuje powiązania wewnątrz grupy jako całość. Dopuszczalna jest sytuacja, w której pewne węzły wewnątrz grupy są słabo połączone, jeśli tylko cała grupa spełnia określone wymagania. Wykrywanie grup przy użyciu metod opartych na kryterium sieciowym bierze pod uwagę globalną topologię sieci jako całości. Węzły sieci dzielone są na pewną liczbę rozłącznych zbiorów. W ostatniej grupie metod (hierarchicznych), do wykrywania grup budowana jest hierarchiczna struktura grupowa, opierająca się na topologii sieci.

6. Miary opisujące grupę

Do opisu grup można wykorzystać kilka miar wyrażających ich najważniejsze cechy, takie jak: gęstość, stabilność i spójność.

Gęstość (ang. *density*) – liczona, jako stosunek liczby powiązań wewnątrz grupy do maksymalnej, możliwej liczby powiązań [12].

$$Density = \frac{2 \sum_i^n \sum_{i < j}^n a(i, j)}{n(n-1)},$$

gdzie:

n – liczba wierzchołków,

$a(i, j) = 1$, gdy istnieje krawędź od i do j , w przeciwnym wypadku 0.

W grupach o dużej gęstości jej członkowie dobrze się znają, każda osoba jest powiązana z wieloma innymi, w ramach grupy.

Inna miara – stabilność (ang. *stability*) – opisuje stałość grupy w czasie, informuje, czy zbiór członków grupy ulega zmianom. Liczona jest jako stosunek liczby osób zawsze obecnych w grupie do liczby wszystkich członków grupy [14]. Mając dane grupy G w punktach czasowych t_1 i t_2 stabilność określona jest wzorem:

$$Stability = \frac{|G_1 \cap G_2|}{|G_1 \cup G_2|},$$

gdzie:

$G_1 \cap G_2$ – osoby należące do grup G_1 i G_2 ,

$G_1 \cup G_2$ – osoby należące do grupy G_1 lub G_2 .

Wysoka stabilność określa grupę trwałą, stałą, w której występują zawsze te same osoby. W grupach o niskiej stabilności często pojawiają się nowe osoby, a znikają „stare”.

Spójność (ang. *cohesion*) – bierze pod uwagę siłę powiązań pomiędzy członkami grupy. Liczona jako stosunek średniej siły powiązań między członkami grupy do średniej siły ich powiązań z osobami spoza grupy [12]:

$$Cohesion = \frac{\frac{\sum_{i \in G} \sum_{j \in G} w(i, j)}{n(n-1)}}{\frac{\sum_{i \in G} \sum_{j \notin G} w(i, j)}{n(N-n)}},$$

gdzie:

$w(i, j)$ – waga powiązania pomiędzy węzłami i oraz j ,

N – węzły należące do całej sieci,

n – węzły należące do grupy.

Grupy o dużej spójności charakteryzują się mocnymi powiązaniem pomiędzy ich członkami.

7. Analiza społeczności w blogosferze

Wśród blogerów mogą utworzyć się różnorodne społeczności, czyli grupy osób w jakiś sposób powiązanych ze sobą. Tymi powiązaniem mogą być np.: udzielone komentarze, linki do innych blogów czy też podobne poglądy lub zainteresowania. Istnienie grup zwiększa

aktywność użytkowników oraz wspomaga tzw. propagację zaufania. Liczebność grup może być bardzo zróżnicowana (od kilku do kilkudziesięciu osób) i zwykle wzrasta wraz z osłabieniem powiązań pomiędzy ich członkami.

7.1. Wyodrębnianie społeczności

W naszych badaniach grupy z poszczególnych procesów wyodrębniane były programem CFinder¹⁹, który jest implementacją algorytmu CPM (Clique Percolation Method) [6, 7]. Algorytm ten wyszukuje grupy w grafie *k-klik* (graf pełny o *k* wierzchołkach, w którym z każdego wierzchołka można bezpośrednio dojść do dowolnego innego), traktowane jako zbiory przylegające, czyli posiadające *k*-1 wspólnych wierzchołków, *k-klik* (ang. *adjacent k-cliques*). Zgodnie z tą definicją jeden wierzchołek może należeć do wielu grup, co dobrze odzwierciedla rzeczywistą sytuację, gdy każdy z blogerów może należeć do wielu grup.

Program CFinder na wejściu przyjmuje plik z listą krawędzi, a w wyniku działania dostaje się listę grup dla określonego parametru *k*. Ze względu na potrzebę analizy grup w czasie, CFinder uruchamiany był wielokrotnie z danymi z kolejnych okresów.

Aby dokonać bardziej wnikliwej analizy społeczności, podczas wyszukiwania grup należy nałożyć dodatkowe warunki. Grupy utworzone za pomocą algorytmu CPM musiały istnieć przez co najmniej 3 okresy (np. 3 miesiące, gdy okresem był 1 miesiąc). Dodatkowo, wymagane jest, aby w kolejnym okresie w grupie pozostawało 50% jej członków. Warunek ten może być rozumiany na 2 sposoby:

- 1) osoby z kolejnego okresu muszą stanowić przynajmniej 50% osób już należących do grupy,
- 2) osoby z kolejnego okresu muszą stanowić przynajmniej 50% osób z poprzedniego okresu.

Do dalszych analiz wykorzystano warunek 2.

Nie powinny być analizowane grupy większe niż 100-osobowe, gdyż takie tworzyły się dla słabo powiązanych użytkowników (dla parametrów $k=3$, dla algorytmu CPM). Dlatego też w systemie istnieje warunek eliminujący duże grupy z dalszego przetwarzania.

W tak powstałych społecznościach można wyliczyć wartość parametrów gęstości oraz stabilności (omówione w rozdziale 6.). Dodatkowo, do oceny grupy wprowadzony został mniej restrykcyjny warunek w stosunku do stabilności: liczymy stosunek liczby osób obecnych przynajmniej w dwóch okresach do liczby wszystkich osób w grupie. Parametr ten określony został jako *lekka stabilność*.

7.2. Opis przeprowadzonych eksperymentów

Eksperymenty przeprowadzono na podstawie sieci zbudowaną na danych portalu salon24.pl, zebranych w latach 2006-2010, w której było 2 825 blogów, prowadzonych przez

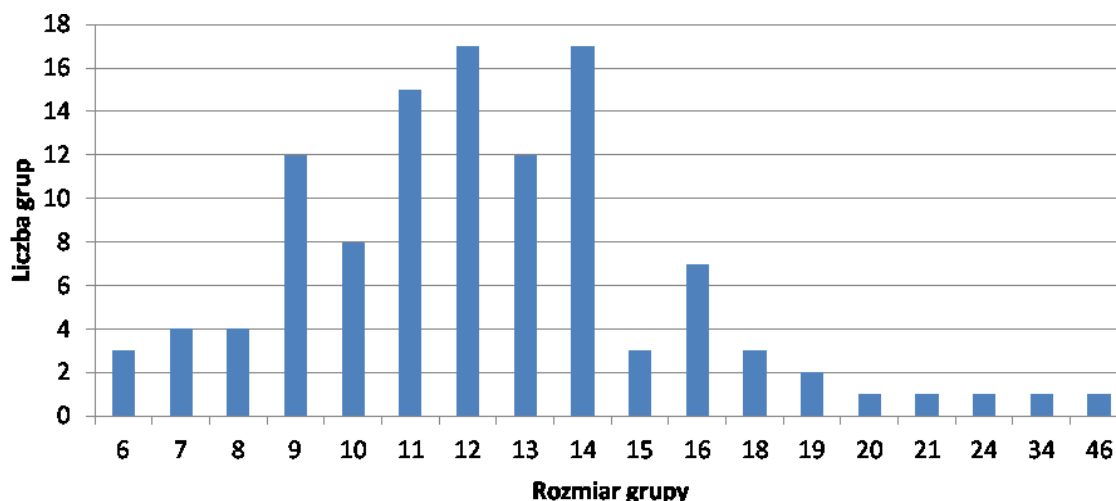
¹⁹ <http://cfinder.org>

7 386 autorów, którzy napisali 16 756 postów, które z kolei dostały 368 084 komentarzy. Najwięcej danych pochodziło z 2009 roku.

Następnie na zebranych danych przeprowadzono wiele analiz, mających na celu wyszukanie wpływowych autorów, analizę dynamiki sieci oraz analizę grup. Pierwszym etapem było wyliczenie miar SNA dla wszystkich autorów i stworzenie na tej podstawie rankingu, mającego na celu wskazanie najważniejszych blogerów. Następnym krokiem była analiza dynamiki sieci, poprzez śledzenie zmian miar SNA dla poszczególnych autorów, szczególnie tych, uznanych za wpływowych dzięki wysokim wartościom miar SNA. W kolejnym etapie spróbowano wykorzystać analizę treści, w celu wykrycia zjawiska wpływu niektórych postów na powstanie kolejnych tekstów, czyli odnalezienie postów zainspirowanych innymi. Do tych analiz również wykorzystano autorów, wyszukanych w poprzednich analizach. Ostatnim zadaniem było wyszukanie grup w sieci oraz analiza ich zmienności w czasie.

Do analizy grup wykorzystano dane z lat 2007-2009, gdyż są one najbardziej kompletne. Przedział ten podzielono na półroczne okresy. Udało się wyodrębnić kilka interesujących grup, które ulegały podziałom, a następnie powtórnym połączeniom. Niewielka grupa użytkowników była obecna zawsze w każdej z faz. Ciekawym trendem jest fakt, że z każdym kolejnym miesiącem społeczność się powiększała i wielu użytkowników pozostawało w grupie.

Kryteria przedstawione w rozdziale 7.1 pozwoliły na znalezienie 112 grup; liczyły one od 6 do 46 osób. Rozkład liczości grup przedstawiono poniżej (rys. 1).

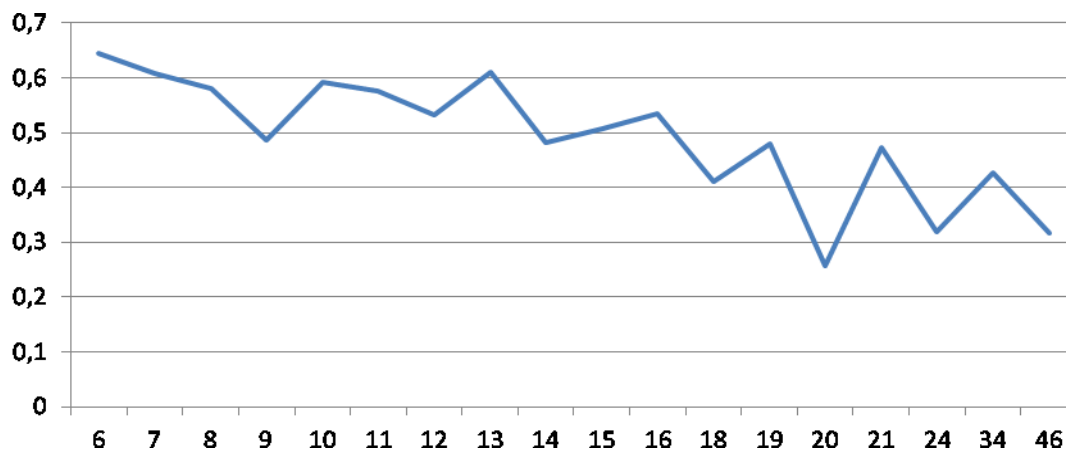


Rys. 1. Zależność liczby grup od ich wielkości

Fig. 1. Dependence between number of groups and their size

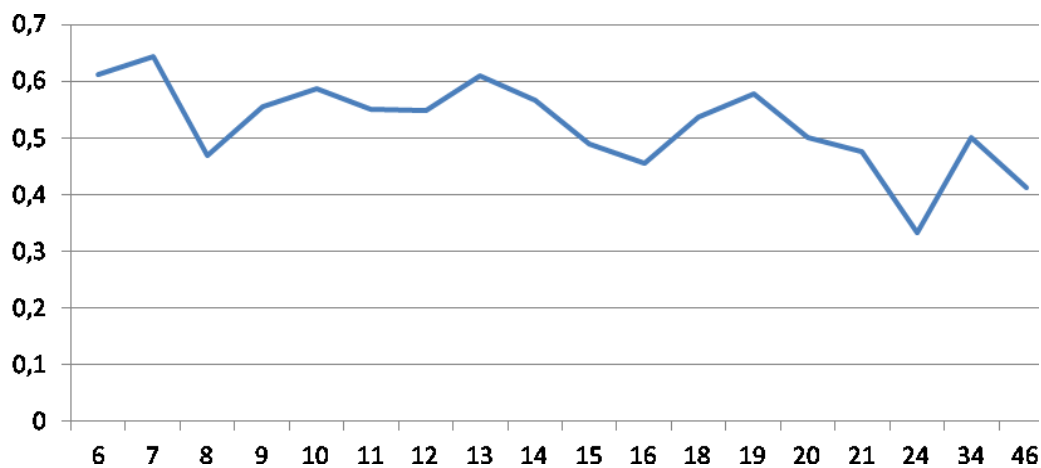
Z powyższego wykresu można wywnioskować, że najczęściej występują średnie grupy (9-14 osób).

Dla wszystkich grup wyliczono 3 parametry: gęstość, stabilność oraz lekką stabilność i zobrazowano je na poniższych wykresach (przedstawiono średnią wartość dla danego rozmiaru grupy).



Rys. 2. Średnie wartości gęstości dla grup o poszczególnych rozmiarach
Fig. 2. Average density vs groups size

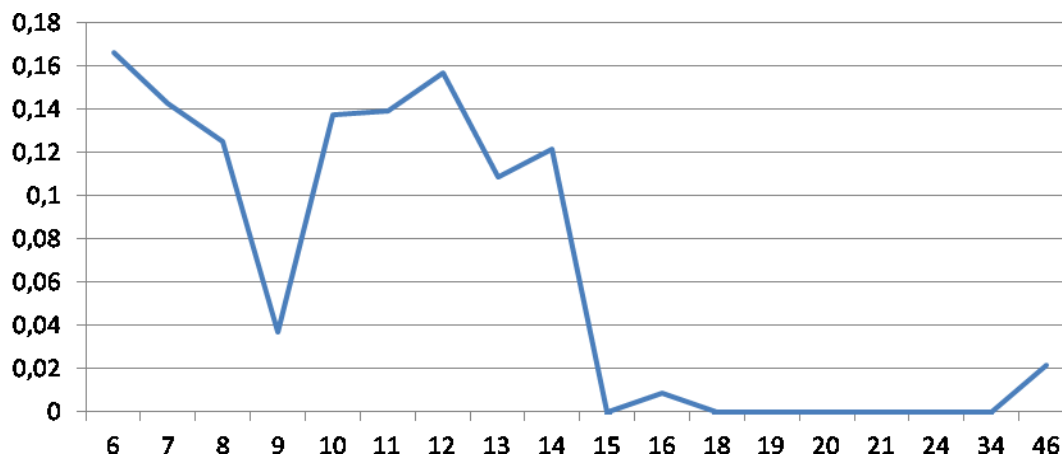
Gęstość również maleje liniowo wraz ze wzrostem liczności grup (rys. 2). Jest to zrozumiałe, gdyż w małych grupach łatwiej nawiązać bezpośredni kontakt z wszystkimi członkami. Największą gęstość – 0,82 osiągnęła grupa 10-osobowa, czyli na 45 możliwych powiązań, w grupie istnieją 37.



Rys. 3. Średnie wartości lekkiej stabilności dla grup o poszczególnych rozmiarach
Fig. 3. Average value of light stability vs group size

Na wykresie (rys. 3) przedstawiono średnie wartości dla grup o danej liczności. Średnie te są z przedziału od 0,64 dla grup 7-osobowych do 0,33 dla grupy 24-osobowej. Największą wartość – 0,9 osiągnęła grupa 12-osobowa, najmniejszą – 0,3 grupa 10-osobowa. Generalnie można zauważyć, że lekka stabilność maleje liniowo wraz z rozmiarem grupy. Lekka stabilność równa 0,33 oznacza, że 33% członków grupy była obecna przez co najmniej 2 okresy.

Stabilność bardzo szybko maleje wraz ze wzrostem liczości grupy (rys. 4). Dla grup większych niż 15 osób jest równa 0. Największą wartość – 0,375 osiągnęła grupa 8-osobowa. Oznacza to, że 37% członków tej grupy było obecnych w każdym okresie trwania grupy.



Rys. 4. Średnie wartości stabilności dla grup o poszczególnych rozmiarach
Fig. 4. Average value of stability vs group size

8. Wnioski

W artykule zostały przedstawione analizy charakterystyk grup formujących się dla blogów udostępnianych na portalu salon24.pl. Zaprezentowane zostały średnie wartości gęstości i dwóch rodzajów stabilności dla grup o poszczególnych rozmiarach. Można zauważyć, że wartości gęstości oraz stabilności maleją wraz ze wzrostem rozmiarów grup, co potwierdza intuicyjne przypuszczenia.

Dalsze prace koncentrują się na użyciu innych algorytmów identyfikacji grup oraz na analizie przyczyn powstawania grup oraz skłaniających poszczególnych użytkowników do przyłączania się do nich. Przewiduje się również dokładniejszą analizę struktury grup, poprzez wyodrębnienie ról pełnionych przez poszczególnych jej członków (rdzeń grupy, aktywni członkowie oraz osoby luźno z grupą związane).

BIBLIOGRAFIA

1. Agarwal N., Liu H.: Modeling and Data Mining in Blogosphere. Morgan & Claypool, 2009.
2. Fortunato S.: Community detection in graphs. Physics Reports 486, 2010.

3. Geard N., Bullock S.: Group formation and social evolution: a computational model, Proc. of the Eleventh International Conference on the Simulation and Synthesis of Living Systems, MIT Press, USA 2008.
4. Girvan M., Newman M. E. J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci., USA 2002.
5. Granovetter M. S.: The Strength of Weak Ties. American Journal of Sociology 78 (6), 1973.
6. Palla G., Abel D., Derényi I., Farkas I., Pollner P., Vicsek T.: K-clique percolation and clustering in directed and weighted networks. Bolayai Society Mathematical Studies, 2005.
7. Palla G., Derényi I., Farkas I., Vicsek T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 2005, s. 814÷818.
8. Palla G., Pollner P., Barabasi A.-L., Vicsek T.: Social group dynamics in networks. Gross T., Sayama H. (ed.): Adaptive networks. Springer Berlin/Heidelberg 2009.
9. Tang L., Liu H.: Community detection and mining in social media. Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan&Claypool, 2010.
10. Tang L., Liu H.: Graph mining applications to social network analysis, [in:] Aggarwal C., Wang X. (ed.): Managing and Mining Graph Data, 2010.
11. Tang L.: Learning with large-scale social media network. PhD thesis, Arizona State University, 2010.
12. Wasserman S., Faust K.: Social Network Analysis: Methods and applications. Cambridge University Press, 1994.
13. White S., Smyth P.: Algorithms for estimating relative importance in networks, Proc. of ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.
14. Xu J., Marshall B., Kaza S., Chen H.: Analyzing and visualizing criminal network dynamics: a case study, ISI, 2004.
15. Zygmunt A., Koźlak J., Krupczak Ł., Małocha B.: Analiza blogów internetowych przy użyciu metod sieci społecznych. Automatyka: półrocznik Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, 13(2), 2009.
16. Zygmunt A., Koźlak J., Krupczak Ł.: Identyfikacja wpływowych jednostek w blogosferze. Studia Informatica, Vol. 31, No. 2A (89), Wyd. Pol. Śląskiej, Gliwice 2010.

Recenzenci: Dr inż. Michał Kozielski

Dr hab. inż. Tadeusz Pankowski, prof. Uniwersytetu im. Adama Mickiewicza

Wpłynęło do Redakcji 31 stycznia 2011 r.

Abstract

In this paper we describe methods of finding groups in social networks. Understanding process of group formation and reasons of their creation can be important for example in marketing, politics. There are many methods of extraction groups from social network. In our work we analyze blogosphere and groups which form around influential bloggers. We choose Clique Percolation Method (CPM) algorithm, based on identification of k -cliques. This algorithm allows finding overlapping groups which is natural in blogosphere: every bloggers can belong to many groups. We try to find stable groups lasted for some period of time.

Adresy

Anna ZYGMUNT: Akademia Górniczo-Hutnicza w Krakowie, Katedra Informatyki,
Al. Mickiewicza 30, 30-059 Kraków, Polska, azygmunt@agh.edu.pl.

Jarosław KOŹLAK: Akademia Górniczo-Hutnicza w Krakowie, Katedra Informatyki,
Al. Mickiewicza 30, 30-059 Kraków, Polska, kozlak@agh.edu.pl.

Łukasz KRUPCZAK: Akademia Górniczo-Hutnicza w Krakowie, Al. Mickiewicza 30,
30-059 Kraków, Polska, lkrupczak@gmail.com.