

**Krzysztof SZKÓŁKA, Cezary BARTOSIAK,
Rafał KASPRZYK, Zbigniew TARAPATA**

Wojskowa Akademia Techniczna, Wydział Cybernetyki
ul. Kaliskiego 2, 00-908 Warszawa

E-mail: kszkolka@wat.edu.pl, cbartosiak@wat.edu.pl, rkasprzyk@wat.edu.pl,
ztarapata@wat.edu.pl

Modele i metody komputerowego rozpoznawania wzorców opisanych ilościowo

1 Wprowadzenie

Badanie podobieństwa grafów i sieci jest dziedziną istotną w wielu obszarach nauki: chemii, biologii, energetyce czy informatyce. Większość systemów można stosunkowo łatwo i intuicyjnie modelować jako graf lub sieć, a następnie tak zbudowany model wykorzystać do badania, jak również optymalizacji tych systemów [2, 3, 4, 5].

Według Słownika Języka Polskiego „*podobieństwo*” to wspólność pewnych cech dwu albo więcej osób, przedmiotów, obiektów itp. W kontekście analizy podobieństwa grafów i sieci można stosować tę definicję intuicyjnie, bez potrzeby jej formalizacji. W literaturze spotyka się liczne taksonomie metod badania podobieństwa grafów i sieci [14, 15]. Najbardziej ogólny podział rozróżnia właśnie metody grafowe i sieciowe. Metody grafowe wykorzystuje się do badania podobieństwa *strukturalnego*, a sieciowe do badania podobieństwa *niestrukturalnego* (nazywanego inaczej *sieciowym* lub *ilościowym*).

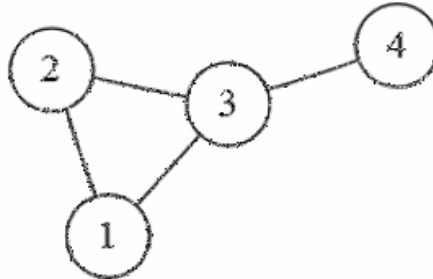
W literaturze można spotkać wiele przykładów metod i miar, które rozwiązują problem strukturalnego podobieństwa grafów. Do klasycznych metod można zaliczyć: izomorfizm oraz odległość edycyjną. W przypadku metod i miar ilościowych rozwiązań jest mało, niewiele z nich jest dobrze udokumentowanych i są to najczęściej metody rozwiązujące konkretny problem z danej dziedziny.

W artykule przedstawiona jest autorska metoda analizy podobieństwa grafów i sieci. Metoda ta została zaimplementowana i jest jedną z funkcjonalności w popularnej aplikacji komercyjnej, oferującej między innymi audyt użytkowników.

2 Definicje i notacje

Graf jest abstrakcyjną reprezentacją zbioru obiektów, między którymi mogą istnieć związki. Obiekty te prezentowane są jako *wierzchołki*, związki zaś jako *gałęzie* (można mówić o *krawędziach*, *łukach* oraz *pętlach*). Przykładowy graf pokazano na rysunku 1.

Formalnie graf G to trójka uporządkowana, $G = \langle V, E, P \rangle$, gdzie V to zbiór wierzchołków grafu G , E to zbiór gałęzi grafu G , a $P \subset \langle V \times E \times V \rangle$ – relacja incydencji [1].



Rys. 1. Graf o 4 wierzchołkach i 4 krawędziach [opracowanie własne]

Fig. 1. Graph with 4 nodes and 4 edges [own preparation]

Siecią nazywamy graf, który jest opisany ilościowo (wierzchołki i gałęzie mogą mieć atrybuty, które je dodatkowo opisują). Formalnie sieć S to trójka uporządkowana:

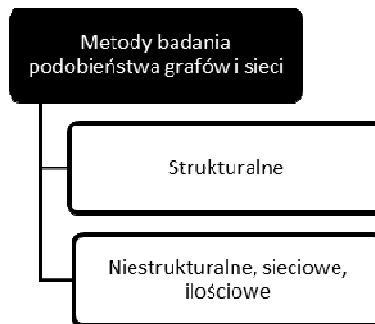
$$S = \left(G, \{f_i(v)\}_{i \in \{1, \dots, LF\}, v \in V}, \{h_j(e)\}_{j \in \{1, \dots, LH\}, e \in E} \right),$$

gdzie G to graf, $f_i(v): V \rightarrow R^n$, to i -ta funkcja opisana na wierzchołku grafu G ($i = 1, \dots, LF$; LF – liczba wszystkich funkcji opisanych na wierzchołkach grafu), $h_j(e): E \rightarrow R^n$ to j -ta funkcja opisana na gałęzi grafu G ($j = 1, \dots, LH$; LH – liczba wszystkich funkcji opisanych na gałęziach grafu). Warto zauważyć, że każdy graf jest siecią, w której $f_i = h_j = \emptyset$.

Przykładem sieci może być więc graf z ponumerowanymi wierzchołkami i gałęziami (na zbiorze wierzchołków i gałęzi są określone funkcje, których wartościami są numery).

3 Metody badania podobieństwa grafów i sieci

Metody badania podobieństwa można podzielić jak na rysunku 2.



Rys. 2. Taksonomia miar podobieństwa [opracowanie własne]

Fig. 2. Taxonomy of similarity measures [own preparation]

Miary strukturalne (ang. *structural similarity measures*) stosuje się zwykle do badania grafów, a ilościowe (ang. *quantitative similarity measures*) do badania sieci. W tych

pierwszych porównywana jest jedynie topologia (struktura) systemu, natomiast miary ilościowe uwzględniają cechy opisane na wierzchołkach i gałęziach grafu. Do najważniejszych metod i miar badania podobieństwa strukturalnego można zaliczyć izomorfizm grafów, miary oparte na odległości edycyjnej, miary topologiczne i metody iteracyjne - strukturalne. Metody badania podobieństwa niestukturalnego opierają się głównie na metodach iteracyjnych – ilościowych. Dalej przedstawione są wybrane metody i miary badania podobieństwa grafów i sieci.

3.1 Izomorfizm

Jeśli pomiędzy dwoma grafami występuje izomorfizm, to są one strukturalnie nieodróżnialne. Formalnie grafy G i H nazywa się izomorficznymi, jeżeli istnieje bijekcja pomiędzy zbiorami ich wierzchołków:

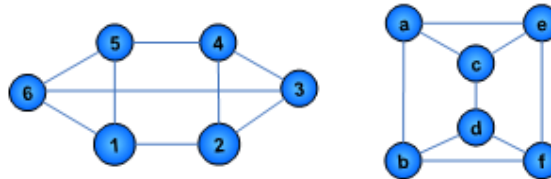
$$f: V_G \rightarrow V_H,$$

gdzie:

V_G – zbiór wierzchołków grafu G ,

V_H – zbiór wierzchołków grafu H ,

taka że dwa dowolne wierzchołki u i v w grafie G są do siebie przyległe wtedy i tylko wtedy, gdy $f(u)$ i $f(v)$ są przyległe w grafie H . Można również rozważać izomorfizm sieci, w takim przypadku muszą zostać zachowane atrybuty (wartości funkcji) wierzchołków i gałęzi. Przykładowe grafy izomorficzne przedstawione są na rysunku 3.



Rys. 3. Grafy izomorficzne [opracowanie własne]

Fig. 3. Isomorphic graphs [own preparation]

Problem rozstrzygnięcia izomorfizmu grafów [9] należy do klasy NP [6], dla których nie istnieje obecnie odpowiedź, czy należą one do podklasy problemów P czy NP-zupełnych.

Miarami izomorficznymi nazywana jest grupa miar badania podobieństwa grafów, które korzystają z izomorfizmu do wyznaczenia wartości tych miar.

Wśród metod badania izomorfizmu grafów za najlepsze uznawane są obecnie metoda E. M. Luksa [7] o złożoności czasowej $exp(O(n \log n))$ oraz metoda B. D. McKaya [8] (złożoność czasową trudno jest oszacować, wiadomo, że jest ona wykładnicza). Warto zaznaczyć, iż wiele metod jest ograniczonych do określonych klas grafów.

3.2 Odległość edycyjna

Odległość edycyjna między dwoma grafami określa minimalną liczbę koniecznych operacji, aby jeden graf „zamienić” w drugi (operacją może być dodawanie lub usuwanie wierzchołków lub gałęzi). Można ją rozumieć jako uogólnienie izomorfizmu. Odległość edycyjna pomiędzy dwoma grafami izomorficznymi wynosi 0.

Zazwyczaj różne modyfikacje związane są z różnymi kosztami. Koszt modyfikacji mówi o tym, jak „trudno” jest ją przeprowadzić.

Do popularnych metod wyznaczania odległości edycyjnej można zaliczyć poszukiwanie największego podgrafu badanych grafów lub najmniejszego wspólnego nadgrafu, czy przedstawioną dalej metodę Papadopoulou i Manolopoulou. Autorzy tej metody zaproponowali w swojej pracy [10] algorytm, który opiera się na dodawaniu, usuwaniu oraz modyfikacji wierzchołków. Autorzy wprowadzili ponadto koncepcję tzw. histogramu grafu. Histogram grafu G budowany jest poprzez wyliczenie stopnia każdego wierzchołka. Następnie każdy wierzchołek zaliczany jest do innego „pojemnika” histogramu (którego wartość jest równa stopniowi odpowiadającego mu wierzchołka). W celu porównania dwóch grafów buduje się dla nich histogramy (ewentualnie dodając „pojemniki” o zerowej wartości, jeśli między grafami występuje różnica w liczbie wierzchołków), następnie szereguje się je w porządku malejącym (według wartości) oraz zwiększa wartości każdego „pojemnika” o jeden. Suma bezwzględnych różnic wartości między odpowiadającymi sobie „pojemnikami” stanowi szacunkową liczbę modyfikacji.

Algorytm ten jest bardzo szybki (złożoność czasowa $O(n \log n)$), lecz nie jest dokładny. Ponadto nadaje się tylko do badania podobieństwa strukturalnego. Mimo to dzięki szybkości może być z powodzeniem używany do wstępnego badania podobieństwa grafów.

3.3 Miary centralności

W przypadku bardzo dużych grafów i sieci może się okazać, że badanie izomorfizmu lub wyliczanie odległości edycyjnej nie będzie możliwe, głównie ze względu na złożoność obliczeniową tych metod, a w konsekwencji ich czasochłonność w praktycznych przypadkach. Wówczas do oceny podobieństwa grafów i sieci wykorzystać można miary odzwierciedlające ich właściwości. Przykładowymi charakterystykami (właściwościami) grafów i sieci są: średnia długość najkrótszej drogi, średnica (droga najkrótsza pomiędzy najbardziej oddalonymi wierzchołkami), współczynnik klasteryzacji, średni stopień wierzchołka, rozkład stopnia wierzchołka, czy w końcu miary centralności.

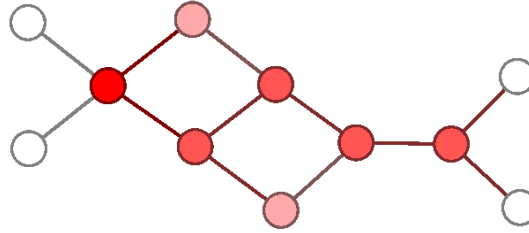
W różnego rodzaju analizach modeli grafowo-sieciowych coraz częściej wykorzystuje się miary centralności, w szczególności w przypadku analizy sieci społecznych i ich zastosowaniu, np. w epidemiologii [17, 18, 19] i analizie siatek przestępczych [20]. Miary centralności ułatwiają udzielenie odpowiedzi na pytanie, który wierzchołek jest najbardziej istotny w badanym grafie lub sieci. Stąd też miary centralności często nazywane są miarami istotności. Najczęściej wykorzystywanymi miarami centralności są: *stopień wierzchołka*, *promień wierzchołka*, *bliskość wierzchołka* i *obciążenie/pośrednictwo wierzchołka*. Każda miara centralności pozwala na ilościową ocenę istotności wierzchołków w grafach lub sieciach niejako z różnej perspektywy. Przedstawione dalej definicje miar centralności zostały opracowane dla tzw. *grafów zwykłych* [1]. Na potrzeby zobrazowania interpretacji poszczególnych miar centralności na rysunkach od 4 do 7 wierzchołki grafu przyjmują kolor o intensywności proporcjonalny do wartości rozpatrywanej miary centralności.

Najprostszą miarą centralności wierzchołka, którą podpowiada intuicja, jest **stopień wierzchołka** (ang. *degree centrality*). Według tej miary wierzchołek jest tym istotniejszy, im ma wyższy stopień:

$$dc_i = \frac{k_i}{|V|-1},$$

gdzie:

k_i - stopień i -tego wierzchołka.



Rys. 4. Zobrazowanie wartości miary stopień [opracowanie własne]

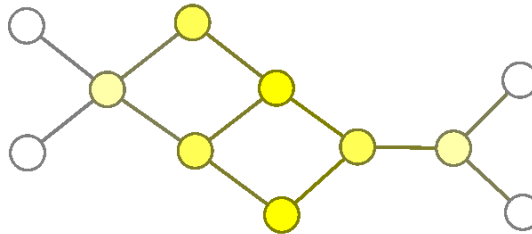
Fig. 4. Visualization of degree centrality [own preparation]

Kolejną miarą centralności jest tzw. **promień wierzchołka** (ang. *radius centrality*), według której wierzchołek jest tym ważniejszy, im jego odległość do najdalszego wierzchołka jest najmniejsza. Miara ta jest niezwykle interesująca z punktu widzenia zarządzania kryzysowego, kiedy istotne jest zapewnienie pewnego minimum w najgorszym możliwym do zaistnienia przypadku:

$$rc_i = \frac{1}{\max_{j \in V} d_{ij}}$$

gdzie:

d_{ij} – droga najkrótsza pomiędzy wierzchołkami i oraz j .

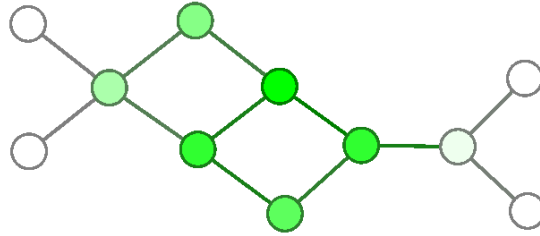


Rys. 5. Zobrazowanie wartości miary promień [opracowanie własne]

Fig. 5. Visualization of radius centrality [own preparation]

Miara **bliskość wierzchołka** (ang. *closeness centrality*) pozwala na wyliczenie średniej odległości rozpatrywanego wierzchołka do wszystkich pozostałych wierzchołów w grafie. Wierzchołek, który „średnio rzecz biorąc”, jest najbliższy wszystkim wierzchołkom jest uznawany za najistotniejszy:

$$cc_i = \frac{|V|-1}{\sum_{j \in V} d_{ij}}.$$



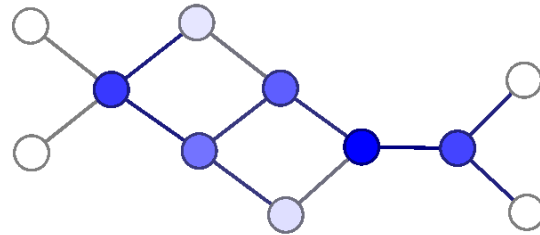
Rys. 6. Zobrazowanie wartości miary „bliskość” [opracowanie własne]
 Fig. 6. Visualization of closeness centrality [own preparation]

Innym sposobem oceny węzła jest wyznaczenie **pośrednictwa/obciążenia wierzchołka** (ang. *betweenness/load centrality*), które można zdefiniować jako procent najkrótszych dróg pomiędzy dowolną parą wierzchołków przechodzących przez rozpatrywany węzeł:

$$bc_i = \frac{\sum_{l \in V} \sum_{k \neq l \in V} \frac{p_{l,i,k}}{p_{l,k}}}{(|V|-2)(|V|-1)},$$

gdzie:

- $p_{l,k}$ – liczba dróg najkrótszych między wierzchołkami l oraz k ,
- $p_{l,i,k}$ – liczba dróg najkrótszych między wierzchołkami l oraz k przechodzących przez i .



Rys. 7. Zobrazowanie wartości miary pośrednictwo [opracowanie własne]
 Fig.7. Visualization of betweenness centrality [own preparation]

Ze względu na prosty sposób obliczania każdej z miar można je zastosować przy badaniu podobieństwa, np. badać podobieństwo rozkładu wybranej miary w jednym grafie do miary w innym grafie.

3.4 Miary iteracyjne

Istnieje szeroka gama metod badających podobieństwo grafów i sieci, które odnoszą się do pytania: *Jak bardzo podobne jest sąsiedztwo pewnego wierzchołka w jednym w grafie do sąsiedztwa innego wierzchołka w drugim grafie?* Metody te opierają się na idei, która mówi, że dwa wierzchołki dwóch różnych grafów są uważane za podobne, gdy ich sąsiedztwo jest do siebie podobne. Definicja ta sugeruje, że metody, które się do niej stosują, muszą mieć iteracyjny charakter, bowiem oceny podobieństwa elementów propagują się na sąsiedztwo tychże w każdym kroku.

Metod iteracyjnych jest bardzo dużo. Wykorzystują one różne techniki do oceny podobieństwa wierzchołków. Do najbardziej popularnych zalicza się na przykład *SimRank* [11] czy iteracyjne strukturalne podobieństwo wierzchołków [12]. Warto

również zwrócić uwagę na sieciową miarę podobieństwa wierzchołków [13, 16]. Autor proponuje miarę, która pozwala oszacować ilościowe podobieństwo wierzchołków sieci. Można także mierzyć ilościowe podobieństwo gałęzi (zamiast funkcji opisanych na wierzchołkach bierze się pod uwagę funkcje opisane na gałęziach) lub zastosować podejście mieszane.

4 Autorska metoda badania podobieństwa sieci

Niech dana będzie sieć: $S = \left(G = \langle V, E \rangle, \emptyset, \{h_k(e)\}_{k \in \{1, \dots, LH\}} \right)_{e \in E}$,

gdzie:

G - graf zwykły [1],

V - zbiór wierzchołków grafu G o mocy $|V|=n$,

$E \subset \{\{v, v'\}: v, v' \in V\}$ - zbiór krawędzi grafu G ,

$h_k: E \rightarrow R$ - k -ta funkcja określona na zbiorze E krawędzi grafu G przyjmująca wartości ze zbioru R (zbiór liczb rzeczywistych).

Opracowana metoda wyznacza podobieństwo każdej pary wierzchołków w sieci zdefiniowanej jak wyżej.

Krok 1. Wyznaczenie wektora macierzy podobieństw wierzchołków $Sim(S)$ ze względu na k -tą funkcję opisaną na krawędziach sieci S :

$$Sim(S) = \langle Sim_1, \dots, Sim_k, \dots, Sim_{LH} \rangle,$$

gdzie:

$$Sim_k = [s_{ij}(k)]_{n \times n}.$$

Wyliczenie każdego z elementów $s_{ij}(k)$ macierzy Sim_k bazuje na podgrafie¹ G'_{ij} , którego zbiorem wierzchołków jest zbiór $NB_{ij} \cup \{i, j\}$, gdzie: NB_{ij} - zbiór wierzchołków przyległych do wierzchołka i lub wierzchołka j . W oparciu o podgraf

G'_{ij} konstruowana jest podsieć $S'_{ij} = \left(G'_{ij} = \langle NB_{ij} \cup \{i, j\}, E'_{ij} \rangle, \emptyset, \{h_k(e)\}_{k \in \{1, \dots, LH\}} \right)_{e \in E'_{ij}}$, co

w konsekwencji pozwala na wyznaczenie elementu $s_{ij}(k)$ według następującego wzoru:

$$s_{ij}(k) = \frac{\sum_{l \in NB_{ij}} \left[1 - \left| \frac{h_k(i,l) - h_k(j,l)}{h_k} \right| \right]}{|NB_{ij}|} \in [0,1],$$

gdzie:

¹ Przypomnijmy, że podgraf G' grafu G powstaje w ten sposób, że wybieramy podzbiór V' zbioru wierzchołków V grafu G , który staje się zbiorem wierzchołków podgrafu G' , a następnie do zbioru E' gałęzi podgrafu G' zaliczamy wszystkie gałęzie grafu G , które występują między wierzchołkami należącymi do V' .

$$h_k^{ij} = \max_{(i,j) \in E_{ij}'} h_k(i,j) .$$

Uwaga! We wzorze na $s_{ij}(k)$ przyjęto założenie, że jeśli wierzchołki v i v' nie są do siebie przyległe, czyli nie istnieje pomiędzy nimi krawędź, wówczas $h_k(v, v') = 0$.

Wartość elementu $s_{ij}(k)$ macierzy Sim_k interpretowana jest jako stopień podobieństwa wierzchołka i do wierzchołka j ze względu na k -tą funkcję opisaną na krawędziach grafu G , biorąc pod uwagę jedynie najbliższe sąsiedztwo wierzchołków i oraz j , tj. wierzchołki przyległe do i lub j . Wartość $s_{ij}(k) = 1$ oznacza, że wierzchołki i oraz j są identyczne z punktu widzenia k -tej funkcji na krawędziach grafu G .

Krok 2. Wyznaczenie skumulowanej macierzy podobieństwa wierzchołków:

Mając obliczone wartości elementów wektora macierzy podobieństwa wierzchołków $Sim(S) = \langle Sim_1, \dots, Sim_k, \dots, Sim_{LH} \rangle$, można obliczyć wartości elementów skumulowanej macierzy podobieństwa wierzchołków Sim^{Total} . W tym celu proponuje się wykorzystać jedną z metod redukcji problemów wielokryterialnych polegającą na utworzeniu kombinacji liniowej wektora kryteriów:

$$Sim^{Total} = \sum_{k=1}^{LH} \alpha_k \cdot Sim_k ,$$

gdzie:

α_k - wagi ustalane przez decydenta spełniające następujące warunki:

$\alpha_k \in [0,1], \forall k = 1, \dots, LH$

$$\sum_{k=1}^{LH} \alpha_k = 1 .$$

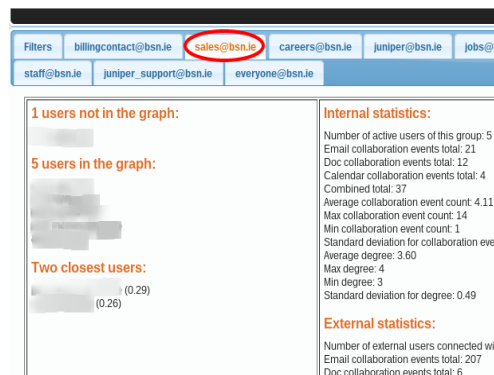
Wyliczona w powyższy sposób macierz Sim^{Total} mówi, w jakim stopniu każda para wierzchołków grafu G jest do siebie podobna przy zadanych arbitralnie przez decydenta wagach.

5 Praktyczne zastosowanie metody

Przedstawiony algorytm badania podobieństwa wierzchołków został zaprojektowana i zaimplementowany na potrzeby jednej z usług systemu *General Audit Tool* (GAT) [21, 22, 23] dla platformy *Google Apps* [24] – rysunek 8. Opracowana usługa systemu GAT wskazuje pracowników, którzy są potencjalnie w stanie najszybciej przejąć od siebie obowiązki (są do siebie najbardziej podobni w sensie realizowanych zadań w firmie), np. podczas ich urlopu lub zmiany pracodawcy – rysunek 9.



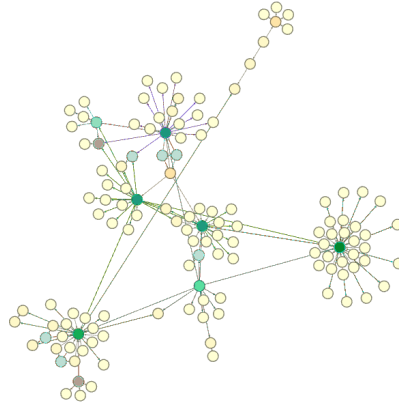
Rys. 8. GAT - panel główny [opracowanie własne]
 Fig. 8. GAT – main panel [own preparation]



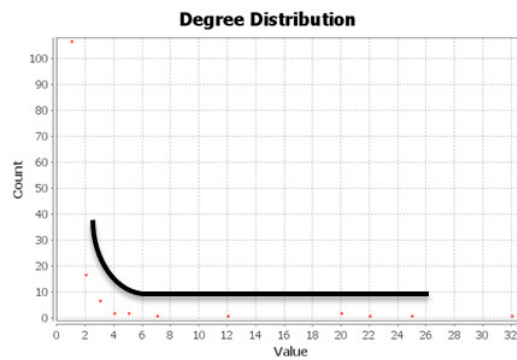
Rys. 9. GAT – analiza podobieństwa pracowników działu sprzedaży
 [opracowanie własne]
 Fig. 9. GAT – similarity analysis of employees in sales department
 [own preparation]

Sieć współpracy pomiędzy pracownikami, będąca podstawą działania prezentowanej usługi, budowana jest między innymi w oparciu o: wymianę wiadomości e-mail, wzajemną pracę nad dokumentami, wspólne spotkania (na podstawie kalendarzy pracowników), jak również wpisy na portalach społecznościowych.

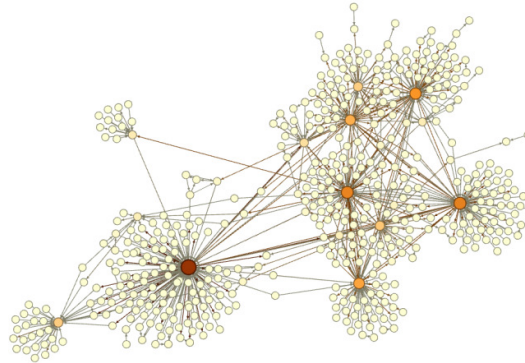
W dalszej części rozdziału zaprezentowano przeprowadzone badania na sieci odzwierciedlające współpracę pracowników rzeczywistej organizacji składającej się z 50-ciu pracowników, w oparciu o zebrane dane z okresu trzech miesięcy w próbkach tygodniowych.



Rys. 10. Graf pracowników – 1-szy miesiąc współpracy [opracowanie własne]
Fig. 10. Employees graph – 1st month of collaboration [own preparation]

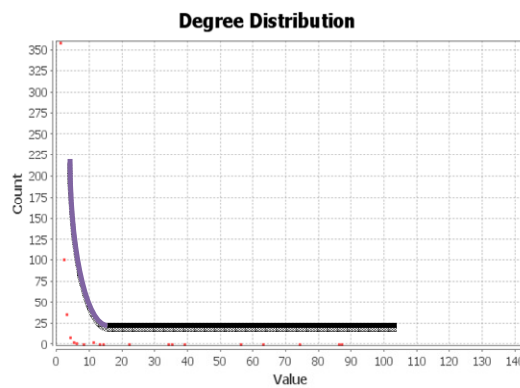


Rys. 11. Rozkład stopni wierzchołków – 1-szy miesiąc współpracy
[opracowanie własne]
Fig. 11. Degree distribution – 1st month of collaboration
[own preparation]



Rys. 12. Graf pracowników organizacji – 3-ci miesiąc współpracy
[opracowanie własne]

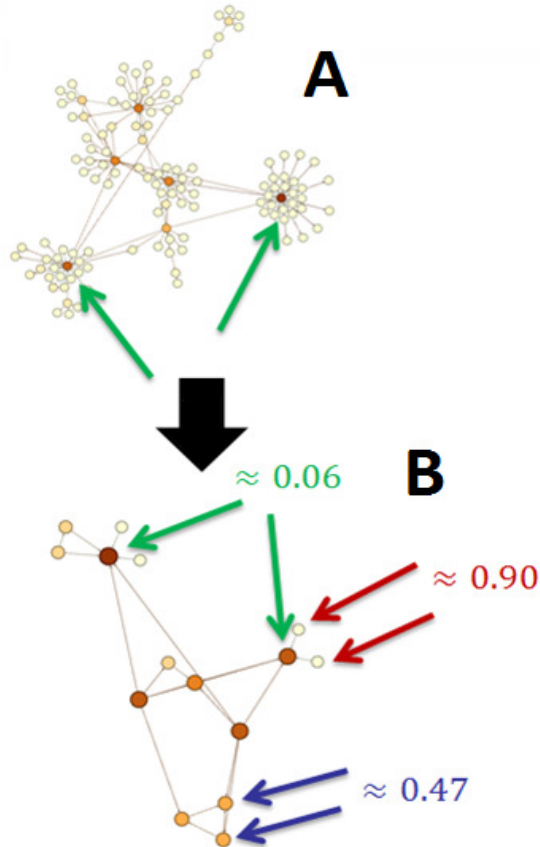
Fig. 12. Employees graph – 3rd month of collaboration [own preparation]



Rys. 13. Rozkład stopni wierzchołków – 3-ci miesiąc współpracy
[opracowanie własne]

Fig. 13. Degree distribution – 3rd month of collaboration [own preparation]

Na rysunku 10 i 12 przedstawiono odpowiednio grafy reprezentujące pracowników i ich współpracę z innymi osobami w organizacji, a także poza nią po 1-szym i po 3-cim miesiącu zbierania danych. Analiza tych sieci pokazała, że mają one charakter sieci bezskalowych (ang. *scale-free*). Sieci te cechują się potęgowym rozkładem stopni wierzchołków (rys. 11 i rys. 13). Większość wierzchołków posiada względnie mały stopień, podczas gdy pewna niewielka ich liczba ma stopień bardzo duży w porównaniu do reszty. Takie wierzchołki nazywa się *hubami*. Dodatkowo badane sieci mają względnie mało krawędzi (są *grafami rzadkimi*) i stosunkowo niewielką średnią odległość pomiędzy wierzchołkami (około 2-3).



Rys. 14. Graf pracowników organizacji – podobieństwo wierzchołków (pracowników)
[opracowanie własne]

Fig. 14. Employees graph – nodes similarity [own preparation]

Na rysunku 14 przedstawiono dwie sieci. Sieć A pokazuje współpracę pracowników między sobą i osobami spoza organizacji. Sieć B reprezentuje tę samą sieć, ale z usuniętymi osobami „zewnętrznymi”. Po zastosowaniu opisanej wcześniej metody wierzchołki wskazywane przez czerwone strzałki otrzymały wysoką wartość współczynnika podobieństwa (około 0,9). Oznacza to, iż pracownicy reprezentowani przez te wierzchołki charakteryzują się bardzo podobnymi cechami. Rzeczywiście może tak być, ponieważ ze struktury sieci wynika, że mają wspólnego sąsiada, którym może być, na przykład kierownik zespołu. Ponadto pracownicy w ramach tego zespołu mogą wykonywać podobne zadania, np. programiści. Natomiast wierzchołkami wskazywanymi przez zielone strzałki mogą być kierownicy dwóch zespołów zajmujących się różnymi, niezwiązanymi ze sobą dziedzinami. Dlatego wartość współczynnika podobieństwa jest niska i wynosi około 0,06.

7 Wnioski

W artykule przedstawiona została autorska metoda analizy podobieństwa grafów i sieci. Otrzymane wyniki badania konkretnego problemu, tj. analizy podobieństwa pracowników, pokazały, że metody ilościowe mogą być z powodzeniem stosowane w takich zadaniach. Trudność analizy polega na identyfikacji istotnych i mierzalnych cech wśród zbioru wszystkich cech, które posiadają obiekty. W opisanym problemie uwaga została skupiona tylko na cechach związanych z „aktywnością” osób w organizacji (wynikało to z potrzeb biznesowych). Badając podobieństwo obiektów, należy zidentyfikować i uwzględnić jedynie te cechy, które są istotne z punktu widzenia celu modelowania, czyli mające znaczenie w danym kontekście.

Ciekawym spostrzeżeniem jest fakt, że analizowana sieć organizacji rzeczywistej ma charakter sieci bezskalowych. Istotnymi wierzchołkami w takiej sieci są huby – świadczy o tym niska wartość współczynnika podobieństwa (im mniejsza wartość, tym obiekt jest bardziej „unikalny”, więc trudniej go zastąpić). Ponadto zmiany w „hubach” znacznie wpływają na rezultaty wyszukiwania podobieństw wśród pracowników. Informacje o hubach mogą być bardzo pomocne. Liczba krawędzi i wartości atrybutów opisanych na nich wskazują, że mogą to być na przykład kierownicy działów, kadra zarządzająca itp. Pozostałe wierzchołki (szczególnie te, których sąsiadem jest ten sam hub) mają wysoki współczynnik podobieństwa. Takie wierzchołki są mniej istotne i można je łatwo zastąpić – w przeciwieństwie do huba.

Warto zauważyć, że jednym z rozszerzeń przedstawionej metody może być zastosowanie jej w połączeniu z technikami symulacji - umożliwienie badania wpływu sposobu zarządzania organizacją (w tym struktury organizacji) na jej elastyczność (łatwość dostosowania się do zmiany pracowników).

Literatura

1. Korzan B.: *Elementy teorii grafów i sieci : metody i zastosowania*, WNT, 1978
2. Newman M. E. J.: *Networks. An Introduction*, Oxford University Press, 2010
3. van Steen M.: *Graph Theory and Complex Networks: An Introduction*, 2010
4. Easley D., Kleinberg J.: *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010
5. Jungnickel D.: *Graphs, Networks and Algorithms*, Springer, 2007
6. Chudy M.: *Wybrane metody optymalizacji*, EXIT, 2014
7. Luks E. M.: Isomorphism of graphs of bounded valence can be tested in polynomial time. *Journal of Computer and System Sciences*, v. 25, 42-65, 1982
8. McKay B. D.: *Practical graph isomorphism*, *Congressus Numerantium*, v. 30, 45-87, 1981
9. Fortin S.: *The Graph Isomorphism Problem*, Technical Raport TR 96-20, Department of Computer Science, The University of Alberta, July 1996
10. Papadopoulos A. N., Manolopoulos Y.: Structure-based similarity search with graph histograms, *DEXA '99 Proceedings of the 10th International Workshop on Database & Expert Systems Applications*, 174-178, 1999

11. Jeh G., Widom J.: SimRank: a measure of structural-context similarity, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, Edmonton, 2002
12. Blondel V. D., Gajardo A., Heymans M., Senellart P., Dooren P. V.: A measure of similarity between graph vertices applications to synonym extraction and web searching, *SIAM Review*, v. 46 (4), 647-666
13. Tarapata Z.: Multicriteria weighted graphs similarity and its application for decision situation pattern matching problem, *Proceedings of the 13th IEEE/IFAC International Conference on Methods and Models in Automation and Robotics*, 1149-1155, 2007
14. Zager L.: *Graph similarity and matching*, PhD thesis, MIT, May 2005
15. Conte D., Foggia P., Sansone C., Vento M.: Thirty years of graph matching in pattern recognition, *International Journal of Pattern Recognition and Artificial Intelligence*, v. 18 (3), 2004
16. Tarapata Z., Kasprzyk R.: An application of multicriteria weighted graph similarity method to social networks analyzing, *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, 366-368, 2009
17. Kasprzyk R.: Diffusion in Networks, *Journal of Telecommunications and Information Technology*, 2/2012, ISSN 1509-4553, 99-106
18. Bartosiak C., Kasprzyk R., Najgebauer A.: The graph and network theory as a tool to model and simulate the dynamics of infectious diseases, *Bio-Algorithms and Med-Systems*. Volume 9, Issue 1, Pages 17–28, ISSN (Online) 1896-530X, ISSN (Print) 1895-9091, DOI: 10.1515/bams-2013-0003, March 2013
19. Kasprzyk R., Najgebauer A., Pierzchała D.: Modelling and Simulation of an Infected Disease in Social Networks, *Lecture Notes in Artificial Intelligence*, 2011, Volume Part I, pp. 388-398, ISBN: 978-3-642-23934-2, Springer-Verlag Berlin, Heidelberg
20. Tarapata Z., Kasprzyk R.: Graph-Based Optimization Method for Information Diffusion and Attack Durability in Networks, RSCTC'2010, *Lecture Notes in Artificial Intelligence*, vol. 6086, Springer, Heidelberg, pp. 698-709, 2010
21. www.bsn.ie
22. www.generalaudittool.com
23. <https://plus.google.com/communities/108178540809040162296>
24. www.google.pl/intx/pl/enterprise/apps/business/

Streszczenie

W niniejszej pracy autorzy poruszają zagadnienia z obszaru badania podobieństwa grafów i sieci ze szczególnym naciskiem na metody ilościowe. Zostaje podjęta próba klasyfikacji wybranych metod grafowo-sieciowych badania podobieństwa. Prezentuje się tu także autorską metodę i rozwiązanie programowe, które jest wykorzystywane w celu wspierania działań decyzyjnych w organizacjach.

Słowa kluczowe: sieć, rozpoznawanie wzorców, miary podobieństwa sieci

Models and methods of quantitative computer patterns recognition

Summary

In this paper, authors describe issues of assessing similarity of graphs and networks, with an emphasis on quantitative methods. An attempt is taken to classify graphs and networks similarity methods. It also presents author's method and software solution that is used to support decision-making processes in organizations.

Keywords: network, pattern recognition, network similarity measures