



Uniwersytet im. A. Mickiewicza w Poznaniu

Wydział Matematyki i Informatyki

kierunek: Informatyka

Praca magisterska

# Analiza algorytmów dla wybranych problemów NP-zupełnych

Kamil Kostyszyn

Promotor:

prof. dr. hab. Zbigniew Palka

Poznań, 2014

Poznań, .....

## Oświadczenie

Ja, niżej podpisany/a .....  
student/ka Wydziału Matematyki i Informatyki Uniwersytetu im. Adama  
Mickiewicza w Poznaniu oświadczam, że przedkładaną pracę dyplomową pt:  
.....  
napisałem/am samodzielnie. Oznacza to, że przy pisaniu pracy, poza niezbęd-  
nymi konsultacjami, nie korzystałem/am z pomocy innych osób, a w szcze-  
gólności nie zlecałem/am opracowania rozprawy lub jej części innym osobom,  
ani nie odpisywałem tej rozprawy lub jej części od innych osób. Oświadczam  
również , że egzemplarz pracy dyplomowej w formie wydruku komputerowego  
jest zgodny z egzemplarzem pracy dyplomowej w formie elektronicznej. Jed-  
nocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało  
się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

.....

# Spis treści

|   |           |
|---|-----------|
| <b>Wstęp</b>  | <b>5</b>  |
| <b>1. Wprowadzenie</b>                                  | <b>6</b>  |
| <b>2. Pierwszy problem NP</b>                           | <b>11</b> |
| 2.1. Wstęp . . . . .                                    | 11        |
| 2.2. Charakterystyka klasy NP-zupełnej . . . . .        | 12        |
| 2.3. Problem spełnialności układów logicznych . . . . . | 12        |
| <b>3. Problem pokrycia wierzchołkowego</b>              | <b>20</b> |
| 3.1. Wstęp . . . . .                                    | 20        |
| 3.2. Przedstawienie problemu . . . . .                  | 20        |
| 3.3. Dowód przez redukcje . . . . .                     | 21        |
| 3.4. Algorytm aproksymacyjny . . . . .                  | 23        |
| 3.4.1. Algorytm VERTEX-COVER-DEGREES-APPROX . . .       | 23        |
| 3.4.2. Algorytm VERTEX-COVER-EDGES-APPROX . . . .       | 26        |
| <b>4. Problem kliki</b>                                 | <b>29</b> |
| 4.1. Wstęp . . . . .                                    | 29        |
| 4.2. Przedstawienie problemu . . . . .                  | 29        |
| 4.3. Dowód przez redukcje . . . . .                     | 30        |
| 4.4. Algorytm aproksymacyjny . . . . .                  | 31        |

---

|  |           |
|--|-----------|
| <b>5. Problem komiwojażera</b>         | <b>33</b> |
| 5.1. Wstęp . . . . .                   | 33        |
| 5.2. Przedstawienie problemu . . . . . | 33        |
| 5.3. Algorytm aproksymacyjny . . . . . | 34        |
| <b>6. Problem sumy podzbioru</b>       | <b>35</b> |
| 6.1. Wstęp . . . . .                   | 35        |
| 6.2. Przedstawienie problemu . . . . . | 35        |
| 6.3. Dowód przez redukcje . . . . .    | 36        |
| 6.4. Algorytm aproksymacyjny . . . . . | 39        |
| <b>Spis ilustracji</b>                 | <b>41</b> |
| <b>Spis tabel</b>                      | <b>42</b> |
| <b>Bibliografia</b>                    | <b>42</b> |

# Wstęp

Zasadniczym celem pracy dyplomowej przygotowywanej przez studenta jest:

- wykazanie się umiejętnością formułowania i rozwiązywania problemów wiążących się z programem odbytych studiów,
- wykazanie się znajomością metod i sposobów prowadzenia analizy oraz redakcyjnego przygotowania pracy w oparciu o umiejętności nabyte w czasie studiów.

Praca dyplomowa musi być samodzielnym opracowaniem autorstwa studenta, przygotowanym przy pomocy promotora. Student, jako autor ponosi pełną odpowiedzialność z tytułu oryginalności i rzetelności zaprezentowanego materiału i powinien uwzględniać wszelkie prawa i dobre obyczaje w tym zakresie.

Tematyka pracy powinna znacząco wykraczać poza materiał omówiony w trakcie studiów; zakłada samodzielny wkład autora w postaci np.: implementacji, opracowania algorytmu, samodzielnego porównania, oceny i analizy istniejących rozwiązań.

# Rozdział 1

## Wprowadzenie

W celu dobrego przedstawienia problemów i algorytmów będących tematem pracy, niezbędne jest przytoczenie podstawowych pojęć oraz zagadnień odnoszących się NP-zupełności.

Bazowym pojęciem, które w pierwszej kolejności należy przedstawić jest sam termin "problemu". Według definicji w [1] problem abstrakcyjny  $Q$  jest relacją dwuargumentową, określoną na zbiorze  $I$  egzemplarzy problemu oraz zbiorze  $S$  rozwiązań problemu. Egzemplarzem jest struktura matematyczna, natomiast podzbiór jej elementów składowych jest rozwiązaniem. Przenosząc to na przykład dla problemu MAXIMUM-INDEPENDENT-EDGE-SET (czyli znalezienia maksymalnego skojarzenia) egzemplarzem jest graf  $G$  złożony ze zbioru wierzchołków  $V$  oraz krawędzi  $E$ . Zatem maksymalny zbiór krawędzi niemających wspólnych wierzchołków jest rozwiązaniem tego problemu. Skoro problem jest relacją, a nie funkcją, możliwa jest sytuacja, że dla jednego egzemplarza może istnieć wiele rozwiązań.

W przypadku poruszania się w obrębie teorii NP-zupełności takie definiowanie problemu jest zbyt ogólne, dlatego najlepiej jest ograniczyć się do

problemów decyzyjnych. W tej formie znana jest instancja problemu oraz wynik będący odpowiedzią. W przypadku INDEPENDENT-EDGE-SET egzemplarz ma postać  $i = \langle G, k \rangle$ , a rozumiany jest następująco: w grafie  $G$  istnieje skojarzenie o mocy co najmniej  $k$ . Rozwiązaniem jest podzbiór zbioru krawędzi  $E' \in E$  o mocy co najmniej  $k$ . Zadanie polega na zweryfikowaniu, czy podany wynik jest rzeczywiście prawdziwy. Nie trudno zauważyć, że dla tak postawionego problemu zbiór odpowiedzi w takim przypadku ogranicza się do dwóch wartości prawdy lub fałszu. Pozornie łatwe podejście w celu rozpatrywania problemu nie jest wcale takie proste, gdy dochodzi do jego realizacji.

Inną formą podejścia do problemu jest podejście optymalizacyjne. Podobnie jak w problemie decyzyjnym na wstępie otrzymuje się instancję problemu, jednak nie występuje sugerowane rozwiązanie. Głównym zadaniem jest teraz znalezienie najlepszego wyniku, w zależności czy zbiór wynikowy ma się minimalizować (w celu zmniejszania pewnych kosztów) czy też chodzi o maksymalizowanie (w celu zwiększenia pewnych zysków). Odpowiedzą tak sformułowanemu problemowi będzie moc zbioru wynikowego, czasami również elementy tego zbioru. Mimo, iż optymalizacyjne podejścia wydają się przy realizacji trudniejsze, można je łatwo zastąpić formą decyzyjną w bardzo prosty sposób. Należy obrać początkową trywialną wartość, a następnie zmieniać ją o jednostkową wartość do momentu, aż przestanie istnieć rozwiązanie dla danej wartości. W przypadku minimalizacji za trywialną wartość obieramy maksymalną wartość dla zbioru, którą później dekrementujemy. Natomiast dla maksymalizacji trywialną wartością startową jest zero, które zwiększamy, aż do momentu w którym przestanie istnieć rozwiązanie.

Większa część problemów należy do klasy złożoności P, dlatego rzadko podejmuje się problemy wybiegające poza tę klasę. Problemy będące tema-

tem tej pracy wybiegają dalej i dotyczą sławnej hipotezy: "czy  $P \neq NP$ ", dlatego obszar który podejmuje praca odnosi się do klas problemów  $P$ ,  $NP$  i  $NPC$ .

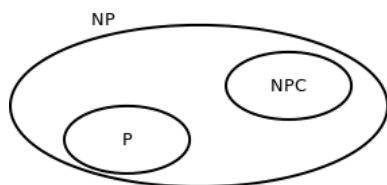
Najbardziej powszechna klasa  $P$  zawiera ten zbiór problemów, dla których istnieje algorytm znajdujący rozwiązanie w czasie wielomianowym. Czas działania algorytmu oznacza się jako funkcję  $f$ , która będzie działać ze zbioru oznaczającego rozmiar danych o ustalonym kodowaniu na zbiór odpowiadający liczbie operacji elementarnych potrzebnych do zakończenia działania algorytmu. Przy tak określonej funkcji, jeśli można ją ograniczyć z góry przez dowolną funkcję wielomianową  $O(nk)$ , to wtedy problem należy do klasy złożoności  $P$ .

O ile  $P \neq NP$ , to klasa  $NP$  zawiera te problemy, które są "weryfikowalne" w czasie wielomianowym. Przez słowo "weryfikowalne" rozumie się, że istnieje takie świadectwo pozytywne (zwane inaczej dowodem), dzięki któremu można sprawdzić, czy otrzymane rozwiązanie jest poprawne. Klasa  $P \subseteq NP$ . Zatem, skoro można rozwiązać problem w czasie wielomianowym, to tym bardziej można w wielomianowym czasie sprawdzić pozytywność świadectwa.

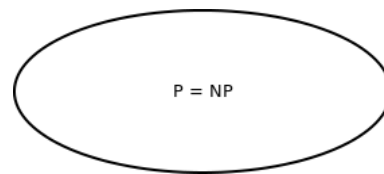
Klasa  $NPC$  jest tym, co różni klasę  $P$  i  $NP$ . W jej skład wchodzi te problemy, które należą do  $NP$  a nie należą do  $P$ . Nasuwa się pytanie: "Czy w ogóle istnieją takie problemy z klasy  $NP$ , których nie można rozwiązać w czasie wielomianowym?". Na obecny stan rzeczy odpowiedź brzmi: nie wiadomo. W czasie powstawania pracy nikt dotąd nie udowodnił, że wszystkie problemy  $Q \in NP$  są rozwiązywalne w czasie wielomianowym, ani też nie pokazał, że istnieje taki problem z klasy  $NPC$ , który nie jest rozwiązany w czasie wielomianowym. Tematem tej pracy są właśnie te problemy o których wiadomo, że są weryfikowalne w czasie wielomianowym, a nie są rozwiązywalne w czasie



wielomianowym.



**Rys. 1.1.**  $P \neq NP \wedge NPC \neq \emptyset$



**Rys. 1.2.**  $P = NP \wedge NPC = \emptyset$

Szczególną własnością problemów NPC jest fakt, iż jeśli choć jeden problem znajdzie się w klasie P, to wszystkie problemy z klasy NPC będą należały do klasy P. Pokazuje to, że jeśli uda się znaleźć algorytm wielomianowy rozwiązujący jakikolwiek problem z klasy NPC to oznaczałoby, że  $P = NP$ . Związanie ze sobą tych wszystkich problemów NPC zawdzięcza się redukcji wielomianowej. Polega ona na tym, że rozwiązanie pewnego problemu można sprowadzić do rozwiązania innego. Odbywa się to za pośrednictwem algorytmu redukcji, który jest w stanie przeprowadzić każdą instancję problemu  $A$  do postaci instancji problemu  $B$  w czasie wielomianowym. Aby to zobrazować, można posłużyć się trywialnym przykładem, w którym problem  $Q$  obliczenia długości wektora w  $\mathbb{R}^2$  zredukuje się do problemu  $Q'$  obliczenia długości wektora w  $\mathbb{R}^3$ . Mając tak zdefiniowane problemy potrzebny jest jeszcze algorytm redukcji. Nasz algorytm będzie rozszerzał wektor o trzecią pozycję i wpisywał w to miejsce 0. Teraz każdy wektor z  $\mathbb{R}^2$  możemy przedstawić w  $\mathbb{R}^3$ , a następnie obliczyć jego długość. Widać, że długość wektora  $[x_1, x_2, 0]$  odpowiada długości wektora  $[x_1, x_2]$ . Pokazuje to, że problem  $Q$  można zredukować do problemu  $Q'$ .

Mimo, iż tak wiele wiadomo o problemach NP-zupełnych, to nadal sprawiają one wyzwanie. O ile algorytmy o złożoności większej od wielomianowej potrafią znaleźć optymalne rozwiązanie, o tyle należy mieć pewność, że dane będą z odpowiednio wąskiego zakresu. Jeśli jednak zakres danych jest zbyt

duży, pozostaje jedynie heurystyczne poszukiwanie wyniku dla mocno sprecyzowanego problemu lub zadowolenie się wynikiem zbliżonym do optymalnego w oparciu o algorytm aproksymacyjny. W przypadku heurystycznego podejścia nie można za wiele powiedzieć, gdyż wszystko zależy od indywidualnej potrzeby, o którą musi zadbać algorytm oraz improwizacji, na którą można sobie pozwolić.

Obecnie najbardziej popularną metodą jest wykorzystanie algorytmów aproksymacyjnych ze względu na ich wielomianową złożoność. Jakość takiego algorytmu zależy od znalezionej wartości  $C$  względem rozwiązania optymalnego  $C^*$ . Współczynnik aproksymacji  $1 \leq \rho(n)$  jest wyznacznikiem jakości i ogranicza z góry stosunek kosztów:  $\max(\frac{C}{C^*}, \frac{C^*}{C}) \leq \rho(n)$

Wzór ten uwzględnia przypadek zarówno maksymalizacji, jak i minimalizacji określając ile razy otrzymane rozwiązanie jest gorsze od optymalnego. Dla maksymalizacji mamy  $0 \leq C \leq C^*$  czyli wartość znalezionej wartości jest nie większa od rozwiązania optymalnego. Natomiast w problemie minimalizacji  $0 \leq C^* \leq C$  wartość znalezionej wartości jest nie mniejsza od rozwiązania optymalnego.

brakuje jakiegoś zakończenia, podsumowania, bo jest urwane w połowie

## Rozdział 2

# Pierwszy problem NP

### 2.1. Wstęp

Przechodząc do rozważań nad konkretnymi problemami z klasy NP-zupełnej należy wprawdzie skupić się nad znaczeniem tej klasy. W teorii NP-zupełności fundamentalne znaczenie ma metoda redukcji pomiędzy dwoma problemami. Podstawą wykorzystania tej metody jest założenie, że klasa ta jest niepusta dzięki czemu można pokazać redukcję problemu, który da się wprowadzić do tej klasy względem problemu znajdującego się już w klasie NP-zupełnej. Tu pojawia się problem: „jak tego dokonać jeśli w początkowych rozważaniach klasa NPC jest pusta?”. Nie można posłużyć się redukowalnością, aby wprowadzić pierwszy „pierwotny” problem. Niezbędne okaże się skorzystanie wprost z definicji, by w ten oto sposób zasilić klasę NP-zupełną o pierwszy problem. Później w oparciu o to będzie można już redukcją wprowadzać pozostałe problemy.

## 2.2. Charakterystyka klasy NP-zupełnej

Definicja NP-zupełności mówi, że, aby problem należał do klasy NPC musi spełniać dwa warunki, które można sprowadzić do następujących własności:

1. świadectwo problemu jest weryfikowalne w czasie wielomianowym
2. jest co najmniej tak "trudny", jak każdy problem z NP.

Własność 1. nie wydaje się względnie trudnym warunkiem, natomiast warunek 2. może już przysporzyć wielu trudności. Własność ta mówi o tym, że każdy problem z klasy NP można zredukować do powyższego problemu. Tu zaczynają się komplikacje, ponieważ należy określić metody redukcji każdego problemu z klasy NP (nie ważne jaki on by nie był) do naszego problemu "pierwotnego". O ile pozornie to zadanie wydaje się nie do zrealizowania o tyle istnieje na to bardzo ciekawe rozwiązanie. Rozwiązanie jest ukryte w komputerach, a dokładnie na relacji pomiędzy modelem danych, a jego niższą warstwą fizycznych danych. Oczywiście jest to, że algorytmy implementujemy w pamięci komputera, zaś ten jest w stanie przeprowadzić jego działanie, tak, aby otrzymać wynik. Natomiast cała idea tego przetwarzania jest realizowana na gigantycznych sieciach logicznych. Przyjrzeć się należy zatem problemowi spełnialności układów logicznych jako potencjalnemu kandydatowi na rolę "pierwotnego" problemu w klasie NPC.

## 2.3. Problem spełnialności układów logicznych

Problem spełnialności układów logicznych (CIRCUIT-SAT) to układ  $C$  zbudowany ze stałej liczby wejść binarnych oraz bramek logicznych. W pod-

stawowym charakterze wyróżnia się bramki AND, OR I NOT. Układ nazywamy spełnialnym, jeśli istnieje takie wartościowanie wejść sprawiające, iż po wykonaniu wszystkich operacji logicznych z bramek na wyjściu otrzyma się prawdę.

Rozwiązanie tego problemu polega na sprawdzeniu wszystkich możliwych odpowiedzi, a dokładniej w obecnym tu przypadku na zbadaniu wszystkich możliwych wartościowań wejścia układu. Zakładając że układ składa się z  $k$  wejść od razu widać, że należy sprawdzić  $2^k$  możliwych przypadków niezależnie od rozmiaru egzemplarza danych wejściowych. Sprawia to, że czas działania takiego algorytmu jest wykładniczy, czyli wyższy od wielomianowego.

Spróbujmy wykazać teraz, że  $\text{CIRCUIT-SAT} \in \text{NPC}$ . W tym celu wpierw skupmy się na pierwszej własności i udowodnimy, że pozytywne świadectwo można weryfikować w czasie wielomianowym, czyli w rezultacie  $\text{CIRCUIT-SAT} \in \text{NP}$ .

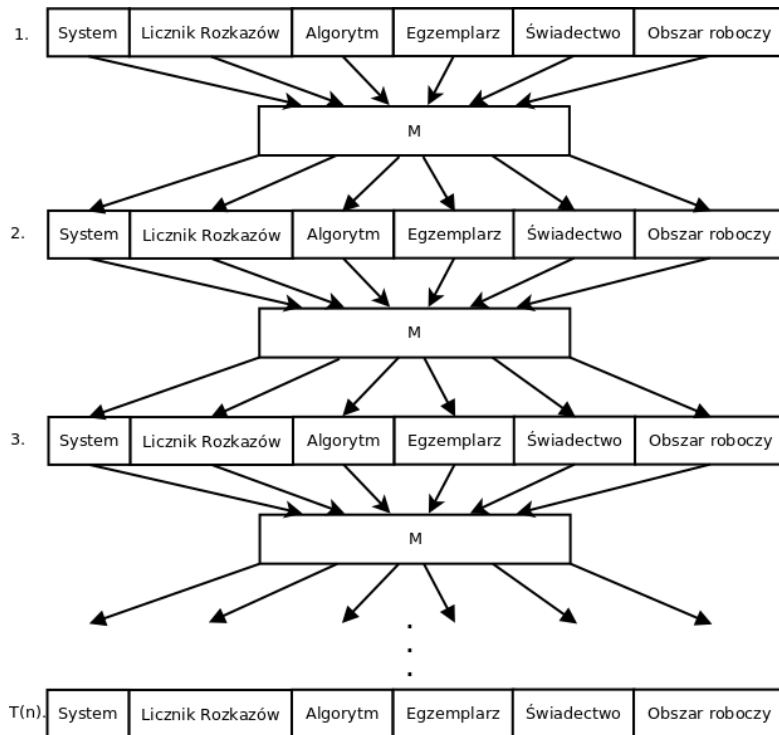
**Lemat 1.**  $\text{CIRCUIT-SAT} \in \text{NP}$

Dowód: Algorytm weryfikujący problem  $\text{CIRCUIT-SAT}$  na wejściu oczekuje dwóch parametrów. Pierwszy to egzemplarz  $E$  pełnej specyfikacji układu logicznego  $C$ , natomiast drugi parametr to świadectwo  $y$  zawierające binarne wartości przypisane do konkretnych wejść. Pierwszym krokiem algorytmu jest sprowadzenie układu logicznego  $C$  do zapisu w formie odwrotnej notacji polskiej. W tym procesie bramki logiczne zostają przekształcone na operandy logiczne w następujący sposób: AND jako koniunkcja, OR jako alternatywa, a NOT negacja. Przy rozsądnej specyfikacji układu logicznego krok ten można wykonać w sposób liniowy. Następną czynnością jest wykonanie tych operacji względem wartości logicznych z świadectwa  $y$ . Do uzyskania poprawnej kolejności wykonywanych operacji na bitach wykorzystuje się

strukturę kolejki i stosu. Po zrealizowaniu wszystkich operacji logicznych na końcu stosu znajduje się wynik będący jedynym elementem stosu. Rozwiązanie jest poprawne, gdyż zmiana realizacji problemu uległa jedynie forma z bramek logicznych na operacje logiczne, a wynik jest jedynie konsekwencją realizacji działań logicznych. Jeśli w wyniku na końcu uzyskaliśmy prawdę (wartość logiczną 1) oznacza to, że rzeczywiście świadectwo  $y$  spełnia układ logiczny  $C$  z egzemplarza  $E$ . W przeciwnym przypadku, kiedy wynikiem byłby fałsz dostajemy informację, że to świadectwo nie świadczy o spełnialności układu logicznego  $C$ , zarazem nie wykluczając jego spełnialności. W przypadku, gdy układ  $C$  nie byłby spełnialny, to nie ważne dla jakiego wartościowania nigdy nie uzyskamy prawdy na końcu przy poprawnym wykonaniu wszystkich operacji logicznych. W kwestii złożoności czasowej, łatwo zauważyć, że algorytm składa się z dwóch kroków, z czego każdy ma złożoność liniową. Sprawia to, że całość algorytmu realizowana jest w czasie liniowym, a zatem i wielomianowym, co kończy dowód i ukazuje, że  $\text{CIRCUIT-SAT} \in \text{NP}$ .

Po wykazaniu, iż  $\text{CIRCUIT-SAT} \in \text{NP}$  następnym krokiem jest pokazanie, że  $\text{CIRCUIT-SAT} \in \text{NP-trudny}$ , czyli w praktyce odpowiada temu realizacji własności 2. Przebieg tego dowodu będzie wykorzystywał ściśle zasadę funkcjonowania komputera. Celem jest, by problem  $\text{CIRCUIT-SAT}$  realizował schemat obliczeniowy dowolnego algorytmu rozwiązującego problem z klasy NP na komputerze. Wpierw należy bliżej nakreślić sytuację, w której realizowany będzie dowód. W pierwszej kolejności zwróćmy uwagę na fakt, że komputer przede wszystkim wykorzystuje pamięć i procesor. Pamięć służy do przechowywania wartości. Natomiast procesor składa się z maszyny, którą można oznaczyć jako  $M$  zawierającej sieć bramek logicznych, dzięki której jest w stanie wykonywać operacje logiczne, a co dalej za tym idzie wykonywać instrukcje. Teraz należy podzielić pamięć na obszary ze względu na przy-

datność do realizacji algorytmu. Zdecydowanie największym obszarem, który zajmuje pamięć są dane związane z obsługą komputera umożliwiające jego działanie. Dodatkowo wyznacza się obszar licznika rozkazów, za pośrednictwem którego znajduje się adres następnej komendy do wykonania. Obszary najbardziej przydatne w tym momencie to obszar pamięci, w której jest zapisany algorytm  $A$ , obszar na egzemplarz problemu  $x$  oraz świadectwo  $y$ . W praktyce potrzeba jeszcze jednego obszaru, którego przeznaczeniem będzie przechowywanie obliczeń częściowych, które będzie wykorzystywał algorytm  $A$  oraz po wykonaniu wszystkich operacji w pewnej ustalonej komórce tego obszaru będzie znajdował się nasz wynik. Konfiguracją komputera nazywa się wszystkie te obszary. Działanie komputera należy opisać w następujący sposób: najpierw działa w sposób wcześniej zaprezentowany, następnie wykonuje obliczenia w oparciu o konfigurację, a rezultat obliczeń zostaje zapisany jako nowa konfiguracja. W rezultacie jedna konfiguracja zostaje przekształcona w kolejną i tak bez końca, dopóki urządzenie funkcjonuje.



Rys. 2.1. Sekwencja konfiguracji

Teraz nastąpi sprawdzenie czy problem spełnialności układów logicznych jest NP-trudny:

**Lemat 2.** *CIRCUIT-SAT*  $\in$  NP-trudny.

Dowód: Aby wykazać, iż CIRCUIT-SAT jest NP-trudny należy pokazać, że każdy problem z klasy NP można zredukować do CIRCUIT-SAT. W tym celu wyznacza się kolejne etapy, które to przedstawia:

1. Zdefiniowanie algorytmu redukcji.

Dla każdego problemu  $Q$  z klasy NP można określić dwuparametrowy algorytm  $A$ , którego zadaniem jest zweryfikowanie świadectwa. W tym przypadku jest możliwość wykazania zależności, iż jeśli świadectwo jest poprawne, to można je przekształcić na odpowiednią postać binarną,



która z kolei jest spełniającym wartościowaniem układu logicznego  $C$ . Również w drugą stronę, jeśli znajdzie się takie wartościowanie, które spełnia układ logiczny  $C$  to będzie go można przeformułować na postać świadectwa, które będzie poprawnie weryfikowane przez algorytm  $A$ .

Skupmy się teraz na konstrukcji, która zamienia proces wykonywania algorytmu  $A$  na tworzenie układu logicznego  $C$ . Jak zostało wcześniej wspomniane, wszystkie dane zawierające informacje na temat komputera i algorytmu  $A$  znajdują się w pamięci komputera i wspólnie stanowią pierwszą konfigurację. Następnie maszyna  $M$  przeprowadza kolejne kroki algorytmu, przechodząc z jednej konfiguracji w drugą, aż do momentu zakończenia działania algorytmu  $A$ . W rezultacie liczba konfiguracji jest dokładnie równa  $T(n)$ , czyli liczbie kroków jakie potrzebuje algorytm  $A$  do uzyskania wyniku końcowego. Jak widać, przez cały ten czas maszyna  $M$  pozostaje niezmienna, a jej postępowanie zależy jedynie od konfiguracji i przez ten czas jak będzie się zmieniała konfiguracja tak również maszyna  $M$  będzie zmieniała różne obszary pamięci.

Aby sprowadzić zasadę działania wykonywania algorytmu  $A$  w komputerze do struktury problemu CIRCUIT-SAT, to konfiguracja  $c_0$  jest przewodami wejściowymi do układu logicznego. Następnie działa na to struktura logiczna maszyny  $M$ . W wyniku takiego działania otrzymać można wyjście, które jest zapisywane jako konfiguracja  $c_1$ . Jednak w niniejszym schemacie nie można zastosować takiego rozwiązania, w wyniku czego po prostu przedłuża się otrzymane ścieżki tak, aby były kolejnym wejściem do powielonej maszyny  $M$ . Kontynuując to w ten sposób, otrzymujemy tyle obszarów bramek logicznych, ile wynosi liczba kroków algorytmu  $A$ . W końcowej fazie musimy wypuścić na wyjście

ścieżkę z ostatniej konfiguracji  $c_{T(n)}$  obszaru pamięci przeznaczanej na wynik algorytmu. Jest to wyjście układu zawierającego odpowiedź na pytanie czy świadectwo jest prawdziwe. W ten oto sposób niejako "sklejone zostały" ze sobą przejścia przez wszystkie konfiguracje, tak by na końcu całego układu znalazła się końcowa konfiguracja  $c_T(n)$ .

## 2. Analiza poprawności

W ten sposób skonstruowany układ  $C$ , jak można zauważyć, jest spełnialny tylko wtedy, gdy świadectwo jest pozytywnie zweryfikowane przez algorytm  $A$ , a nie może być inaczej, gdyż układ  $C$  symuluje obliczeniową naturę komputera. Z drugiej strony jeśli na wejście układu  $C$  przedstawimy pewne wartościowanie które powoduje spełnialność układu oznacza to, że ta interpretacja binarna danych przekłada się na poprawne świadectwo, gdyż algorytm  $A$  poprawnie weryfikuje to świadectwo, a jak wiadomo  $A$  jest poprawnym algorytmem. Zatem algorytm redukcji  $F$  jest skonstruowany w sposób poprawny.

## 3. Analiza złożoności

Ostatnim elementem dowodu jest wykazanie wielomianowej złożoności algorytmu redukcyjnego  $F$  względem rozmiaru danych wejściowych  $n = |x|$ . Niewątpliwie rozmiar potrzebny na dane związane ze stanem maszyny, obszarem licznika rozkazów i implementacją maszyny  $M$  jest zależny od typu komputera, który symuluje. Mimo to jest stały za względu na rozmiar danych przychodzących. Podobnie rozmiar algorytmu  $A$  jest wartością stałą. Kolejnym czynnikiem są dane wejściowe egzemplarza  $x$  oraz świadectwa  $y$ . Świadectwo będące podzbiorem egzemplarza jest nie większe rozmiarem od egzemplarza  $x$ , zaś rozmiar egzemplarza wynosi  $n$ . Ostatnim elementem przekładającym

się na strukturze algorytmu  $F$  jest obszar roboczy, w którym znajdzie się wynik oraz który służy na pomocnicze obliczenia algorytmu złożoności wielomianowej  $A$ , czyli nie może być większy niż wielomianowa od rozmiaru danych. Zatem cała konstrukcja jednego kroku algorytmu redukcyjnego  $F$  jest wielomianowa.

Teraz należy zwrócić uwagę na liczbę kroków, którą potrzebuje zrealizować algorytm  $F$ . Cała idea konstrukcji polega na symulowaniu działania algorytmu  $A$  przez algorytm  $F$ . Zatem każda pojedyncza instrukcja algorytmu  $A$  przekłada się na jeden wielomianowy krok algorytmu  $F$ . Skoro więc wielomian od wielomianu nadal jest wielomianem, to całość działania algorytmu  $F$  jest ostatecznie wielomianowa.

#### 4. Podsumowanie

Podsumowując udało się pokazać, że algorytm redukcji  $F$  w poprawny sposób symuluje działanie każdego algorytmu dającego przedstawić się w strukturze komputerowej. Ponadto jeśli złożoność czasowa symulowanego algorytmu nie jest większa od wielomianowej to złożoność czasowa algorytmu redukcyjnego  $F$  też jest wielomianowa. Oznacza to, że problem CIRCUIT-SAT jest tak samo trudny jak każdy problem należący do klasy NP, czyli CIRCUIT-SAT  $\in$  NP-trudny.

**Twierdzenie 3.** *CIRCUIT-SAT  $\in$  NP-zupełny*

Dowód: Reasumując, to, że CIRCUIT-SAT  $\in$  NP i CIRCUIT-SAT  $\in$  NP-trudny otrzymujemy, że CIRCUIT-SAT  $\in$  NP-zupełny.

Wynika z tego wniosek, że o ile klasa NP-zupełna jest nie pusta (czyli ciągle nie rozstrzygnięte przez ludzkość pytanie czy NP = P) to niewątpliwie CIRCUIT-SAT należy do tej klasy.

## Rozdział 3

# Problem pokrycia wierzchołkowego

### 3.1. Wstęp

Tematem rozdziału będzie przedstawienie problemu minimalnego pokrycia wierzchołkowego. Nieformalnie pokryciem wierzchołkowym nazywa się zbiór wierzchołków, w którym każda krawędź z grafu ma co najmniej jeden koniec. Samo określenie problemu wydaje się być niezwykle proste, jednak rozwiązanie tego problemu jest NP-zupełne. Formalnie problem można zdefiniować następująco.

### 3.2. Przedstawienie problemu

Problem minimalnego pokrycia wierzchołkowego (VERTEX-COVER) polega na znalezieniu w grafie nieskierowanym  $G = (V, E)$  takiego podzbioru  $V' \subseteq V$ , że dla każdej krawędzi  $\{v, w\} \in E : v \in V' \vee w \in V'$ .

**Twierdzenie 4.**  $VERTEX-COVER \in NP\text{-}zupelnych$ .

### 3.3. Dowód przez redukcje

Dowód:

W pierwszej kolejności należy pokazać, że  $VERTEX-COVER \in NP$ . Wejściem dla świadectwa jest graf  $G = (V, E)$  oraz podzbiór wierzchołków  $V' \subseteq V$ . Zadanie polega na sprawdzeniu, czy każda krawędź ma co najmniej jeden koniec w zbiorze  $V'$ . Jeśli tak jest, to jest to poprawne pokrycie wierzchołkowe. Nie trudno zauważyć, że czas potrzebny na to jest liniowy względem liczby krawędzi. Natomiast jeśli graf jest stosunkowo gęsty, to złożoność jest kwadratowa względem liczby wierzchołków. Tak czy inaczej jest wielomianowa w ujęciu rozmiaru danych, a co za tym idzie  $VERTEX-COVER \in NP$ .

Teraz należy pokazać, że  $VERTEX-COVER \in NP\text{-}trudnych$ . W tym celu trzeba skorzystać z redukcji  $VERTEX-COVER$  do  $CLIQUE$ . Dopełniając graf posiadający klikę otrzymuje się graf, którego pokryciem wierzchołkowym jest wszystko poza kliką. Dopełnienie grafu należy rozumieć jako odwrócenie wszystkich krawędzi. Precyzując, wszystkie krawędzie, które należały do grafu  $G$  nie należą do grafu dopełnionego  $\bar{G}$ , a te krawędzie, które nie istniały w grafie  $G$  istnieją w dopełnionym grafie  $\bar{G}$ . W dalsze rozważania wykorzystują fakt, że suma krawędzi z grafu  $G$  i jego dopełnienia  $\bar{G}$  daje graf pełny. Jak również to, że dopełnienie grafu dopełnionego daje graf pierwotny. Wszystko to razem oznacza, że problem maksymalnej kliki jest dualny z problemu minimalnego pokrycia wierzchołkowego. Zatem redukcja opierać się będzie na tezie: graf  $G$  zawiera klikę na wierzchołkach  $V'$  wtedy i tylko wtedy, gdy  $V - V'$  jest pokryciem wierzchołkowym dopełnionego grafu  $\bar{G}$ . W celu udowodnienia tej tezy zostanie to rozbite na dwie implikacje.

W pierwszej kolejności zostanie pokazane, że jeśli graf  $G$  zawiera klikę na wierzchołkach  $V'$ , to  $V - V'$  jest pokryciem wierzchołkowym dopełnionego grafu  $\bar{G}$ . Dla wierzchołków  $u, v \in V$ , pomiędzy którymi nie istnieje krawędź w  $G$  zachodzi, że co najmniej jeden z nich nie należy do kliky, czyli należy do zbioru  $V - V'$ . Jako, że w grafie dopełnionym  $\bar{G}$  pomiędzy wierzchołkami  $v$  i  $u$  istnieje krawędź to zbiór  $V - V'$  pokryje ją. Jako, że krawędź została wybrana w sposób dowolny to rozumowanie jest poprawne dla każdej krawędzi z  $\bar{G}$ , co kończy dowodzenie pierwszej implikacji.

Teraz pozostało pokazać, że: jeśli  $V - V'$  jest pokryciem wierzchołkowym dopełnionego grafu  $\bar{G}$  to graf  $G$  zawiera klikę na  $V'$  wierzchołkach. Należy zwrócić uwagę na to, że zbiór  $V'$  zawiera wszystkie te wierzchołki, które nie należą do pokrycia wierzchołkowego, zatem w grafie dopełnionym  $\bar{G}$  między nimi nie występuje żadna krawędź. Zachodzi tak, ponieważ jeśli którąś parę łączyłaby krawędź to co najmniej jeden z tych wierzchołków musiałby należeć do pokrycia wierzchołkowego. Skoro w grafie dopełnionym  $\bar{G}$  między zbiorem tych wierzchołków nie występuje krawędź to w grafie  $G$  między wszystkimi tymi wierzchołkami będą występowały krawędzie, co z kolei oznacza, że na zbiorze  $V'$  mamy klikę. To stwierdzenie kończy drugą implikację, ostatecznie wykazując prawdziwość tezy postawionej wcześniej.

Ostatnim krokiem jest określenie złożoności sprowadzenia VERTEX-COVER do CLIQUE. Przy takim postępowaniu należy wpierw stworzyć graf pełny na identycznej liczbie wierzchołków, by później iterując po każdej krawędzi pozostawiać jedynie te krawędzie, które nie należały do pierwotnego grafu. Na koniec klika będzie znajdowała się na tych wierzchołkach, które nie należały do pokrycia wierzchołkowego. Jak widać algorytm ten jest złożoności kwadratowej ze względu na ilość wierzchołków, czyli wielomianowa, co kończy cały dowód wykazując że VERTEX-COVER  $\in$  NP-zupełny.

## 3.4. Algorytm aproksymacyjny

### 3.4.1. Algorytm VERTEX-COVER-DEGREES-APPROX

Można odnieść wrażenie, że metoda zachłanna będzie przynosiła wyniki bardzo zbliżone do optymalnego, zatem można spróbować w ten oto sposób podejść do problemu. Możemy wyróżnić dwa sposoby podejścia zachłannego. Pierwszy to taki, który pokrywałby wierzchołki o najwyższym stopniu, idąc za myślą, że taki wierzchołek pokryje najwięcej krawędzi. Zatem naturalnie powinien wchodzić w skład pokrycia wierzchołkowego. Drugie podejście zachłanne to takie, w którym można przyjrzeć się końcowi (brzegom) grafu, czyli takim nie izolowanym wierzchołkom, które mają możliwie najmniejszy stopień wierzchołka. Tutaj naturalną jest myśl, że jeśli należy pokryć te incydentne krawędzie, to należałoby do pokrycia wybrać wierzchołek z drugiego końca krawędzi. Obie metody w swojej idei wykorzystują najlepsze rozwiązanie lokalne co wydaje się prowadzić do dobrego przybliżenia problemu globalnego. Niniejszy fragment pracy magisterskiej skupia się na drugim podejściu dlatego, że to pierwsze jest popularnie rozważane w literaturze. Oto algorytm realizujący to podejście:

---

**Algorithm 1** VERTEX-COVER-DEGREES-APPROX

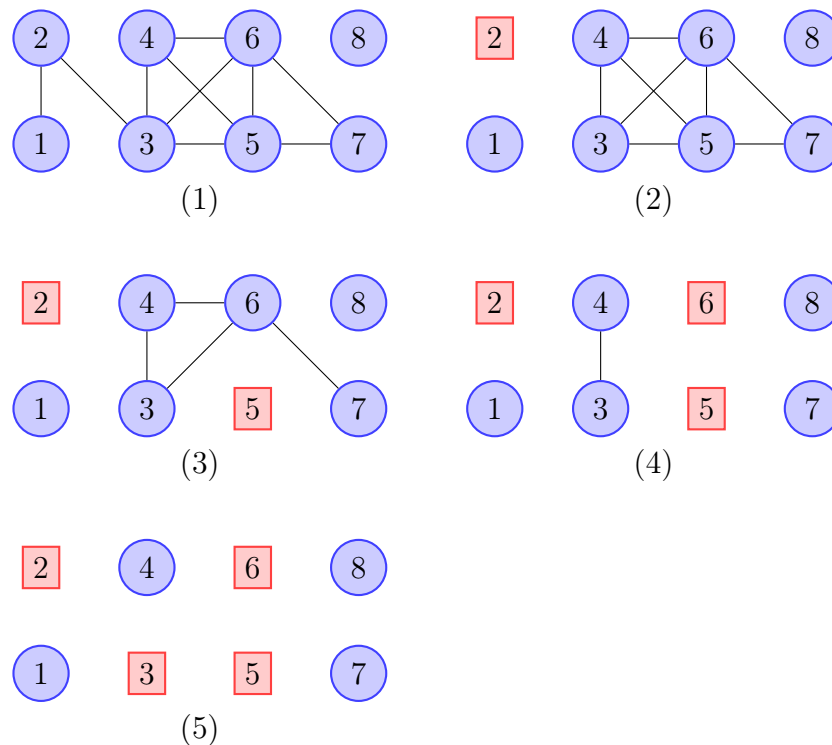
---

```
1: function MYPROCEDURE
2:    $C \leftarrow \emptyset$ 
3:    $E' \leftarrow G.E$ 
4:   while  $E' \neq \emptyset$  do
5:     weź dowolny nieizolowany wierzchołek  $v$  o najmniejszym stopniu w grafie  $G$ 
6:     weź incydentny wierzchołek  $u$  z wierzchołkiem  $v$  o największym stopniu
7:      $C \leftarrow C \cup \{u\}$ 
8:     usuń z  $E'$  incydentne krawędzie z  $u$ 
9:   return  $C$ 
```

---

Rysunek 3.1 przedstawia kolejne kroki wykonywania algorytmu VERTEX-

COVER-DEGREES-APPROX. W pierwszym kroku wierzchołek o najmniejszym stopniu to 1 a jedyny incydentny z nim wierzchołek, który wchodzi do pokrycia to 2. W drugim kroku wierzchołek o najmniejszym stopniu to 7 z nim incydentne są 5 i 6. Oba są tego samego stopnia, zatem do pokrycia wejdzie dowolny z nich w tym przypadku 5. Dalej sytuacja jest klarowna i za pośrednictwem 7 do pokrycia wchodzi 6. Ostatni przebieg to pojedyncza krawędź w rezultacie do pokrycia wchodzi dowolnie wybrany 3. Reasumując pokrycie wierzchołkowe dla tego grafu wynosi 4 (wierzchołki: 2,3,5,6) i jest to optymalne rozwiązanie dla tego grafu.

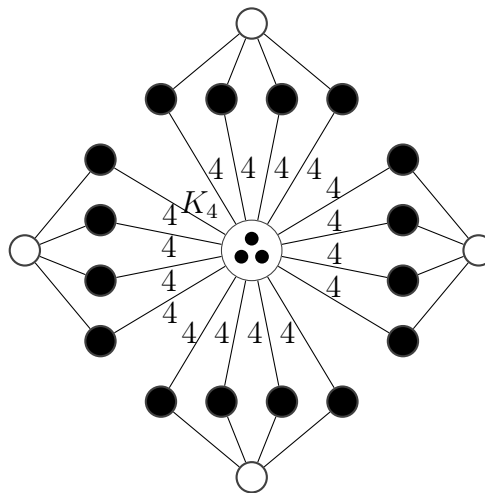


**Rys. 3.1.** Przykład wykonania algorytmu VERTEX-COVER-DEGREES-APPROX

Zastanowić się należy teraz nad konstrukcją grafu, który będzie wpuklał tezę, że pokrywanie wierzchołków wyższego stopnia daje gorsze rozwiązanie. Bardzo dobrym przykładem jest graf, który nazwać można płatkami śnie-



gu(snowflake). Taki graf składa się z rdzenia i wielu ramion. Poniżej mieści się schemat takiego grafu, którego rdzeniem jest graf  $K_4$  i składa się z 4 ramion. Możemy zauważyć, że liczba ramion dla takiego rdzenia może zbiegać do nieskończoności, natomiast liczba wierzchołków incydentnych do rdzenia nie może być większa niż najniższy stopień wierzchołków, do których przylegają. Wraz ze wzrostem rdzenia grafu można do niego przyłączyć większe ramiona.



Rys. 3.2. Graf snowflake(4,4)

Wierzchołki oznaczone kolorem czarnym zostały nominowane do pokrycia wierzchołkowego grafu plus dodatkowo 3 wierzchołki z  $K_4$  (nie można mniej dla grafu pełnego). Natomiast można łatwo zauważyć, że jeśli do pokrycia zaliczymy białe wierzchołki plus dodatkowo klikę, to również pokryje to graf przy znacznie mniejszej liczbie wierzchołków. Teraz oszacować można współczynnik aproksymacji dla tej rodziny grafów gdzie  $k$  będzie rozmiarem rdzenia a  $m$  liczbą ramion. Łatwo zauważyć, że przypadek pesymistyczny to taki, gdzie liczbę i rozmiar ramion będziemy maksymalizowali, natomiast rozmiar rdzenia będziemy minimalizowali. Aby własność doboru nieodpowiednich wierzchołków była zachowana to liczba wierzchołków połączonych

z rdzeniem musi być nie większa od rozmiaru rdzenia. Minimalny graf o jak największej liczbie krawędzi to graf pełny. W skład pokrycia optymalnego będzie wchodził zawsze cały rdzeń grafu, czyli  $k$  wierzchołków a dodatkowo  $m$  wierzchołków z każdego ramienia po tym wierzchołku, który nie jest połączony z rdzeniem. W przypadku rozwiązania proponowanego przez algorytm jest  $k - 1$  wierzchołków z rdzenia oraz  $m * k$  wierzchołków ze wszystkich ramion. Wynik jest następujący:  $\frac{C}{C^*} = \frac{k-1+m*k}{k+m} \leq m * k$ . Zatem zależność jest liniowa, co świadczy o słabym współczynniku aproksymacji.

### 3.4.2. Algorytm VERTEX-COVER-EDGES-APPROX

Można też inaczej podejść do problemu. Proponowane rozwiązanie to wybranie losowej krawędzi, dodanie dwój jej końców do pokrycia wierzchołkowego, po czym usunąć wszystkie incydentne z nimi krawędzie. Oto algorytm realizujący to podejście:

---

#### Algorithm 2 VERTEX-COVER-EDGES-APPROX

---

```

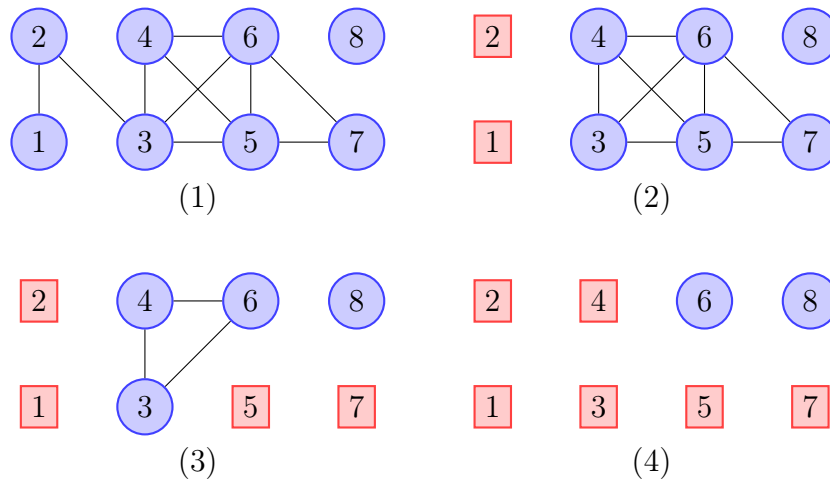
1: function MYPROCEDURE
2:    $C \leftarrow \emptyset$ 
3:    $E' \leftarrow G.E$ 
4:   while  $E' \neq \emptyset$  do
5:     weź dowolna krawędź  $(u, v) \in E'$ 
6:      $C \leftarrow C \cup \{u, v\}$ 
7:     usuń z  $E'$  incydentne krawędzie z  $u$  lub  $v$ 
8:   return  $C$ 

```

---

Teraz wykonać można ten sam algorytm dla poprzedniego grafu. W pierwszym kroku w dowolny sposób wybrać należy krawędź  $\{1,2\}$ , dodać te wierzchołki do pokrycia i usunąć wszystkie incydentne krawędzie. W drugim kroku analogicznie została wybrana dowolna krawędź  $\{5,7\}$ , wierzchołki dodane do pokrycia, a następnie usunięte wszystkie incydentne krawędzie. Całe rozumowanie jest poprawne, ponieważ dla każdego z ramion  $G_i$  algorytm wybrał krawędź z ramienia, która nie jest połączona z rdzeniem. W tym przypadku nie ma to znaczenia, ponieważ w tym momencie nie ma już krawędzi łączących rdzeń z ramieniem.

wanie jest powtarzane, aż w rezultacie do pokrycia wierzchołkowego zostały przypisane wierzchołki: 1,2,3,4,5,7.



**Rys. 3.3.** Przykład wykonania algorytmu VERTEX-COVER-EDGES-APPROX

Porównując te dwa podejścia można odnieść wrażenie, że ostatni algorytm nie jest godny uwagi, lecz można spróbować określić współczynnik aproksymacji dla algorytmu VERTEX-COVER-EDGES-APPROX.

**Twierdzenie 5.** *VERTEX-COVER-DEGREES-APPROX jest 2 aproksymowalne [do po polsku]*

Dowód: W pierwszej kolejności pokazać można, że zbiór z wyjścia algorytmu VERTEX-COVER-EDGES-APPROX jest rzeczywiście pokryciem wierzchołkowym. Pętla z 4. linii wykonuje się do momentu, aż wszystkie krawędzie nie zostaną pokryte. Wewnątrz w linii 5. do pokrycia dodaje się oba końce losowo wybranej krawędzi, następnie usuwając te krawędzie, które pokryte są przez dodane wierzchołki. W rezultacie zostają jedynie usunięte te krawędzie które są pokryte, a do zbioru wyjściowego wchodzi te wierzchołki, przy których krawędź jeszcze nie została pokryta. W tym procesie otrzymany zbiór wierzchołków jest prawidłowym pokryciem wierzchołkowym dla

zadanego grafu.

Co do jakości tego rozwiązania wygląda to następująco. Algorytm VERTEX-COVER-EDGES-APPROX w linii 5. dodaje dwa wierzchołki z jednej krawędzi do pokrycia wierzchołkowego. W linii 7. zostają usunięte wszystkie incydentne krawędzie w rezultacie. Tak postępując można zaobserwować, że żadne losowo wybrane krawędzie nie mają wspólnego wierzchołka. Zatem zbiór zawiera tyle wierzchołków, ile wynosi podwojona liczba losowych krawędzi. Warto zauważyć, że pokrycie optymalne musi analogicznie pokryć wszystkie krawędzie, w tym również te, które zostały wybrane losowo. W rezultacie pokrycie generowane przez algorytm VERTEX-COVER-EDGES-APPROX jest nie więcej niż dwukrotnie większe od optymalnego.

Choć algorytm VERTEX-COVER-EDGES-APPROX posiada znacząco niższy współczynnik aproksymacji od algorytmu VERTEX-COVER-DEGREES-APPROX to widać, że swoją przewagę zawdzięcza jedynie szczególnym przypadkom. Wnioskując algorytm VERTEX-COVER-EDGES-APPROX będzie zwracał gorsze rozwiązanie od algorytmu VERTEX-COVER-DEGREES-APPROX w średnim przypadku.

# Rozdział 4

## Problem kliki

### 4.1. Wstęp

Kliką nazywamy taki podzbiór wierzchołków  $V'$  ze zbioru  $V$ , że występujące w nim wszystkie wierzchołki są połączone między sobą krawędziami. Jeżeli moc zbioru  $|V'| = k$  to mówimy że klika jest rozmiaru  $k$ , oczywiście każdy graf pełny  $K_n$  jest kliką rozmiaru  $n$ .

### 4.2. Przedstawienie problemu

Problem maksymalnej kliki (CLIQUE) polega na znalezieniu w grafie nieskierowanym  $G$  kliki o największym rozmiarze.

**Twierdzenie 6.** *CLIQUE*  $\in$  NP-zupełnych.

### 4.3. Dowód przez redukcje

Dowód: Sprawdzenie, że problemu CLIQUE należy do klasy NP odbywa się poprzez sprawdzenie podzbioru  $V'$  [zawarty bądź równy]  $V$  będącego świadectwem. Jeśli w podzbiorze  $V'$  między każdą parą wierzchołków występuje krawędź to oznacza to że tworzą one klikę. Taki algorytm sprawdzający wykonuje dokładnie  $teta - oznaczona(n) = n^2$  operacji dla każdego wierzchołka, czyli jest złożoności wielomianowej. Dowodzi to że CLIQUE [należy do] NP.

Kolejnym krokiem dowodu jest pokazanie że problem CLIQUE jest Np-trudny. Aby tego dokonać wykorzystam fakt że problem 3-CNF jest NP-zupełny. Problem 3-CNF jest formułą logiczną zawierającą klauzule połączone koniunkcją, natomiast każda klauzula składa się z dokładnie trzech literałów połączonych alternatywą. Zadanie polega na znalezieniu takich wartościowań literałów dzięki którym formuła jest spełniana.

Przedstawię teraz sposób sprowadzenia problemu 3-CNP do problemu znalezienia klik w grafie. Każda klauzula będzie określała oddzielną grupę wierzchołków, a w każdej z nich będą się znajdowały dokładnie trzy wierzchołki odpowiadające konkretnym literałom. Krawędzią połączymy te wierzchołki które spełniają jednocześnie dwa warunki: - wierzchołki nie należą do wspólnej klauzuli - literały odpowiadające wierzchołkom nie są sprzeczne. Szukać będziemy takiej klik która połączy ze sobą wszystkie klauzule. Jak łatwo zauważyć transformacja egzemplarza 3-CNF na egzemplarz grafowy można bez większych problemów wykonać w czasie wielomianowym.

Przyjrzyjmy się teraz związkowi jaki tworzą między sobą te dwa zadania. Nie połączyliśmy ze sobą wierzchołków w tej samej klauzuli co oznacza że wystarczy że znajdziemy tylko jeden literał prawdziwy, co sprawi że dzięki wewnętrznym alternatywą cała klauzula będzie prawdziwa. Łączymy ze so-

bą tylko nie sprzeczne literały co oznacza że zmienne nie mogą jednocześnie posiadać dwóch różnych wartości. Zbierając to wszystko do siebie chcę pokazać że znalezienie kliku na tych klauzulach oznacza znalezienie niesprzecznych wartości literałów które w rezultacie spełniają całą formułę. Natomiast z drugiej strony znalezienie wartościowań literałów pozwalających uczynić formułę spełnialną oznacza że w takiej konstrukcji grafu musi istnieć klika wielkości liczbie klauzul w formule.

---

Świadectwem dla problemu CLIQUE jest podzbiór wierzchołków  $V'$  zawierających klikę. Aby sprawdzić autentyczność tego świadectwa wystarczy jedynie sprawdzić, czy dla każdej pary wierzchołków z tego podzbioru istnieje krawędź. To sprawdzenie ma złożoność kwadratową, czyli wielomianową.

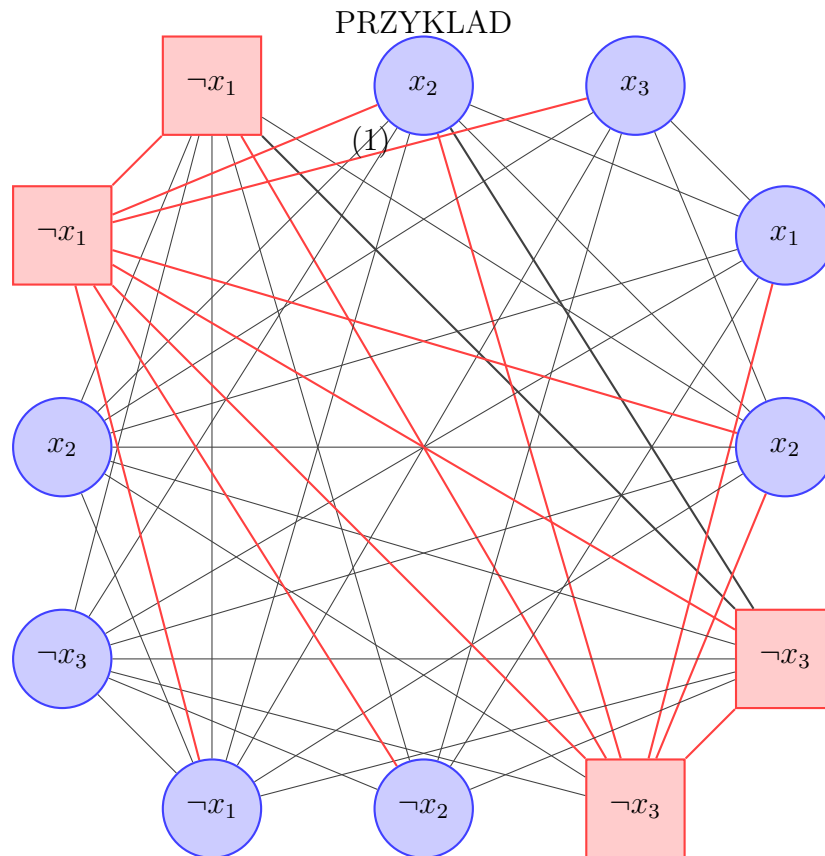
Aby pokazać, że CLIQUE jest NP-trudny podeprzemy się faktem że problem 3-CNF jest NP-trudny i przedstawię redukcję 3-CNF  $[ \leq ]$  p CLIQUE.  $[kolki = C1 \text{ i } C2 \text{ i } \dots Ck]$  oznaczmy jak formułę problemu 3-CNF składającą się z  $r = 1, 2, \dots k$  klauzul. Na każdą klauzulę składają się dokładnie 3 literały  $l^r, 1, l^r, 2, l^r, 3$ . Skonstruujemy graf  $G$  z formuły  $[kolko]$  w ten sposób że formuła jest spełniana wtedy i tylko wtedy, gdy  $G$  zawiera klikę o rozmiarze  $k$ .

... redukcja z 3CNF

koniec

## 4.4. Algorytm aproksymacyjny

Dzięki temu że problem CLIQUE jest dualny do problemu VERTEX-COVER, to algorytmy z poprzedniego rozdziału również po odpowiednim przeformułowaniu wejścia zwracają rozwiązania problemu CLIQUE



**Rys. 4.1.** Przykład wykonania algorytmu VERTEX-COVER-EDGES-APPROX. Przykład przedstawia model prezentujący formułę 3-CNF [ozkreską = C1 i C2 i C3 i C4], C1 =  $x_1$  lub  $x_2$  lub  $x_3$ , C2 =  $x_1$  lub  $x_2$  lub  $\neg x_3$ , C3 =  $\neg x_1$  lub  $\neg x_2$  lub  $\neg x_3$ , C4 =  $\neg x_1$  lub  $x_2$  lub  $\neg x_3$  w postaci grafu G problemu CLIQUE. Dla tak określonej formuły wartościowanie  $x_1 = 0$ ,  $x_3 = 0$  spełnia formułę, natomiast zmienna  $x_2$  pozostaje dowolna. Na grafie jest to zobrazowane czerwonymi wierzchołkami zawierającymi klikę.

zakomentarzowany tekst



# Rozdział 5

## Problem komiwojażera

### 5.1. Wstęp

### 5.2. Przedstawienie problemu

Problem komiwojażera (TSP) polega na wyznaczeniu cyklu w grafie pełnym  $G_n$  z zastrzeżeniem, że sumaryczny koszt krawędzi w cyklu powinien być minimalny i zawierać wszystkie  $n$  wierzchołków.

**Twierdzenie 7.**  $TSP \in NP\text{-}zupelnych$ .

Dowód:

Tradycyjnie dowód zaczyna się od pokazania, że TSP *in* NP. Świadectwem dla problemu TSP jest ciąg wszystkich  $n$  wierzchołków grafu  $G_n$ , natomiast weryfikacja świadectwa polega na sprawdzeniu czy każdy wierzchołek z grafu występuje dokładnie raz oraz czy między wierzchołkami występują krawędzie. Złożoność sprawdzenia tych warunków jest wielomianowa.

Teraz druga część dowodu w której zostanie pokazane że TSP jest NP-trudny. W tym celu posłużymy się problemem HAM-CYCLE, czyli znalezie-

nia cyklu w grafie po wszystkich wierzchołkach. Czyli zostanie przedstawiona redukcja  $\text{HAM-CYCLE} \leq \text{TSP}$ . Oznaczmy  $G = (V, E)$  jako egzemplarz problemu HAM-CYCLE, natomiast  $G' = (V, E')$  będzie egzemplarzem problemu TSP. Graf  $G'$  jest konstruowany jako graf pełny na tym samym zbiorze wierzchołków. Koszt krawędzi wynosi 0 dla tych krawędzi które należały do grafu  $G$ , natomiast dla pozostałych wynosi on 1. Złożoność skonstruowania takiego grafu jest wielomianowa.

Założmy że graf  $G$  zawiera cykl Hamiltona i pokażmy że wtedy w grafie  $G'$  musi istnieć marszruta o całkowitym koszcie 0. Skoro graf  $G$  zawiera cykl Hamiltona, to wszystkie te krawędzie w  $G'$  mają koszt 0, zatem ten cykl jest marszrutą o całkowitym koszcie 0. Teraz z kolei założmy że  $G'$  zawiera marszrutę o koszcie 0 i pokażmy że wtedy w grafie  $G$  musi istnieć cykl Hamiltona. Zatem jeżeli w grafie występuje marszruta o całkowitym koszcie 0 to wszystkie te krawędzie muszą istnieć w  $G$ , dalej te wszystkie krawędzie tworzą ścieżkę przechodzącą przez wszystkie wierzchołki a początkowy wierzchołek tej ścieżki jest jednocześnie wierzchołkiem końcowym, czyli tworzy ona cykl Hamiltona. Wnioskowanie to kończy dowód.

### 5.3. Algorytm aproksymacyjny

# Rozdział 6

## Problem sumy podzbioru

### 6.1. Wstęp

Ostatnim problem który zostanie poruszony w tej pracy jest problem sumy podzbioru. Problem ten jest natury numerycznej. W tym problemie należy skonstruować szukaną sumę w oparciu o liczby z dostępnego zbioru. Jest to kolejny problem który można łatwo określić, a przez to podejrzewać że istnieje prosty algorytm rozwiązujący ten problem. Jednak zostanie przedstawione jest to problem NP-zupełny.

### 6.2. Przedstawienie problemu

Problem sumy podzbioru(SUBSET-SUM) polega na wyznaczeniu podzbioru  $S' \subseteq S$  dla którego suma elementów jest maksymalna, ale nie większa od zadanej wartości  $t$ .

**Twierdzenie 8.** *SUBSET-SUM  $\in$  NP-zupełnych.*

## 6.3. Dowód przez redukcje

Dowód: Problem SUBSET-SUM  $\in$  NP, ponieważ dla zadanego zbioru liczb  $S' \subseteq S$  możemy w czasie liniowym zsumować je a otrzymaną sumę porównać z wartością  $t$ .

Następnie zostanie pokazana redukcja z problemu VERTEX-COVER do problemu SUBSET-SUM, czyli VERTEX-COVER  $\leq_p$  SUBSET-SUM. Co udowodni że SUBSET-SUM  $\in$  NPC.

Dla dowolnego egzemplarza  $\langle G, k \rangle$  problemu pokrycia wierzchołkowego zostanie przedstawiona konstrukcja odpowiadającego mu egzemplarza  $\langle S, t \rangle$  problemu sumy podzbioru.

Dla dowolnego grafu  $G(V, E)$  skonstruujemy macierz  $T$  wymiaru  $|E+1| * |V + E|$  której wiersze będą odzwierciedlały kolejne elementy zbioru  $S$  zakodowane w układzie czwórkowym. Macierz  $T$  składa się z dwóch grup  $x_i$  oraz  $y_i$ .

Grupę  $x_i$  determinują wierzchołki grafu i jest ona równoliczna ze zbiorem wierzchołków  $|V|$ . Grupę tę wypełnia macierz incydencji  $M$  grafu  $G$  od strony najmniej znaczącego bitu, natomiast kolumna z najbardziej znaczącym jest wypełniona przez wartość 1. [wzór na element]

Grupa  $y_i$  to macierz jednostkowa  $I_{|E|}$  począwszy od najmniej znaczącego bitu, a wypełniona wartością 0 na pozycji najbardziej znaczącego bitu. [wzór na element]

Ostatnim elementem konstrukcji jest wyznaczenie wartości  $t$ :

$$t = k4^{|E|} + \text{sum}(\text{odj} = 0, \text{do}|E| - 1)2 * 4^i.$$

Elementy składające się na sumę  $t$  z grupy  $x_i$  to te  $x_i$  którym odpowiadają wierzchołki wchodzące do pokrycia wierzchołkowego. Natomiast z grupy  $y_i$  są to te  $y_i$  odpowiadające krawędzią dla których dokładnie jeden z końców znajduje się w pokryciu wierzchołkowym.

Przy tak określonej redukcji zostanie pokazane, że graf  $G$  będzie miał pokrycie wierzchołkowe  $k$  wtedy i tylko wtedy, gdy zbiór  $S$  będzie zawierał część elementów sumujących się do  $t$ .

W pierwszej kolejności założmy, że graf  $G$  ma pokrycie wierzchołkowe  $v_1, v_2, v_3 \dots$  *wiekszyodst*  $v_k \subset V$ . Na podstawie grafu  $G$  otrzymujemy macierz w oparciu o przedstawioną wyżej konstrukcję. Z grupy  $X$  wybrane są elementy  $x_i$  odpowiadające pokryciu wierzchołkowemu, natomiast z grupy  $Y$  te  $y_i$  których odpowiadające krawędzie mają koniec w dokładnie jednym wierzchołku pokrycia wierzchołkowemu. Udowodnimy dalej że po wybraniu dokładnie tych elementów suma ich będzie równa zdefiniowanej przez nas wartości  $t$ . W grupie  $X$  wartość najbardziej znaczącego bitu wynosi dokładnie  $k$  elementów, ponieważ one zostały wybrane z pokrycia wierzchołkowego. Natomiast w grupie  $Y$  na najbardziej znaczącym bicie występują zawsze wartości 0. Pozostałe bity zdeterminowane są przez krawędzie w grafie, a wartość  $t$  w każdej kolumnie wynosi 2. Każda kolumna w macierzy odpowiada krawędzi grafu, zatem w każdej kolumnie w grupie  $X$  będą dokładnie dwie wartości 1. Jednak uwzględniając tylko wierzchołki z pokrycia, to w każdej kolumnie może występować co najwyżej raz wartość 1. Spoglądając szerzej jeśli krawędź jest pokryta wyłącznie przez jeden wierzchołek to w kolumnie znajduje się wyłącznie jedna wartość 1. W takim przypadku do zbioru  $S'$  dodajemy jednak element z grupy  $Y$ , który w tej pozycji ma wartość 1. W innym przypadku jak krawędź jest pokryta przez dwa wierzchołki to w grupie  $X$  w odpowiedniej kolumnie występują dokładnie dwie wartości 1. Dla takiego przypadku nie zostaje dodany odpowiadający mu element z grupy  $Y$ . Podsumowując w kolumnie najbardziej suma bitów zawsze wynosi  $k$ , a w pozostałych kolumnach suma zawsze wynosi 2. Oznacza to że elementy ze zbioru  $S'$  zawsze będą się w ten sposób sumowały do wartości  $t$ .

Teraz udowodnijmy przeciwne wnioskowanie i załóżmy że istnieje taki podzbiór  $S'$  zawierający  $S$ , którego elementy sumują się do wartości  $t$ . Wartość  $t$  zdefiniowana jest w ten sposób że składa się z wartości 2 na każdym bicie w kodzie czwórkowym poza najbardziej znaczącym bitem. Najbardziej znaczący bit oznaczmy jako  $m$ . Dzięki temu można odczytać ile krawędzi znajduje się w grafie  $G$ . Dalej podzielmy podzbiór  $S'$  na dwie grupy. Grupa  $Y$  zawiera podzbiory  $S'$ , których suma wynosi wartość potęgi liczby 4. Pozostałe elementy stanowią podzbiory których suma należy do grupy  $X$ , podzbiory charakteryzuje to że ich najbardziej znaczący bit jest związany z taką potęgą liczby 4 co wartość  $m$ . Dzięki temu możemy podzielić zbiór  $S'$  w taki sposób  $x_{i1}, x_{i2}, \dots, x_{im}, y_{j1}, y_{j2}, \dots, y_{jp}$ . Dodatkowo zbiór  $S$  może zawierać dodatkowe elementy które wypełnią wiersze macierzy z konstrukcji. W rezultacie w każdej kolumnie poza najbardziej znaczącym bitem muszą znajdować się trzy wartości 1. Dwie w grupie  $X$  i jedna w grupie  $Y$ , tak by krawędź w grafie miała dokładnie 2 końce. Takie ograniczenia jednoznacznie określają macierz z konstrukcji tym samym opisując dokładnie postać grafu  $G$ . Elementy  $x_{i1}, x_{i2}, \dots, x_{im}$  wyznaczają pokrycie wierzchołkowe.

Pokażemy że  $m = k$  oraz że elementą  $x_i$  będą odpowiadały wierzchołki  $v_i$  tworzące pokrycie wierzchołkowe w grafie  $G$ .

W takim przypadku aby można było otrzymać poprawny graf w grupie  $X$  musi znajdować się co najmniej  $m$  elementów, tak by suma najbardziej znaczących bitów odpowiadała wartości  $t$ . W przypadku mniej znaczących bitów w każdej kolumnie muszą znajdować się trzy wartości 1.

W tak zdefiniowanym zbiorze odpowiadającym zasadą konstrukcji wszystkie elementy możemy przedstawić w systemie czwórkowym

Ciąg dalszy nastąpi... że redukcja jest wielomianowa.

Zbiór  $S'$  tworzą liczby z pokrycia wierzchołkowego, zatem każda krawędź

jest pokryta przez 2 lub 1 wierzchołek.

Wybrane wiersze sumowały się do wartości 2 w kolumnie

Liczbę  $t$  tworzą dwie grupy  $X$  i  $Y$ .

## 6.4. Algorytm aproksymacyjny

# Spis rysunków

|  |    |
|--|----|
| 1.1. $P \neq NP \wedge NPC \neq \emptyset$ . . . . .   | 9  |
| 1.2. $P = NP \wedge NPC = \emptyset$ . . . . .   | 9  |
| 2.1. Sekwencja konfiguracji . . . . .  | 16 |
| 3.1. Przykład wykonania algorytmu VERTEX-COVER-DEGREES-<br>APPROX . . . . .  | 24 |
| 3.2. Graf snowflake(4,4) . . . . .   | 25 |
| 3.3. Przykład wykonania algorytmu VERTEX-COVER-EDGES-<br>APPROX . . . . .  | 27 |
| 4.1. Przykład wykonania algorytmu VERTEX-COVER-EDGES-<br>APPROX. Przykład przedstawia model prezentujący formułę<br>3-CNF [ozkreską = C1 i C2 i C3 i C4], C1 = nx1 lub x2 lub<br>x3, C2= x1 lub x2 lub nx3, C3 = nx1 lub nx2 lub nx3, C4 =<br>nx1 lub x2 lub nx3 w postaci grafu G problemu CLIQUE. Dla<br>tak określonej formuły wartościowanie $x1 = 0$ , $x3 = 0$ spełnia<br>formułę, natomiast zmienna $x2$ pozostaje dowolna. Na grafie<br>jest to zobrazowane czerwonymi wierzchołkami zawierającymi<br>klikę. . . . . | 32 |



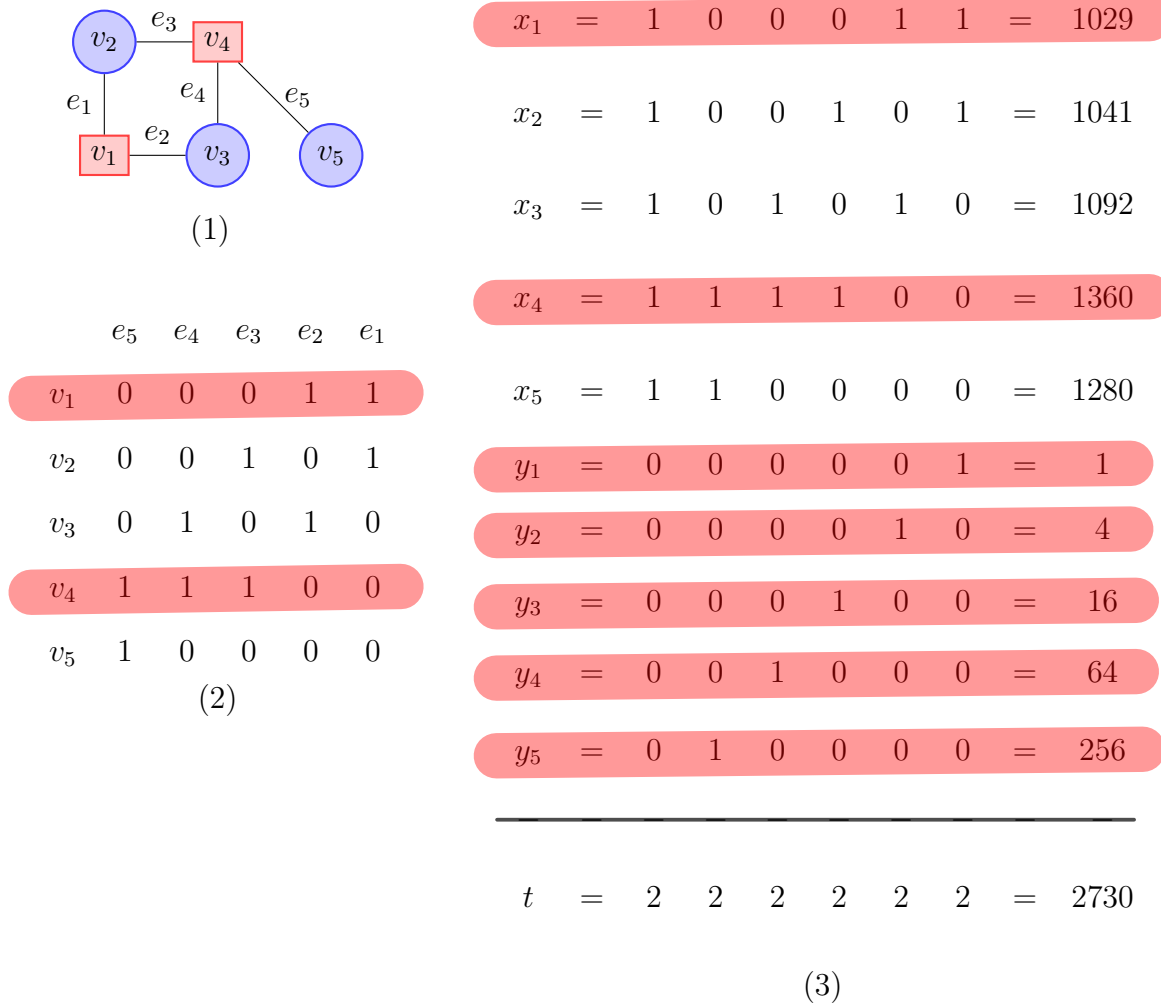
6.1. Redukcja problemu VERTEX-COVER do problemu SUBSET-SUM.

(1) Przykład grafu nieskierowanego. Wierzchołki minimalnego pokrycia wierzchołkowego prezentowane są na czerwono

(2) Macierz incydencji dla przykładowego grafu. Kolorem czerwonym zostały oznaczone wierzchołki z minimalnego pokrycia wierzchołkowego.

(3) Rozwiązanie problemu SUBSET-SUM odpowiadające przykładowemu grafu dla problemu VERTEX-COVER. Elementy typu  $x$  odpowiadają wierzchołkowi, a kolorem oznaczone są te których wierzchołki wchodzi do pokrycia wierzchołkowego. Elementy typu  $y$  którym odpowiadają krawędzie, kolorem czerwonym oznaczone te które mają koniec w dokładnie jednym wierzchołku pokrycia wierzchołkowego. Najbardziej znaczący bit tworzący sumę  $t$  jest równy rozmiarowi pokrycia wierzchołkowego, zaś reszta bitów ma wartość 2. Wszystkie elementy tworzą zbiór  $S$ , natomiast te oznaczone kolorem czerwonym tworzą szukaną sumę. Elementy oznaczone . . . . . 43

Spis tabel



**Rys. 6.1.** Redukcja problemu VERTEX-COVER do problemu SUBSET-SUM.

(1) Przykład grafu nieskierowanego. Wierzchołki minimalnego pokrycia wierzchołkowego prezentowane są na czerwono

(2) Macierz incydencji dla przykładowego grafu. Kolorem czerwonym zostały oznaczone wierzchołki z minimalnego pokrycia wierzchołkowego.

(3) Rozwiązanie problemu SUBSET-SUM odpowiadające przykładowemu grafu dla problemu VERTEX-COVER. Elementy typu  $x$  odpowiadają wierzchołkowi, a kolorem oznaczone są te których wierzchołki wchodzi do pokrycia wierzchołkowego. Elementy typu  $y$  którym odpowiadają krawędzie, kolorem czerwonym oznaczone te które mają koniec w dokładnie jednym wierzchołku pokrycia wierzchołkowego. Najbardziej znaczący bit tworzący sumę  $t$  jest równy rozmiarowi pokrycia wierzchołkowego, zaś reszta bitów ma wartość 2. Wszystkie elementy tworzą zbiór  $S$ , natomiast te oznaczone kolorem czerwonym tworzą szukaną sumę. Elementy oznaczone