3b. Grafy dwudzielne i skojarzenia

Grzegorz Kosiorowski

Uniwersytet Ekonomiczny w Krakowie

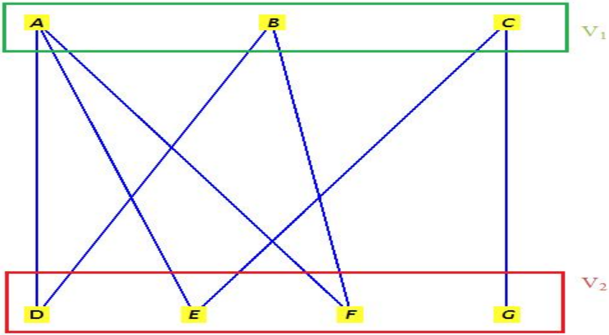
Graf dwudzielny

W tej prezentacji będziemy się zajmować tylko grafami spójnymi, nieskierowanymi i bez wag (choć poszczególne pojęcia można uogólnić na wszystkie grafy, jednak te uogólnienia nie niosą ze sobą nic szczególnie ciekawego).

Graf dwudzielny

Graf dwudzielny to graf $G = (V, E)$, w którym zbiór wierzchołków V da się podzielić na dwa rozłączne podzbiory V_1 oraz V_2 tak, by żadne dwa wierzchołki w obrębie tego samego podzbioru V_i nie były sąsiadami (czyli V_1 i V_2 są antyklikami). Dla podkreślenia takiego podziału, graf dwudzielny będziemy oznaczać przez $(V_1 \cup V_2, E)$.

Graf dwudzielny - przykład



Zazwyczaj jeśli chcemy podkreślić, że jakiś graf jest dwudzielny, rysujemy zbiór V_1 na górze, a V_2 na dole (albo V_1 po lewej, a V_2 po prawej).

Grafy dwudzielne - zastosowania

Grafy dwudzielne pozwalają modelować zjawiska, w których mamy do czynienia z obiektami dwóch typów. Najczęściej chodzi o to, by badać możliwe połączenia pomiędzy dwoma typami obiektów.

- Zbiór klientów biura matrymonialnego/portalu randkowego deklarujących heteroseksualność. Wierzchołkami grafu będą wtedy kobiety (V_1) i mężczyźni (V_2). Krawędź między dwoma wierzchołkami istnieje wtedy i tylko wtedy, gdy osoby te spełniają nawzajem swoje oczekiwania (jeśli chodzi o wybór „pary”). Oczywiście, nie ma krawędzi pomiędzy wierzchołkami tej samej grupy (z założenia o heteroseksualności).

Grafy dwudzielne - zastosowania

- W klinice transplantologicznej jako graf dwudzielny można zinterpretować zbiór dostępnych do transplantacji organów (V_1) i zbiór pacjentów, którzy oczekują na przeszczep (V_2). Połączenia między wierzchołkami z tych grup występują, gdy dany organ kwalifikuje się do przeszczepienia dla danego pacjenta z odpowiednio małym ryzykiem odrzucenia.
- Na uczelni, jedną grupą wierzchołków mogą być pracownicy dydaktyczni, a drugą - zajęcia, które trzeba poprowadzić. Połączenia między wierzchołkami z tych grup występują, gdy dane zajęcia mogą być prowadzone przez danego pracownika.

Trzecim przykładem szczegółowo zajmiemy się później, ale dwa pierwsze sugerują nam typowe zagadnienie związane z grafem dwudzielnym: znalezienie tzw. pełnego skojarzenia, czyli „znalezienia pary” dla każdego wierzchołka z grupy pierwszej w grupie drugiej.

Grafy dwudzielne - zastosowania

- Typowym przykładem w badaniu sieci społecznych jest tak zwana sieć przynależności, w której jedną grupę wierzchołków symbolizują pracownicy, a drugą - firmy dla których pracują lub pracowali.
- Znanym zagadnieniem teorii grafów jest optymalizacja sieci kolejowej, gdzie jeden zbiór wierzchołków symbolizuje zaplanowane trasy pociagów, a drugi - stacje przez które przejeżdżają.

Sprawdzanie dwudzielności

Dwudzielność i cykle

Graf jest dwudzielny wtedy i tylko wtedy, gdy każdy jego cykl ma parzystą długość.

Uzasadnienie: skoro cykl musi się kończyć w tym samym zbiorze (V_1 lub V_2), w którym się zaczynał, a każda jego krawędź przechodzi pomiędzy tymi zbiorami, to cykl musi mieć parzystą liczbę krawędzi.

Wniosek

Graf zawierający 3-klikę nigdy nie jest dwudzielny.

Sprawdzanie dwudzielności

Na podstawie twierdzenia o dwudzielności i cyklach można skonstruować prosty algorytm badania, czy graf jest dwudzielny. Złożoność tego algorytmu to $O(|V(G)|)$.

Sprawdzanie dwudzielności

Dane: Graf $G = (V(G), E(G))$.
Zmienne: V_1, V_2 - zbiory wierzchołków (na początku puste).

- I. Wybierz dowolny wierzchołek v i dołącz go do zbioru V_1.

Sprawdzanie dwudzielności

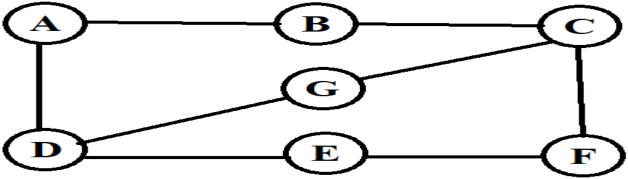
Sprawdzanie dwudzielności

- II. Dopóki zajdzie jedna z poniższych okoliczności: albo wierzchołek, który jest już w zbiorze V_i ma sąsiada w zbiorze V_i albo $V_1 \cup V_2 = V(G)$, wykonuj:
- IIa. Jeśli ostatnio dołączano wierzchołki do zbioru V_i, przypisz nieprzypisanych sąsiadów wierzchołków zbioru V_i do zbioru V_{3-i}.
- III. Jeśli powyższa pętla została przerwana dlatego, że wierzchołek, który jest już w zbiorze V_i, ma sąsiada w zbiorze V_i, graf nie jest dwudzielny.
- IV. Jeśli powyższa pętla została przerwana dlatego, że $V_1 \cup V_2 = V(G)$ (ale nie ze względu na krok III), graf jest dwudzielny.

Sprawdzanie dwudzielności - komentarze

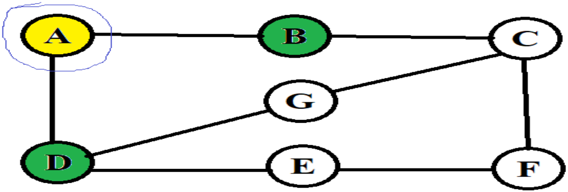
- Krok III wykonujemy przed krokiem IV, więc jeśli wyjdzie nam jednocześnie, że np. wierzchołek z V_1 ma sąsiada z V_1 i że $V_1 \cup V_2 = V(G)$, to graf nie jest dwudzielny.
- W zasadzie w kroku 2a nie musimy patrzeć na cały zbiór V_i, tylko na te wierzchołki, które przypisaliśmy tam w poprzednim obiegu pętli (ćwiczenie - zmodyfikować pseudokod, by tak było), bo sąsiedzi dawniej przypisanych wierzchołków są już przypisani.
- Jak zauważymy w przyszłości, algorytm sprawdzania dwudzielności jest szczególnym przypadkiem algorytmu przeszukiwania grafu wszerz.
- Przerwanie pętli w kroku III pozwala nam też znaleźć cykl o nieparzystej długości (łącząc najkrótsze drogi z wierzchołka startowego do wierzchołków sąsiadujących, które przerwały pętlę) co jest warunkiem równoważnym tego, że graf nie jest dwudzielny.

Sprawdzanie dwudzieloności - przykład 1



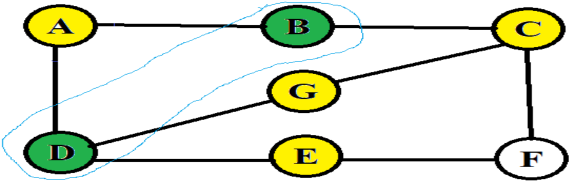
Spróbujmy algorytmicznie sprawdzić dwudzielność powyższego grafu. W kolejnych krokach na żółto będę zaznaczać wierzchołki przypisane do zbioru V_1, na zielono wierzchołki przypisane do V_2, a w niebieską obwódkę będę brał wierzchołki, których przypisywałem do jednego z tych zbiorów w ostatnim kroku.
Zaczynam od przypisania wierzchołka A do zbioru V_1.

Sprawdzanie dwudzieloności - przykład 1



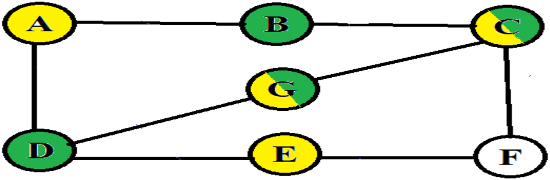
Sąsiadów zbioru V_1 (żółtego), czyli wierzchołki B i D, przypisuję do zbioru V_2 (zielonego). Nie są one swoimi sąsiadami, więc nie przerywamy algorytmu.

Sprawdzanie dwudzielności - przykład 1



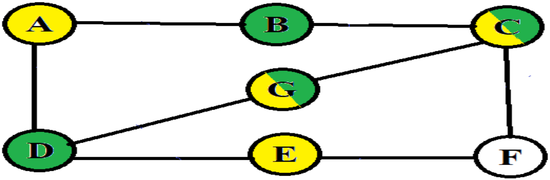
Nieprzypisanych sąsiadów ostatnio dodanych wierzchołków zbioru V_2 (B i D), czyli wierzchołki C, G i E, przypisuję do zbioru V_1 (żółtego). Nie muszę tam przypisywać wierzchołka A, bo już jest przypisany.

Sprawdzanie dwudzieloności - przykład 1



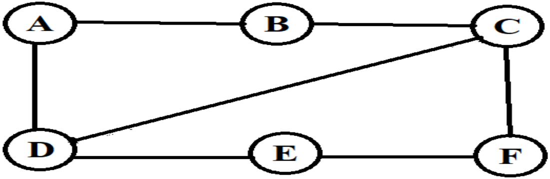
Wierzchołki E i G, które właśnie zostały przypisane do zbioru V_1 są swoimi sąsiadami. Z tego powodu algorytm się kończy i możemy powiedzieć, że przedstawiony graf nie jest dwudzielny.

Sprawdzanie dwudzieloności - przykład 1



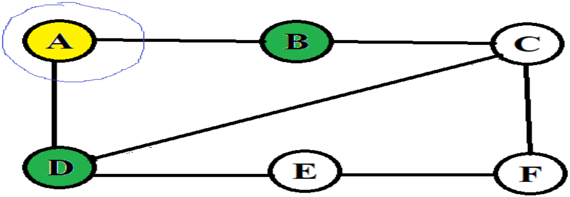
Sam konflikt wierzchołków C i G wskazuje nam nie tylko brak dwudzielności, ale też cykl, który go powoduje: wierzchołki C i G pierwszy raz były przypisane jak sąsiedzi wierzchołków odpowiednio B i D, a te z kolei były sąsiadami A. Cała ta grupa tworzy cykl ABCGDA nieparzystej długości (dokładnie długości 5).

Sprawdzanie dwudzielności - przykład 2



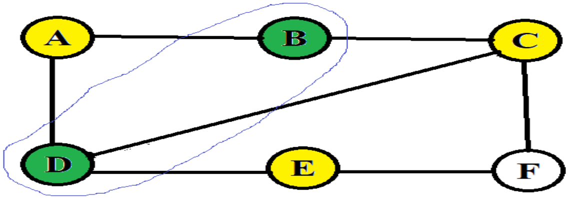
Analogicznie testujemy graf przedstawiony powyżej: zaczynam od przypisania wierzchołka A do zbioru V_1

Sprawdzanie dwudzieloności - przykład 2



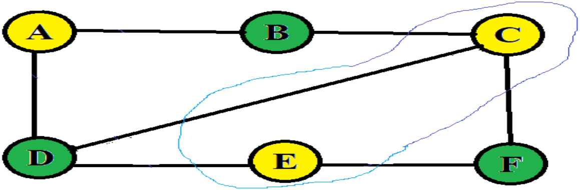
Sąsiadów zbioru V_1 (żółtego), czyli wierzchołki B i D, przypisuję do zbioru V_2 (zielonego). Nie są one swoimi sąsiadami, więc nie przerywamy algorytmu.

Sprawdzanie dwudzieloności - przykład 2



Nieprzypisanych sąsiadów ostatnio dodanych wierzchołków zbioru V_2 (B i D), czyli wierzchołki C i E, przypisuję do zbioru V_1 (żółtego).Nie są one swoimi sąsiadami (ani innych wierzchołków z V_1), więc nie przerywamy algorytmu.

Sprawdzanie dwudzieloności - przykład 2



Nieprzypisanych sąsiadów ostatnio dodanych wierzchołków zbioru V_1 (C i E), czyli wierzchołek F, przypisuję do zbioru V_2 (zielonego). Nie jest on sąsiadem żadnego wierzchołka z V_2. W ten sposób przypisaliśmy (pokolorowaliśmy) wszystkie wierzchołki grafu, więc graf jest dwudzielny z podziałem: $V_1 = \{A, C, E\}$, $V_2 = \{B, D, F\}$.

Macierz sąsiedztwa grafu dwudzielnego

Jeśli w grafie dwudzielnym $G = (V_1 \cup V_2, E)$ takim, że $|V_1| = n$ a $|V_2| = k$ ustawimy wierzchołki w takiej kolejności, że wierzchołki z V_1 poprzedzają wierzchołki z V_2 to macierz sąsiedztwa takiego grafu będzie miała postać:

$$B(G) = \begin{bmatrix} 0_{n,n} & B \\ B^T & 0_{k,k} \end{bmatrix},$$

gdzie $0_{n,n}$ i $0_{k,k}$ to macierze złożone z samych zer, wymiaru odpowiednio $n \times n$ i $k \times k$. W takiej sytuacji, macierz B (która zawiera wszystkie informacje o grafie dwudzielnym) nazywa się czasem macierzą *dwusąsiedztwa*.

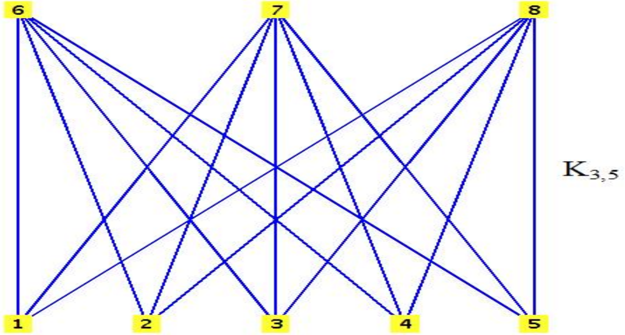
Grafy dwudzielne - przykłady

- Antykliki są dwudzielne, a kliki K_n dla $n \geqslant 3$ nie są.
- Grafy-drogi są dwudzielne (bo nie mają cykli), a grafy-cykle są dwudzielne wtedy i tylko wtedy, gdy mają parzystą liczbę wierzchołków.
- Drzewa są dwudzielne (bo nie mają cykli).
- Z grafów platońskich tylko graf sześcianu jest dwudzielny.
- Graf Petersena nie jest dwudzielny, choć nie zawiera 3-kliki (podobnie jak graf dwunastościanu).

Grafy dwudzielny pełny

Pełny graf dwudzielny

Jeśli w grafie dwudzielnym $G = (V_1 \cup V_2, E)$ pomiędzy wszystkimi parami wierzchołków należących do różnych zbiorów (V_1, V_2) istnieje krawędź, graf taki nazywamy *pełnym grafem dwudzielnym* lub kliką dwudzielną i oznaczamy $K_{n,m}$ gdzie $|V_1| = n$ i $|V_2| = m$.



Grafy dwudzielne i hamiltonowskie

Wiemy, że w ogólnym przypadku trudno stwierdzić, kiedy graf jest hamiltonowski. Dla grafów dwudzielnych sprawdzenie tego jest dużo prostsze.

Twierdzenie o cyklu Hamiltona w grafach dwudzielnych

Niech G $= (V_1 \cup V_2, E)$ będzie grafem dwudzielnym. Jeśli G ma cykl Hamiltona, to $|V_1| = |V_2|$.
Dla **pełnych** grafów dwudzielnych zachodzi też implikacja w przeciwną stronę, tj. jeśli $|V_1| = |V_2|$, to G ma cykl Hamiltona.

Skojarzenie

W grafie dwudzielnym często szukamy skojarzeń, czyli doboru w pary wierzchołków z różnych grup.

Skojarzenie

Skojarzenie w grafie $\mathrm{G} = (V, E)$ to podzbiór krawędzi $M \subset E(G)$, w którym żadne dwie $v_1v_2, u_1u_2 \in M$ nie są incydentne tym samym wierzchołkiem (czyli M to zbiór rozłącznych par elementów połączonych krawędziami).

W szczególności w grafie dwudzielnym końce krawędzi należących do skojarzenia należą do różnych zbiorów V_1 i V_2.

Wierzchołek skojarzony

Dla skojarzenia M $v \in V(G)$ jest *skojarzony*, jeśli istnieje $w \in V(G)$ taki, że krawędź $vw \in M$.

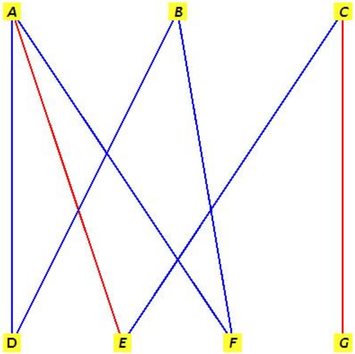
Skojarzenie pełne

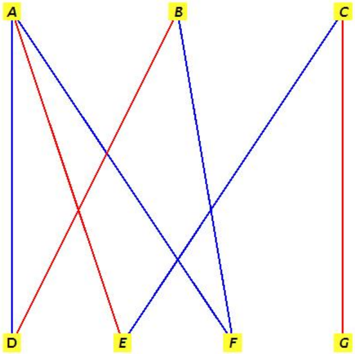
Skojarzenie pełne

Skojarzenie pełne zbioru V_1 w grafie dwudzielnym $\mathrm{G} = (V_1 \cup V_2, E)$ to skojarzenie, w którym każdy wierzchołek z V_1 jest skojarzony.

By skojarzenie pełne istniało, musi zachodzić $|V_1| \leqslant |V_2|$.

Skojarzenia - przykład



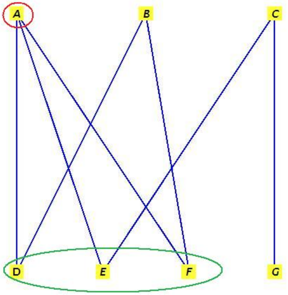


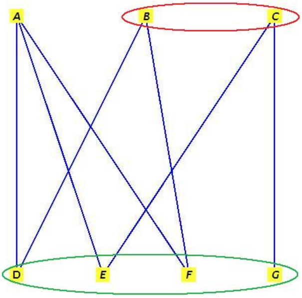
Czerwone krawędzie tworzą skojarzenie, ale nie jest ono pełne.

Czerwone krawędzie tworzą skojarzenie pełne.

Twierdzenie Halla - funkcja Φ

Funkcja Φ każdemu zbiorowi $A \subset V_1$ przyporządkowuje zbiór tych wierzchołków V_2, które są sąsiednie z przynajmniej jednym wierzchołkiem w A.





$\Phi(\{A\}) = \{D, E, F\}$.

$\Phi(\{B, C\}) = \{D, E, F, G\}$.

Twierdzenie Halla

Twierdzenie Halla

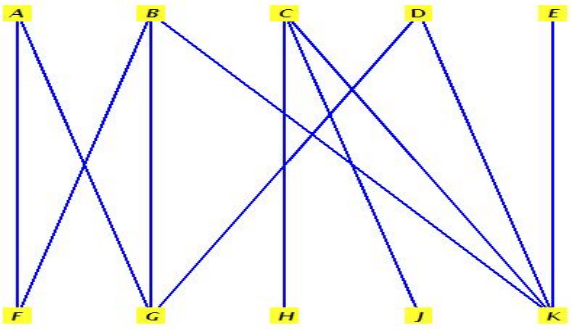
Niech $G = (V_1 \cup V_2, E)$ będzie grafem dwudzielnym. Wówczas pełne skojarzenie w G istnieje wtedy i tylko wtedy, gdy $|X| \leqslant |\Phi(X)|$ dla każdego podzbioru X zbioru V_1.

Wniosek

Grafy dwudzielne pełne mają pełne skojarzenia wtedy i tylko wtedy, gdy $|V_1| \leqslant |V_2|$.

Twierdzenie Halla-przykład

Nie istnieje pełne skojarzenie dla grafu z rysunku poniżej:



Zgodnie z twierdzeniem Halla, ten graf nie ma pełnego skojarzenia dla $V_1 = \{A, B, C, D, E\}$, $V_2 = \{F, G, H, J, K\}$, gdyż jeśli rozważymy $X = \{A, B, D, E\}$, to $\Phi(X) = \{F, G, K\}$, a zatem $|X| = 4 > 3 = |\Phi(X)|$.

Algorytm tworzenia skojarzenia pełnego

Twierdzenie Halla jest w pewnym sensie konstruktywne, gdyż jeśli jego założenie jest spełnione (czyli $|X| \leqslant |\Phi(X)|$), to skojarzenie można skonstruować. Jeden z algorytmów tworzenia skojarzenia pełnego w grafie dwudzielnym poznamy w jednej z kolejnych prezentacji jako szczególny przypadek algorytmu Edmondsa-Karpa znajdowania przepływu maksymalnego w sieci.

Grafy dwudzielne i Nobel z ekonomii

Grafy dwudzielne badał Lloyd Shapley, który w 2012 roku otrzymał nagrodę Nobla z ekonomii między innymi za modelowanie i rozwiązanie zagadnienia optymalizacyjnego zwanego *problemem stabilnego małżeństwa*. Jest to zadanie, w którym każdy „kawaler" (element zbioru V_1) i „panna na wydaniu" (element zbioru V_2) posiadają swój ranking płci przeciwnej i trzeba ich tak połączyć w „pary małżeńskie" (stworzyć skojarzenie pełne), żeby nie doszło do „pokusy zdrady", czyli sytuacji, gdy pewna para (niemałżeńska) woli siebie nawzajem od swoich partnerów w małżeństwie. Okazuje się, że bez względu na ranking indywidualnych preferencji, zawsze istnieje szczęśliwe rozwiązanie tego problemu.

Grafy dwudzielne i Nobel z ekonomii

Problem stabilnego małżeństwa, mimo nieco żartobliwej podstawowej interpretacji, ma wiele poważnych zastosowań. Za pomocą wydajnego algorytmu, który Shapley stworzył z Davidem Gale'em, można optymalizować skojarzenia:

- Przypisania kandydatów na lekarzy do szpitali (stosowany w praktyce w USA)
- Przypisywanie użytkowników do serwerów w dużych podsieciach Internetu.
- Przypisanie kierunków do kandydatów na studia.
- Hebrew Union College za pomocą tego algorytmu przypisuje kończących go rabinów do potrzebujących ich społeczności.

Algorytm Gale'a-Shapleya nie opiera się o teorię grafów, więc zainteresowanych zachęcam do samodzielnego znalezienia informacji o nim.

Zaawansowane zastosowania grafów dwudzielnych

Ponadto, grafy dwudzielne są używane w:

- Analizie kodów (szczególnie w kryptoanalizie i kompresji danych)
- Teorii sieci Petriego użytecznych w modelowaniu systemów rozproszonych.