

Piotr WAWRZYŃIAK, Piotr FORMANOWICZ
Instytut Informatyki
Politechnika Poznańska

PROBLEMY REALIZOWALNOŚCI GRAFÓW I ICH ZASTOSOWANIA W IDENTYFIKACJI STRUKTURY ZWIĄZKÓW CHEMICZNYCH*

Streszczenie. Realizowalność grafów jest to problem decyzyjny, który w swojej klasycznej postaci odpowiada na pytanie, czy skończony ciąg nierosnących liczb dodatnich jest również ciągiem stopni wierzchołków w grafie. Problem ten znajduje różne zastosowania praktyczne, między innymi w biologii. Gdy jednak chcemy zastosować ten problem, np. do rozpoznawania metabolitów za pomocą spektrometrii mas, musimy zmodyfikować jego definicję tak, by przystawała do analizowanej rzeczywistości biologicznej. W niniejszej pracy omówiony jest podstawowy problem realizowalności grafów oraz jego rozszerzenie znajdujące zastosowanie w identyfikacji struktury związków chemicznych.

GRAPH REALIZATION PROBLEMS AND THEIR APPLICATIONS IN THE IDENTIFICATION OF THE STRUCTURE OF CHEMICAL COMPOUNDS

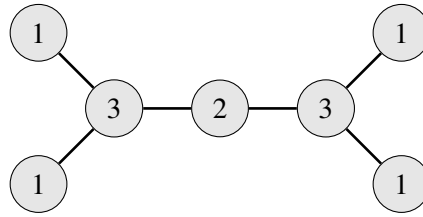
Summary. The graph realization problem is a decision problem that, in its classic form, answers the question of whether a finite sequence of non-increasing positive numbers is also a sequence of degrees of vertices in the graph. This problem has various practical applications, among others in biology. However, when we want to apply this problem, e.g., to the recognition of metabolites using mass spectrometry, we must modify its definition to match the analyzed biological reality. In this paper, the basic graph realization problem as well as its extension suitable for the structural identification of chemical compounds is presented.

1. Wprowadzenie

Realizowalność grafów to jeden z problemów decyzyjnych teorii grafów. Polega on na sprawdzeniu czy ciąg dodatnich liczb całkowitych może być jednocześnie ciągiem stopni wierzchołków w grafie (przykład na rysunku 1).

Grafy możemy podzielić na różne klasy, ze względu na cechy jakie one posiadają. Na potrzeby tej pracy omówimy takie cechy, jak: pętle własne (rys. 2), równoległe krawędzie (rys. 3) oraz spójność (rys. 4). Są to cechy istotne w procesie sprawdzania warunków realizowalności grafów.

*Praca częściowa sfinansowana ze środków statutowych Politechniki Poznańskiej.



Rys. 1. Graf dla ciągu 3, 3, 2, 1, 1, 1, 1 (stopnie podane jako etykiety wierzchołków grafu).

2. Realizowalność grafów prostych

W klasycznej wersji problemu realizowalności, opisaney w 1955 przez Havela [4], mowa jest o grafach bez równoległych krawędzi i pętli własnych. Grafy takie określamy jako grafy proste. Instancja problemu, w tej klasycznej wersji, zdefiniowana jest jako skończony ciąg nierosnących dodatnich liczb całkowitych:

$$D = d_1, d_2, \dots, d_n \text{ gdzie } d_i \geq d_{i+1} \text{ dla } i = 1, 2, \dots, n - 1 \quad (1)$$

a odpowiedź brzmi TAK, jeśli ciąg ten jest również ciągiem stopni wierzchołków w grafie. Taki ciąg nazywamy ciągiem graficznym.

Jednocześnie Havel podał rekurencyjny algorytm, o złożoności wielomianowej, rozwiązujący ten problem. Algorytm polega na odejmowaniu jednościami od k elementów ciągu, następujących po pierwszym elemencie ciągu, który to ma wartość k . Algorytm ten opiera się na następującym spostrzeżeniu [4].

$$\begin{aligned} &\text{Ciąg } D = d_1, d_2, \dots, d_n \text{ jest graficzny} \\ &\text{wtedy i tylko wtedy, gdy} \\ &\text{ciąg } D' = d_{1+1} - 1, \dots, d_{d_1+1} - 1, d_{d_1+2}, \dots, d_n \text{ również jest graficzny.} \end{aligned} \quad (2)$$

Dla przykładu z rysunku 1, poszczególne iteracje algorytmu wyglądają następująco:

```

3, 3 , 2 , 1 , 1, 1, 1 #ciąg do sprawdzenia
3-1, 2-1, 1-1, 1, 1, 1 #zastosowanie algorytmu
2 , 1 , 0 , 1, 1, 1 #wykonanie odejmowania

2, 1 , 1 , 1, 1 #sortowanie, kolejny ciąg do sprawdzenia
1-1, 1-1, 1, 1 #zastosowanie algorytmu
0 , 0 , 1, 1 #wykonanie odejmowania

1, 1 #sortowanie, kolejny ciąg do sprawdzenia
  
```

gdzie ciąg z dwoma jednościami jest w oczywisty sposób graficzny (dwa wierzchołki połączone krawędzią).

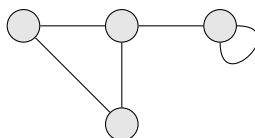
Niezależnie podobne rozwiązanie podał Hakimi [3], a Erdős i Gallai [1] w 1960 zaproponowali inny, wydajniejszy algorytm opisany zbiorem nierówności:

$$\sum_{i=1}^k d_i \leq k(k-1) + \sum_{i=k+1}^n \min(k, d_i) \quad \text{dla każdego } k : 1 \leq k \leq n \quad (3)$$

3. Klasy grafów i warunki ich realizowalności

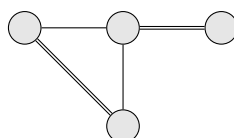
Rozpatrując grafy możemy też rozważać ich wersje inne niż grafy proste. W tym celu możemy wyróżnić cechy grafów, takie jak pętle własne (rys. 2), równoległe krawędzie (rys. 3) lub spójność (rys. 4):

- a) dopuszczenie istnienia pętli własnych; czyli krawędzi z końcami w tym samym wierzchołku;



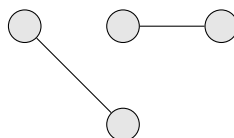
Rys. 2. Graf z pętlą własną.

- b) dopuszczenie istnienia równoległych krawędzi;



Rys. 3. Graf z równoległymi krawędziami.

- c) dopuszczenie istnienia podgrafów rozdzielnych, to jest niepołączonych krawędziami, czyli dopuszczenie grafów niespójnych;



Rys. 4. Graf niespójny z dwoma składowymi.

Dla każdej z powyższych cech można wskazać pewne warunki, które muszą być spełnione, aby ciąg liczb opisywał równocześnie zbiór stopni wierzchołków dla grafów danej klasy.

Jeśli nie mamy żadnych ograniczeń, aby skonstruować graf wystarczy spełnienie lamatu o uściskach dłoni (Handshaking lemma), z którego to wynika, że suma stopni w grafie musi być parzysta [2]. Jest to oczywisty warunek wynikający wprost z faktu, że każda krawędź jest incydentna z dwoma wierzchołkami:

$$\sum_{i=1}^n d_i = 2e, e \in \mathbb{N} \quad (4)$$

Wprowadzając ograniczenie na brak pętli własnych musimy móc upewnić się, że dla wierzchołka o maksymalnym stopniu będzie istniała wystarczająca liczba innych wierzchołków, z którymi może być on połączony za pomocą krawędzi [5]. Wynika to

z faktu, że w przypadku braku pętli własnych, wierzchołki nie mogą łączyć się same ze sobą. Wzór wyrażający ten warunek ma postać:

$$\sum_{i=1}^n d_i \geq 2d_1 \quad (5)$$

Zakaz występowania krawędzi równoległych, zgodnie z pracą Erdős i Gallai [1], można wyrazić następującą zbiorem nierówności:

$$\sum_{i=1}^k d_i \leq k(k-1) + \sum_{i=k+1}^n \min(k, d_i) \quad \text{dla każdego } k : 1 \leq k \leq n \quad (6)$$

Ostatni z warunków, na spójność grafów, sprawdza czy z podanego ciągu liczb można potencjalnie stworzyć graf spójny i jest wyrażony wzorem [3]:

$$\sum_{i=1}^n d_i \geq 2(n-1) \quad (7)$$

Powyższe warunki, wraz z ich przypisaniem do interesujących nas klas grafów, są zebrane w tabeli 1.

Tabela 1

Podsumowanie omówionych klas grafów i warunków ich realizowalności

Pętle własne zakazane	Krawędzie równoległe zakazane	Spójność wymagana	Wymagane warunki realizowalności	Opis
-	-	-	(4)	Grafy bez ograniczeń
-	-	✓	(4) ∧ (7)	
-	✓	-	(4) ∧ (6)	
-	✓	✓	(4) ∧ (6) ∧ (7)	
✓	-	-	(4) ∧ (5)	Multigrafy
✓	-	✓	(4) ∧ (5) ∧ (7)	Grafy molekularne
✓	✓	-	(4) ∧ (5) ∧ (6)	Grafy proste
✓	✓	✓	(4) ∧ (5) ∧ (6) ∧ (7)	Grafy spójne

4. Rozpoznawanie cząsteczek i grafy molekularne

W medycynie, czy też szerzej biologii i chemii, jednym z problemów jest rozpoznawanie cząsteczek metabolitów, czyli niewielkich cząsteczek organicznych lub nieorganicznych produkowanych przez komórki. Rozpoznanie metabolitów umożliwia zdobycie wiedzy o działaniu komórek. Gdy znamy ich prawidłowe działanie, możemy rozpoznać również ich dysfunkcje, a dalej opracować na podstawie tej wiedzy leki, czy też ocenić działanie różnych terapii leczniczych.

Jedną z technik służących rozpoznawaniu metabolitów jest spektrometria masowa. Zwraca ona masę cząsteczkową badanych substancji. Ponieważ masy atomowe poszczególnych pierwiastków są znane, można zastosować problem wydawania reszty

[6] i obliczyć, ile razy poszczególne pierwiastki występują w cząsteczce. Problem wydania reszty, również klasyczny problem informatyki, odpowiada na pytanie jakich i ilu monet należy użyć, aby móc wydać resztę w określonej kwocie. Jeśli otrzymaną masę cząsteczkową M potraktujemy jako kwotę reszty do wydania, znane masy atomowe pierwiastków m_i potraktujemy jako nominały monet, a krotność atomu w cząsteczce x_i będzie odpowiadała krotności monet w wydawanej reszcie, to po uwzględnieniu błędu ε , otrzymamy wzór na znalezienie wzoru sumarycznego związku chemicznego:

$$M - \varepsilon \leq \sum_{i=1}^k m_i x_i \leq M + \varepsilon \quad (8)$$

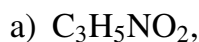
Przykładowo, dla masy $M = 87,077$, błędu $\varepsilon = 0,1$, oraz ograniczając się tylko do czterech podstawowych pierwiastków chemii organicznej, można znaleźć wzory sumaryczne zamieszczone w tabeli 2.

Tabela 2

Przykładowe rozkłady masy w cząsteczce o masie 87,077

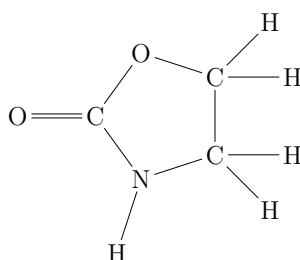
Pierwiastek	Masa atomowa	x	Łączna masa	x	Łączna masa
H	1,0078250	5	5,0391250	3	3,0234750
C	12,0000000	3	36,0000000	7	84,0000000
N	14,0030740	1	14,0030740	-	
O	15,9898292	2	31,9898292	-	
		razem	87,0320282	razem	87,0234750

W omówionym przypadku otrzymujemy dwa następujące wzory sumaryczne:



Czy obydwa ze znalezionych wzorów sumarycznych są prawidłowe? Na to pytanie możemy odpowiedzieć wykorzystując problem realizowalności grafów.

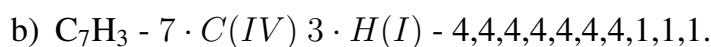
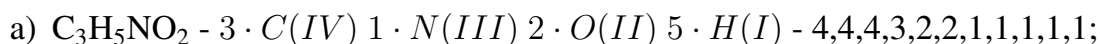
Cząsteczka chemiczna, reprezentowana w postaci wzoru strukturalnego, jest niczym innym jak grafem. Graf ten jest spójnym multigrafem, gdzie wielokrotne krawę-



Rys. 5. Przykładowy graf molekularny - wzór strukturalny cząsteczki.

dzie reprezentują krotności wiązań, a spójność zapewnia, że będzie to jedna cząsteczka. Takie spójne multigrafy nazywane grafami molekularnymi, a przykład takiego grafu pokazano na rysunku 5.

Wzór sumaryczny związku chemicznego to natomiast nic innego jak ciąg stopni wierzchołków, gdzie stopień wierzchołka odpowiada walencyjności danego pierwiastka:



Jak podano w poprzednim rozdziale, aby móc skonstruować graf molekularny (t.j.: spójny multigraf) muszą być spełnione następujące warunki:

(4) dowolny graf: $\sum_{i=1}^n d_i = 2e, e \in \mathbb{N}$,

(5) multigraf: $\sum_{i=1}^n d_i \geq 2d_1$,

(7) spójny graf: $\sum_{i=1}^n d_i \geq 2(n-1)$.

Jak łatwo obliczyć, dla przypadku a) wszystkie te warunki są spełnione:

(4) dowolny graf: $4 + 4 + 4 + 3 + 2 + 2 + 1 + 1 + 1 + 1 + 1 = 24 = 2 \cdot 12$,

(5) multigraf: $4 + 4 + 4 + 3 + 2 + 2 + 1 + 1 + 1 + 1 + 1 = 24 \geq 8 = 2 \cdot 4$,

(7) spójny graf: $4 + 4 + 4 + 3 + 2 + 2 + 1 + 1 + 1 + 1 + 1 = 24 \geq 20 = 2(11 - 1)$;

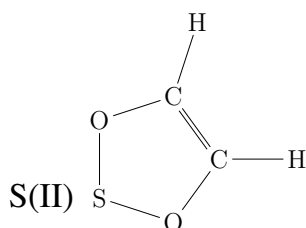
natomiast dla przypadku b) warunek (4) nie jest spełniony

(4) dowolny graf: $4 + 4 + 4 + 4 + 4 + 4 + 4 + 1 + 1 + 1 = 31 \neq 2e, e \in \mathbb{N}$,

co eliminuje przypadek C_7H_3 z puli rozwiązań.

5. Realizowalność sekwencji zbiorów stopni

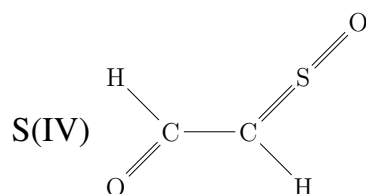
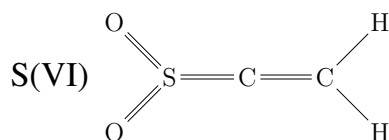
Przypadki opisane w poprzednim rozdziale pomijały jedną istotną cechę pierwiastków chemicznych. Część pierwiastków ma różne wartościowości, co w grafie molekularnym przekłada się na różne stopnie, które mogą mieć wierzchołki odpowiadające danemu pierwiastkowi. Przykładowo, dla cząsteczki o wzorze sumarycznym $C_2H_2O_2S$, w zależności od tego, czy rozpatrujemy siarkę o wartościowości II (rysunek 6), o wartościowości IV (rysunek 7) czy o wartościowości VI (rysunek 8), możemy znaleźć jako rozwiązanie różne wzory strukturalne:



Rys. 6. Wzór strukturalny cząsteczki $C_2H_2O_2S$ z atomem siarki o wartościowości II.

Uwzględnienie różnych wartościowości pierwiastków wymaga odejścia od klasycznej definicji problemu realizowalności grafu i zastąpienia pojedynczych liczb opisujących stopnie wierzchołków w grafie, zbiorami dopuszczalnych wartości dla stopni wierzchołków.

W taki sposób otrzymujemy problem realizowalności ciągu zbiorów stopni. Możemy go zdefiniować następująco:

Rys. 7. Wzór strukturalny cząsteczki $C_2H_2O_2S$ z atomem siarki o wartościowości IV.Rys. 8. Wzór strukturalny cząsteczki $C_2H_2O_2S$ z atomem siarki o wartościowości VI.

Ciąg zbiorów dodatnich liczb całkowitych D :

$$D = (D_1, D_2, \dots, D_n), \text{ gdzie } D_i = \{d_{i,1}, d_{i,2}, \dots, d_{i,k_i}\} \quad (9)$$

nazywamy realizowalnym, gdy istnieje graf o ciągu stopni R :

$$R = (r_1, r_2, \dots, r_n) \text{ takim, że } \forall_{1 \leq i \leq n} r_i \in D_i.$$

Dla podanego wcześniej przykładu cząsteczki o wzorze sumarycznym $C_2H_2O_2S$, instancja nowo zdefiniowanego problemu realizowalności ciągu zbiorów stopni wyglądałaby następująco:

$$D = (\{2, 4\}, \{2, 4\}, \{1\}, \{1\}, \{2\}, \{2\}, \{2, 4, 6\}) \quad (10)$$

Ponieważ zamiast liczb, składowymi ciągu są teraz zbiory, nie możemy mówić o ciągu nierosnącym. Wymaga to też przekonstruowania warunków sprawdzania realizowalności, które to warunki dla grafów molekularnych przedstawiają się następująco:

- dowolny graf, odpowiednik warunku (4) dla klasycznego problemu:

$$\sum_{i=1}^n r_i = 2e, e \in \mathbb{N} \quad (11)$$

- multigraf, odpowiednik warunku (5) dla klasycznego problemu:

$$\sum_{i=1}^n r_i \geq 2r_f, r_f = \max_{1 \leq x \leq n} r_x \quad (12)$$

- spójny graf, odpowiednik warunku (7) dla klasycznego problemu:

$$\sum_{i=1}^n r_i \geq 2(n-1) \quad (13)$$

Sprawdzenie, czy dany ciąg zbiorów liczb umożliwia skonstruowanie grafu o stopniach równych dowolnym liczbom ze zbiorów dopuszczalnych wartości, wymaga oczywiście podania tylko jednego wynikowego ciągu liczb R spełniających zaprezentowane warunki. Można również zauważyć, że ciąg R jest elementem iloczynu kartezjańskiego wszystkich zbiorów stopni \mathcal{R} , mianowicie:

$$R \in \mathcal{R} \text{ gdzie } \mathcal{R} = D_1 \times D_2 \times \dots \times D_n \quad (14)$$

a liczba wszystkich możliwych sekwencji R wynosi:

$$|\mathcal{R}| = \prod_{i=1}^n |D_i| \quad (15)$$

Jak widać, jest to wartość, która rośnie wykładniczo wraz ze wzrostem wielkości zbiorów wejściowych.

6. Podsumowanie

W pracy przedstawiono zagadnienie realizowalności grafów, z uwzględnieniem zbioru warunków niezbędnych, aby poszczególne wybrane klasy grafów mogły być zbudowane z ciągu liczb opisujących stopnie wierzchołków. Rozszerzono również problem tak, aby odpowiadał rzeczywistym potrzebom w chemii obliczeniowej. W przyszłości autorzy zamierzają zająć się tematem efektywnych algorytmów rozwiązujący tak zdefiniowany problem.

LITERATURA

1. Erdős P., Gallai T.: Graphs with Prescribed Degrees of Vertices. *Matematikai Lapok*, 1960, pp. 264–274.
2. Euler, L.: Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, vol. 8, 1736, pp. 128–140
3. Hakimi S.L.: On realizability of a set of integers as degrees of the vertices of a linear graph. I. *Journal of the Society for Industrial and Applied Mathematics*, vol. 10(3), 1962, pp. 496–506.
4. Havel, V.: A remark on the existence of finite graphs. *Casopis Pest. Mat.*, vol. 80, 1955, pp. 477–480.
5. Meierling, D., Volkmann, L.: A remark on degree sequences of multigraphs. *Mathematical Methods of Operations Research*, vol.69(2), 2009, pp. 369–374.
6. Wright, J. W.: The change-making problem., *Journal of the ACM (JACM)*, vol. 22(1), 1975, pp. 125–128.