ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2008
Seria: AUTOMATYKA z. 151 Nr kol. 1797

Adam KOZAK, Tomasz GŁOWACKI, Piotr FORMANOWICZ
Politechnika Poznańska

GRAFY BAZOWO-ETYKIETOWALNE JAKO JEDNOLITA REPREZENTACJA GRAFÓW SPRZĘŻONYCH

Streszczenie. Praca stanowi wprowadzenie do klasy grafów (α, k, σ)-bazowo-etykietowalnych, które przedstawiają ujednoliconą reprezentację matematyczną grafów (α, k)-etykietowalnych oraz (α, k)-dowolnie-etykietowalnych (które są równoważne klasie grafów sprzężonych [1]). Omówione są podstawowe zależności między tymi klasami, zależności między parametrami k i σ oraz znaczenie tych klas dla kombinatoryki wraz z praktycznymi przykładami zastosowań w bioinformatyce.

BASE-LABELED GRAPHS AS UNIFIED REPRESENTATION OF ADJOINTS

Summary. This work introduces a class of (α, k, σ)-base-labeled graphs that presents unified mathematical representation of (α, k)-labeled and (α, k)-free-labeled graphs (that are equivalent to adjoints [1]). It describes basic relationships between these classes, relationship between parameters k and σ and the meaning of these classes for computing science. There are also some practical examples of application in bioinformatics.

1. Wprowadzenie do grafów etykietowalnych i sprzężonych

Ważnym problemem w informatyce jest problem szukania cyklu Hamiltona w grafie. Cykl Hamiltona jest to cykl zawierający każdy wierzchołek grafu dokładnie raz. Prawdopodobnie nie istnieje algorytm efektywnie rozstrzygający (w czasie wielomianowym) istnienie takiego cyklu w dowolnym grafie. Istnieje jednak podklasa grafów, dla których cykl Hamiltona jest równoważny cyklowi Eulera w pewnej transformacji tego grafu. Cykl Eulera jest to cykl zawierający wszystkie krawędzie grafu dokładnie raz. Zarówno rozstrzygnięcie istnienia tego cyklu, jak i jego znalezienie jest problemem prostym, dla którego istnieją efektywne algorytmy. Grafy, dla których zachodzi taka równoważność, są grafami sprzężonymi i należą do klasy grafów (α, k)-dowolnie-etykietowalnych. Przed omówieniem tych klas należy wprowadzić następujące definicje (domyślnie graf będzie oznaczać graf skierowany):

Definicja 1. *Grafem (1-grafem)* skierowanym nazywamy uporządkowaną parę $G = (V, E)$, gdzie $V = \{v_1, v_2, \ldots, v_n\}$ jest zbiorem wierzchołków, a E jest zbiorem krawędzi: $E \subseteq \{(v_1, v_2) : v_1, v_2 \in V\}$.

Definicja 2. *Multigrafem* nazywamy trójkę $G = (V, E, \varphi)$, gdzie (V, E) jest grafem, a φ jest funkcją: $\varphi : E \to N$. Jeśli $e \in E$ jest krawędzią, to $\varphi(e)$ nazywamy krotnością.

Multigraf jest *p-grafem*, jeśli maksymalna krotność krawędzi jest nie większa niż p, czyli $\varphi : E \to \{1, 2, \ldots, p\}$.

Notacja. Niech $N^+(v)$ oznacza zbiór następników wierzchołka v, a $N^-(v)$ oznacza zbiór poprzedników wierzchołka v.

Definicja 3. [2] 1-Graf skierowany $H(E, U)$ jest *grafem sprzężonym* grafu $G(V, E)$ o zbiorze wierzchołków E i takim zbiorze krawędzi U, że między wierzchołkami $x, y \in E$ w grafie H występuje krawędź wtedy i tylko wtedy, gdy wierzchołek końcowy krawędzi x w grafie G jest wierzchołkiem początkowym krawędzi y w grafie G.

Równoważność między cyklem Eulera w grafie $G(V, E)$ a cyklem Hamiltona w grafie $H(E, U)$ wynika bezpośrednio z definicji 3. Praktyczne wykorzystanie tego faktu do szukania cyklu Hamiltona wymaga możliwości efektywnego sprawdzenia, czy dany graf jest grafem sprzężonym. Możliwość taka istnieje i jest określona przez poniższe twierdzenie:

Twierdzenie 1. *[2] 1-graf $H = (V, E)$ jest grafem sprzężonym innego grafu wtedy i tylko wtedy, gdy dla każdej pary wierzchołków $x, y \in V$ spełniony jest następujący warunek:*

$$N^+(x) \cap N^+(y) \neq \emptyset \Rightarrow N^+(x) = N^+(y)$$

Z definicji 3 wynika, że każdy 1-graf G posiada swój graf sprzężony H. Jednak nie każdy graf sprzężony H posiada odpowiadający mu 1-graf. Jeśli H jest grafem sprzężonym 1-grafu, to jest *skierowanym grafem liniowym* (dla multigrafów również istnieją odpowiadające im grafy sprzężone). Podobnie istnieje warunek, jaki muszą spełniać wierzchołki grafu, aby był grafem liniowym:

Twierdzenie 2. *[1] 1-graf $G(V, E)$ jest skierowanym grafem liniowym wtedy i tylko wtedy, gdy spełniony jest następujący warunek:*

$$\forall x, y \in V : \left(x \neq y \wedge N^+(x) \cap N^+(y) \neq \emptyset\right) \Rightarrow \left(N^+(x) = N^+(y) \wedge N^-(x) \cap N^-(y) = \emptyset\right)$$

Można zauważyć, że jeśli w danym grafie przyporządkowana zostanie wierzchołkowi k-literowa etykieta (nad alfabetem o mocy α), a krawędzie zostaną poprowadzone między takimi wierzchołkami (v_1, v_2), że $k-1$ ostatnich liter etykiety v_1 jest równych $k-1$ literom etykiety v_2, to zostanie spełniony warunek twierdzenia 1, czyli graf będzie grafem sprzężonym (wynika to z faktu, że dowolne dwa wierzchołki o identycznych $k-1$ ostatnich literach etykiety będą miały dokładnie ten sam zbiór następników – czyli zbiór wierzchołków, dla których pokrywa się $k-1$ pierwszych liter etykiety). Definiuje się w ten sposób grafy (α, k)-dowolnie-etykietowalne.

Definicja 4. [1] Niech $k > 1, \alpha > 0$ będą liczbami całkowitymi. Wtedy 1-graf $H = (V, E)$ jest grafem (α, k)-*dowolnie-etykietowalnym* (należy do klasy $\mathcal{L}_k^\alpha$), jeśli możliwe jest przyporządkowanie każdemu wierzchołkowi $x \in V$ etykiety $(l_1(x), l_2(x), \ldots, l_k(x))$ o długości k takiej, że $\forall i \forall x \in V : l_i(x) \in \{0, \ldots, \alpha - 1\}$, oraz istnieje krawędź między x i y wtedy i tylko wtedy, gdy $(l_2(x), \ldots, l_k(x)) = (l_1(y), \ldots, l_{k-1}(y))$.

Można dowieść ([1]), że dodanie do definicji 4 warunku unikalności etykiet powoduje uzyskanie warunku twierdzenia 2, czyli grafy liniowe są grafami (α, k)-etykietowalnymi.

Definicja 5. [1] Niech $k > 1, \alpha > 0$ będą liczbami całkowitymi. Wtedy 1-graf $H(V, E)$ jest grafem (α, k)-*etykietowalnym* (należy do klasy $\mathscr{L}_k^\alpha$), jeśli jest grafem (α, k)-dowolnie-etykietowalnym oraz wszystkie etykiety są różne, tj. $\forall x \neq y$: $(l_1(x), \ldots, l_k(x)) \neq (l_1(y), \ldots, l_k(y))$.

2. Grafy (α, k, σ)-bazowo-etykietowalne

Dla grafów z klas $\widetilde{\mathscr{L}}_k^\alpha$ oraz $\mathscr{L}_k^\alpha$ dla istnienia krawędzi wymagane jest nałożenie etykiet między wierzchołkami równe $k-1$. Grafy (α, k, σ)-bazowo-etykietowalne różnią się tym od grafów (α, k)-etykietowalnych, że nałożenie β (zwane tutaj długością bazy krawędzi) może przyjmować wartości: $\beta \in \{0, \ldots, k-1\}$ (dla $\beta = k$ graf całkowicie traci spójność, a przez to też większość własności). Parametr σ jest *stopniem swobody* krawędzi i jest równy $k - \beta$. Grafy te można zdefiniować następująco:

Definicja 6. Niech $k \geqslant 1, \alpha \geqslant 2, \sigma \in \{1, \ldots, k\}, \beta = k - \sigma$ będą liczbami całkowitymi. Wtedy 1-graf skierowany $\ell_{k,\sigma}^\alpha(V, E)$ jest grafem (α, k, σ)-*bazowo-etykietowalnym* (należy do klasy $\mathscr{L}_{k,\sigma}^\alpha$) o wymiarze k, stopniu swobody σ oraz długości bazy krawędzi β, jeśli możliwe jest przyporządkowanie każdemu wierzchołkowi $x \in V$ etykiety $(l_1(x), l_2(x), \ldots, l_k(x))$ o długości k takiej, że:

1) $\forall i \forall x \in V : l_i(x) \in \{0, \ldots, \alpha - 1\}$,
2) wszystkie etykiety są różne, tj. $\forall x \neq y : (l_1(x), \ldots, l_k(x)) \neq (l_1(y), \ldots, l_k(y))$,
3) istnieje krawędź między x i y wtedy i tylko wtedy, gdy β końcowych liter etykiety wierzchołka x jest równych β początkowym literom etykiety wierzchołka y, tj.:
$(x, y) \in E \Leftrightarrow (l_{\sigma+1}(x), \ldots, l_k(x)) = (l_1(y), \ldots, l_\beta(y))$.

Notacja. Dla każdego wierzchołka $x \in V$ grafu $\ell_{k,\sigma}^\alpha(V, E)$ wprowadza się następujące symbole i oznaczenia etykiet ($\beta = k - \sigma$):

1. $\epsilon(x)$ oznacza etykietę wierzchołka x, a $l_i(x) = l_i(\epsilon(x))$ oznacza i-tą literę etykiety.
2. Δ_x oznacza *bazę krawędzi wchodzących*, czyli pierwszych β liter $\epsilon(x)$.
3. δ_x oznacza *swobodę krawędzi wchodzących*, czyli ostatnich σ liter $\epsilon(x)$.
4. Λ_x oznacza *bazę krawędzi wychodzących*, czyli ostatnich β liter $\epsilon(x)$.
5. λ_x oznacza *swobodę krawędzi wychodzących*, czyli pierwszych σ liter $\epsilon(x)$.

W grafie $\ell_{k,\sigma}^\alpha(V, E)$ o wierzchołkach $x, y \in V$ (oznaczonych w powyższej notacji) warunek istnienia krawędzi zgodnie z definicją 6 można sformułować następująco:

$$(x, y) \in E \Leftrightarrow \Lambda_x = \Delta_y \tag{1}$$

Twierdzenie 3. *Grafy bazowo-etykietowalne są grafami sprzężonymi.*

Dowód. Zgodnie z twierdzeniem 1, graf $G(V,E)$ jest grafem sprzężonym wtedy i tylko wtedy, gdy:

$$\forall x,y \in V : N^+(x) \cap N^+(y) \neq \emptyset \Rightarrow N^+(x) = N^+(y) \tag{2}$$

Niech etykiety wierzchołków $x,y \in V$ w grafie $G(V,E) \in \mathscr{L}^{\alpha}_{k,\sigma}$ zostaną oznaczone w następujący sposób: $\epsilon(x) = (\lambda_x, \Lambda_x), \epsilon(y) = (\Delta_y, \delta_y)$. Można zauważyć, że dla wierzchołków x,y zachodzi następująca równoważność:

$$N^+(x) \cap N^+(y) \neq \emptyset \Leftrightarrow \Lambda_x = \Lambda_y \tag{3}$$

Stosując warunek (1) dla wierzchołków x,y, otrzymuje się następujące zbiory następników:

$$N^+(x) = \{v \in V : \Lambda_x = \Delta_v\} \qquad N^+(y) = \{v \in V : \Lambda_y = \Delta_v\} \tag{4}$$

Jeśli dla x,y spełniony jest warunek $\Lambda_x = \Lambda_y$, to na mocy równań (4) zachodzi $N^+(x) = N^+(y)$ – w połączeniu z zależnością (3) implikuje to warunek (2), co kończy dowód. □

Grafy bazowo-etykietowalne są więc jednocześnie grafami dowolnie-etykietowalnymi. Wynika to z faktu, że obie klasy pokrywają się z klasą grafów sprzężonych (dla grafów dowolnie-etykietowalnych wykazano to w pracy [1]). Można zauważyć, że dla $\sigma = k$ graf bazowo-etykietowalny jest grafem pełnym, więc nie może być skierowanym grafem liniowym (nie jest spełniony warunek twierdzenia 2). Jednocześnie dla $\sigma = 1$ graf (α,k,σ)-bazowo-etykietowalny jest grafem (α,k)-etykietowalnym, więc jest skierowanym grafem liniowym. Oznacza to, że parametr σ determinuje przynależność grafu do klasy $\mathscr{L}^{\alpha}_{k}$. Zależność tę określa poniższe twierdzenie.

Twierdzenie 4. *Jeśli $k \geqslant 2\sigma$, to grafy (α,k,σ)-bazowo-etykietowalne są skierowanymi grafami liniowymi.*

Dowód. Niech dla grafu $\ell^{\alpha}_{k,\sigma}(V,E)$ zachodzi $k \geqslant 2\sigma$. Na mocy twierdzenia 2 graf $\ell^{\alpha}_{k,\sigma}(V,E)$ jest grafem liniowym wtedy i tylko wtedy, gdy

$$\forall x,y \in V : \big(x \neq y \wedge N^+(x) \cap N^+(y) \neq \emptyset\big) \Rightarrow \big(N^+(x) = N^+(y) \wedge N^-(x) \cap N^-(y) = \emptyset\big) \tag{5}$$

Niech etykiety wierzchołków $x,y \in V$ w grafie $\ell^{\alpha}_{k,\sigma}(V,E)$ zostaną oznaczone w następujący sposób: $\epsilon(x) = (\lambda_x, \Lambda_x), \epsilon(y) = (\Delta_y, \delta_y)$.
Jeśli $k \geqslant 2\sigma$, to zachodzą implikacje:

$$1.\ \Lambda_x = \Lambda_y \Rightarrow \Delta_x \neq \Delta_y \qquad 2.\ \Delta_x = \Delta_y \Rightarrow \Lambda_x \neq \Lambda_y \tag{6}$$

Zależności (6) wynikają z faktu, że jeśli $k \geqslant 2\sigma$, to:

$$(\Lambda_x = \Lambda_y \wedge \Delta_x = \Delta_y \Rightarrow \lambda_x = \lambda_y \wedge \delta_x = \delta_y) \Rightarrow (\Lambda_x = \Lambda_y \wedge \Delta_x = \Delta_y \Leftrightarrow \epsilon(x) = \epsilon(y))$$

Dla $x \neq y$ nie może jednak zachodzić $\epsilon(x) = \epsilon(y)$ ze względu na definicję 6 (implikacje (6) nie zachodzą dla $\sigma < k < 2\sigma$).

Analogicznie do warunku (3) z dowodu twierdzenia 3 dotyczącego następników można zapisać warunek dotyczący poprzedników:

$$N^-(x) \cap N^-(y) \neq \emptyset \Leftrightarrow \Delta_x = \Delta_y \tag{7}$$

Łącząc zależności (6) (implikacja 1) oraz (7), otrzymuje się implikację:

$$\Lambda_x = \Lambda_y \Rightarrow N^-(x) \cap N^-(y) = \emptyset \tag{8}$$

Ostatecznie z warunków (8) oraz warunków (3), (4) z dowodu twierdzenia 3 wynika warunek (5), czego należało dowieść. □

Dla grafu $\ell^{\alpha}_{k,\sigma}(V,E)$ (w którym parametry są w relacji $\sigma < k < 2\sigma$) zachodzi zależność:

$$\forall x, y \in V : (\Lambda_x = \Lambda_y \wedge \Delta_x = \Delta_y) \Rightarrow (N^+(x) = N^+(y) \wedge N^-(x) = N^-(y)) \tag{9}$$

Niech $H(V_H, E_H)$ będzie grafem dowolnie-etykietowalnym odpowiadającym grafowi $\ell^{\alpha}_{k,\sigma}(V,E)$. Można zauważyć, że wierzchołki $x, y \in V_H$, dla których zachodziłaby implikacja (9), musiałyby mieć identyczne etykiety (wynika to z faktu, że zbiór poprzedników oraz następników tych wierzchołków jest taki sam). Graf ten nie byłby więc skierowanym grafem liniowym.

3. Zastosowanie grafów z klas $\widetilde{\mathscr{L}}^{\alpha}_{k}, \mathscr{L}^{\alpha}_{k}, \mathscr{L}^{\alpha}_{k,\sigma}$

Omawiane klasy grafów mają istotne znaczenie dla teorii złożoności obliczeniowej, ze względu na możliwość efektywnego szukania cyklu Hamiltona. Znane algorytmy szukające cyklu Hamiltona dla dowolnego grafu mają złożoność wykładniczą względem liczby wierzchołków [3], natomiast algorytmy szukające cyklu Eulera mają złożoność liniową względem liczby krawędzi. Grafy te mogą być modelem danych dla dowolnego problemu, w którym odwzorowuje się podobieństwo między elementami pewnego zbioru polegające na odpowiednim nałożeniu etykiet tych elementów.

Problemy takie są często spotykane w bioinformatyce, gdzie analiza łańcuchów DNA wymaga wstępnego cięcia tych łańcuchów, a następnie odwzorowania uzyskanych fragmentów w całość. Przykładem jest ustalanie kolejności nukleotydów w łańcuchu DNA metodą sekwencjonowania przez hybrydyzację [4]. Wynikiem fazy biochemicznej jest informacja o zbiorze wszystkich fragmentów badanego łańcucha o długości l. Elementy tego zbioru zachodzą na siebie, dając graf (4,l)-etykietowalny (graf DNA [1]). Odtworzenie badanego łańcucha następuje przez znalezienie cyklu Hamiltona w tym grafie.

Motywacją do zdefiniowania grafów bazowo-etykietowalnych był problem budowy biblioteki oligonukleotydów, które mają możliwie małe powinowactwo do hybrydyzacji między sobą. Elementy takich bibliotek mogą być stosowane do kodowania instancji w algorytmach opartych na DNA (komputery DNA [5]). W procesie konstrukcji tego typu bibliotek budowany jest graf (α, k, σ)-bazowo-etykietowalny, który następnie modyfikowany jest tak, że krawędzie grafu zostają w pewnej zdefiniowanej relacji. W tak przygotowanym grafie szukany jest cykl Eulera. Następnie sumuje się swobody krawędzi wchodzących (δ_x) etykiet wierzchołków występujących kolejno w cyklu, budując łańcuch wynikowy, który dalej jest cięty zgodnie z liczbą potrzebnych elementów biblioteki. W zależności od wyboru parametru σ powstają różne jakościowo rozwiązania, gdyż każdy wierzchołek reprezentuje inną liczbę oligonukleotydów przeniesionych do łańcucha wynikowego.

4. Podsumowanie

Grafy bazowo-etykietowalne są klasą grafów, które łączą istniejące już klasy w jednym modelu matematycznym. Nie posiadają więc nowych własności istotnych dla złożoności obliczeniowej istniejących problemów, ale mogą wprowadzić zmiany jakościowe w modelach matematycznych niektórych problemów. Wspomniany przypadek budowy biblioteki oligonukleotydów pokazuje, że oparcie modelu na grafach bazowo-etykietowalnych pozwala na bardziej ogólne sformułowanie problemu i analizę nowych własności rozwiązań.

BIBLIOGRAFIA

1. Błażewicz J., Hertz A., Kobler D., de Werra, D.: On some properties of DNA graphs. Discrete Applied Mathematics, 98, 1999, p. 1–19.
2. Berge C.: Graphes. Dunod, Paris 1970.
3. Papadimitriou Ch. H.: Złożoność obliczeniowa. WNT, Warszawa 2002.
4. Kasprzak M.: On the link between DNA sequencing and graph theory. Computational Methods in Science and Technology, 10, 2004, p. 39–47.
5. Adleman L. M.: Molecular computation of solutions to combinatorial problems. Science, 266, 1994, p. 1021–1024.

Recenzent: Prof. dr hab. inż. Andrzej Polański

Abstract

An adjoint of graph $G(V,E)$ is graph $H(E,U)$ where set of vertices is equal to set of arcs in graph $G(V,E)$. Transformation $G(V,E) \rightarrow H(E,U)$ has a property, that Hamiltonian circuit in graph H is equivalent to Eulerian circuit in graph G. Every adjoint of a graph is also (α,k)-free-labeled graph. This means that every vertex has a label of length k over alphabet of length α. There is an arc between vertices v_1, v_2 if and only if last $k-1$ letters of v_1 label are equal to first $k-1$ letters of v_2 label (an overlap is $k-1$ long). Moreover, if every label is unique, the graph is (α,k)-labeled.

This work introduces class of (α,k,σ)-base-labeled graphs that presents unified mathematical representation of (α,k)-labeled and (α,k)-free-labeled graphs. In graphs of this class an overlap between labels must be $k-\sigma$ long. Relationship between parameters k and σ allows to determine if graph is (α,k)-labeled or (α,k)-free-labeled.

The need of modifying overlap constraint comes from a bioinformatics problem – building an oligonucleotide library that contains DNA chains which have minimal tendency to hybridize witch each other. Libraries of this type are used in DNA computing to encode problem instances.