Θεωρητική Πληροφορική Ι: Αλγόριθμοι και Πολυπλοκότητα

Chapter 9: NP-Complete Problems

9.3 Graph-Theoretic Problems (Συνέχεια)

9.4 Sets and Numbers

Γιώργος Αλεξανδρίδης
gealexan@mail.ntua.gr

Επισκόπηση Παρουσίασης

- 9.3 Θεωρητικά Προβλήματα Γράφων (Συνέχεια)
  - **HAMILTON PATH** είναι **NP-πλήρες**
    - Απόρροια: **TSP (D)** είναι **NP-πλήρες**
  - **3-COLORING** είναι **NP-πλήρες**
- 9.4 Σύνολα και Αριθμοί
  - **TRIPARTITE MATCHING** είναι **NP-πλήρες**
    - **EXACT COVER BY 3-SETS**, **SET COVERING** και **SET PACKING** είναι **NP-πλήρη**
  - **INTEGER PROGRAMMING** και **LINEAR PROGRAMMING**
  - **KNAPSACK** είναι **NP-πλήρες**
  - Ψευδοπολυωνυμικοί Αλγόριθμοι
  - Ισχυρή NP-πληρότητα
  - **BIN PACKING** είναι **NP-πλήρες**

HAMILTON PATH είναι NP-πλήρες: Επισκόπηση Απόδειξης

- **Αναγωγή** του **3SAT** σε **HAMILTON PATH**
  – Δοσμένης **έκφρασης *φ*** σε **CNF** με μεταβλητές $x_1,\ldots,x_n$ και **προτάσεις** (clauses) $C_1,\ldots,C_m$ η κάθε μια από τις οποίες αποτελείται από **3 κατηγορήματα** (literals), θα κατασκευαστεί **γράφος *R(φ)*,** ο οποίος θα διαθέτει **μονοπάτι Hamilton αν και μόνο αν** η **έκφραση** είναι **ικανοποιήσιμη**
- Για την **μετάβαση** από το **πεδίο** (domain) του **3SAT** στο πεδίο του **HAMILTON PATH** πρέπει να εξασφαλίζονται:
  – Η **επιλογή** (**choice**) των τιμών των μεταβλητών (τιμές **true** και **false**)
    - Χρήση του *choice gadget*
  – Η **συνέπεια** (**consistency**) στην εμφάνιση των μεταβλητών (όλες οι εμφανίσεις του ***x*** πρέπει να έχουν την **ίδια τιμή** και όλες οι εμφανίσεις του **¬*x*** πρέπει να έχουν την **αντίθετη τιμή**)
    - Χρήση του *consistency gadget*
  – Οι **περιορισμοί** (**constraints**) που επιβάλλονται από τις προτάσεις
    - Χρήση του *constraint gadget* (triangle)

HAMILTON PATH είναι NP-πλήρες: Choice Gadget (τιμές μεταβλητών)

- Αυτός ο **υπό-γράφος** (subgraph) **συνδέεται** με τον υπόλοιπο γράφο **μόνο** μέσω των **endpoints** (γεμισμένες τελείες)
- Το μονοπάτι Hamilton προσεγγίζει τον γράφο από πάνω και επιλέγει μια από τις δύο παράλληλες ακμές (και άρα μια τιμή αληθείας)

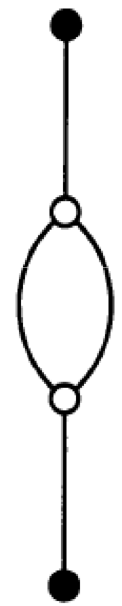


Figure 9-4. The choice gadget.

HAMILTON PATH είναι NP-πλήρες:
Consistency Gadget (Συνέπεια μεταβλητών)

- Ο **a** αποτελεί υπό-γράφο του G και συνδέεται με αυτόν μόνο στα endpoints
- Αν ο G έχει μονοπάτι Hamilton (και το μονοπάτι αυτό δεν αρχίζει από κάποιον εσωτερικό κόμβο του a), τότε ο **a μπορεί να διασχιστεί** με έναν από τους δύο **τρόπους που φαίνονται στην περίπτωση b και c**
  - Σε κάθε μονοπάτι, η πλευρά μπορεί να διασχιστεί μόνο μια φορά
  - Λειτουργία ως πύλη **XOR** (σχήμα d)

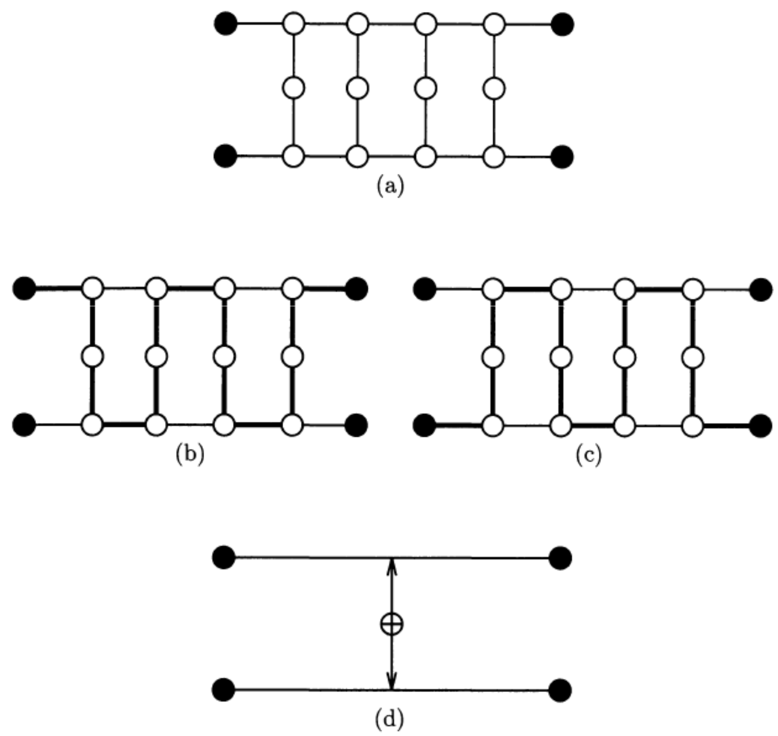


Figure 9-5. The consistency gadget.

HAMILTON PATH είναι NP-πλήρες:
Constraint Gadget (περιορισμοί προτάσεων)

- Μια πλευρά του τριγώνου για κάθε κατηγόρημα (literal) στην πρόταση (clause)
- Λειτουργία
  - Αν υποθέσουμε ότι οι πλευρές του τριγώνου διασχίζονται αν το αντίστοιχο κατηγόρημα έχει την τιμή **false**, τότε θα πρέπει **τουλάχιστον ένα** κατηγόρημα να έχει την τιμή **true**
    - Διαφορετικά και οι τρεις πλευρές του τριγώνου θα διασχιστούν και συνεπώς δεν θα υπάρχει μονοπάτι Hamilton

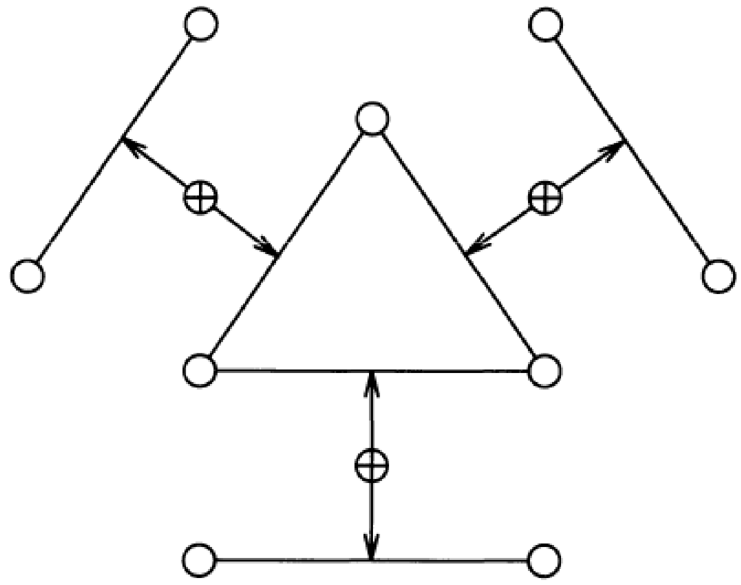


Figure 9-6. The constraint gadget.

HAMILTON PATH είναι NP-πλήρες: Κατασκευή γράφου R(ϕ)

- Ο γράφος G έχει
  - ***n*** αντίγραφα του **choice gadget**, ένα για κάθε **μεταβλητή**
  - ***m*** αντίγραφα του **constraint gadget** (τρίγωνα), ένα για κάθε **πρόταση**
  - Οι ***3m*** κόμβοι των **τριγώνων**, ο **τελευταίος κόμβος** του choice gadget και **ένας νέος κόμβος (3)** διασυνδέονται μεταξύ τους με όλους τους δυνατούς τρόπους σχηματίζοντας ένα μεγάλο clique
- Περιπτώσεις
  - Υπάρχει μονοπάτι Hamilton στον γράφο *R(φ)*
    - Τότε, ορίζεται truth assignment (αποτίμηση) *T* που ικανοποιεί το *φ*
  - *T* ικανοποιεί το *φ*
    - Τότε, μπορεί να βρεθεί μονοπάτι Hamilton στον γράφο *R(φ)*

$(x_1 \lor x_2 \lor x_3) \land (\neg x_1 \lor \neg x_2 \lor \neg x_3) \land (\neg x_1 \lor \neg x_2 \lor x_3)$

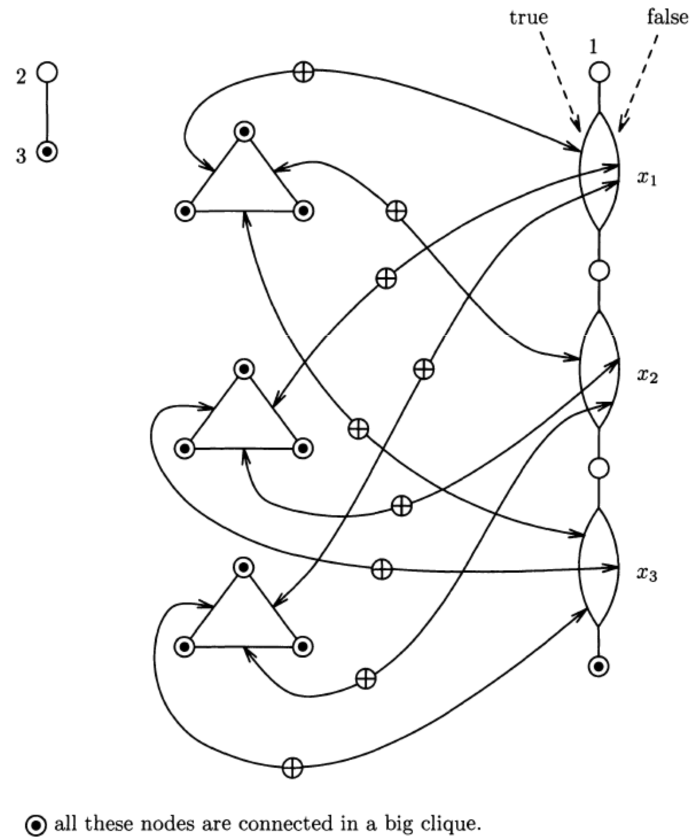


Figure 9-7. The reduction from 3SAT to HAMILTON PATH.

HAMILTON PATH είναι NP-πλήρες: TSP (D) είναι NP-πλήρες

- **Αναγωγή** του **HAMILTON PATH** σε **TSP (D)**
  - Δοσμένου **γράφου *G*** με ***n*** κόμβους, σχεδιάζεται πίνακας **αποστάσεων d_{ij}** και προϋπολογισμός **(budget) *B*** τέτοιος ώστε να **υπάρχει διαδρομή μήκους *B*** (ή και μικρότερη) **αν και μόνο αν** ο γράφος ***G*** έχει **μονοπάτι Hamilton**
  - *n* πόλεις (μια για κάθε κόμβο του γράφου)
  - Η **απόσταση** μεταξύ των **πόλεων *i*** και ***j*** είναι **1** αν **υπάρχει ακμή *[i, j]*** στον *G* και **2 διαφορετικά**
  - Το *B* τίθεται ίσο με *n + 1*

3-COLORING είναι NP-πλήρες: Επισκόπηση Απόδειξης

- k-COLORING
  - Χρωματίζουμε τους κόμβους ενός γράφου με διαφορετικά χρώματα, έτσι ώστε γειτονικοί κόμβοι να μην έχουν το ίδιο χρώμα
- **Αναγωγή** του **NAESAT** σε **3-COLORING**
  - NAESAT: Δοσμένου συνόλου προτάσεων $C_1,\ldots,C_m$, η κάθε μια από τις οποίες περιέχει 3 κατηγορήματα και μεταβλητών $x_1,\ldots,x_n$, υπάρχει αποτίμηση τέτοια στις μεταβλητές, ώστε καμία από τις προτάσεις να μην έχει όλα τα κατηγορήματα της true ή όλα false;
- Κατασκευή **γράφου *G***, ο οποίος μπορεί να χρωματιστεί με **χρώματα *{0,1,2}* αν και μόνο αν** τα **κατηγορήματα** σε όλες τις **προτάσεις** του μπορούν να πάρουν **διαφορετικές τιμές**

3-COLORING είναι NP-πλήρες: Απόδειξη

- Απόδειξη με χρήση τριγώνων
  - Χρησιμοποιούμε και τα **3 χρώματα** στις **ακμές τους**
  - Για κάθε **μεταβλητή x_i** κατασκευάζεται **τρίγωνο *[α,x_i,-x_i]***
    - Όλα τα τρίγωνα μοιράζονται μεταξύ τους τον κόμβο *α*
  - Κάθε **πρόταση C_i** αναπαρίσταται από **τρίγωνο *[C_{i1},C_{i2},C_{i3}]*** (κάτω μέρος της εικόνας)
  - Οι ακμές μεταξύ των δύο ομάδων τριγώνων υποδηλώνουν την ύπαρξη των κατηγορημάτων στις προτάσεις
- Ο G μπορεί να τρι-χρωματιστεί αν και μόνο αν η συγκεκριμένη περίπτωση (instance) του NAESAT ικανοποιείται

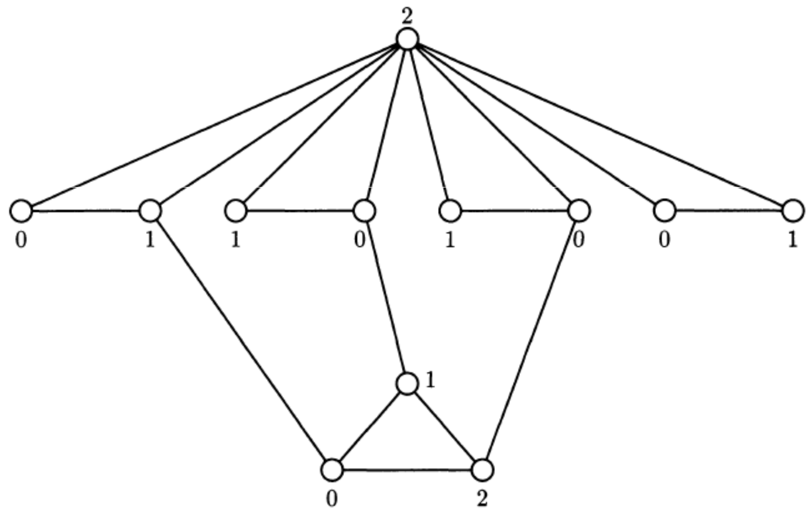


Figure 9-8. The reduction to 3-COLORING.

TRIPARITE MATCING είναι NP-πλήρες: Επισκόπηση Απόδειξης

- TRIPARTITE MATCHING επέκταση του BIPARTITE MATCHING (ενότητα 1.2)
  - Δίνονται **3 σύνολα *B,G* και *H*** (boys, girls και homes) n στοιχείων το κάθε ένα και **τριαδική** (ternary) **σχέση** που τα συνδέει $\boldsymbol{T} \subseteq B \times G \times H$
  - Ζητείται το , δηλαδή κάθε αγόρι αντιστοιχίζεται με διαφορετικό κορίτσι και κάθε ζευγάρι έχει το δικό του σπίτι
- Αναγωγή του 3SAT στο TRIPARTITE MATCHING
  - 3SAT: Έκφραση φ σε CNF με μεταβλητές $x_1,\ldots,x_n$ και προτάσεις $C_1,\ldots,C_m$, με την κάθε πρόταση να έχει ακριβώς 3 κατηγορήματα
  - Χρήση gadget που συνδυάζει την επιλογή και την συνέπεια
    - Για κάθε μεταβλητή *x* στην έκφραση

TRIPARITE MATCING είναι NP-πλήρες: Κατασκευή Απόδειξης

- Επιλογή και συνέπεια
  - *k* boys, *k* girls αλλά *2k* homes
    - *k* εμφανίσεις μεταβλητής *x*
  - Οι μονοί δείκτες του *h* υποδηλώνουν τις εμφανίσεις της μεταβλητής *x* (οι ζυγοί τις εμφανίσεις του ¬*x*)
  - Αν υπάρχει αντιστοίχηση, τότε:
    - b_i αντιστοιχίζεται με g_i και h_{2i}
      - Θεωρούμε τότε ότι *T(x)* = true
    - b_i αντιστοιχίζεται με g_{i-1} και h_{2i-1}
      - Θεωρούμε τότε ότι *T(x)* = false
- Περιορισμοί
  - Για κάθε πρόταση *c* υπάρχουν 3 τριπλέτες *(b, g, h)*, όπου το *h* αντιστοιχεί στις 3 εμφανίσεις των κατηγορημάτων της κάθε πρότασης *c*
  - Αν κάποιο από τα *h* μείνει χωρίς αντιστοίχηση, τότε αυτό αντιστοιχεί σε true literal
    - Τότε η c ικανοποιείται
  - Αν και τα 3 κατηγορήματα της πρότασης είναι false τότε τα *b, g* δεν μπορούν να αντιστοιχηθούν με *h*

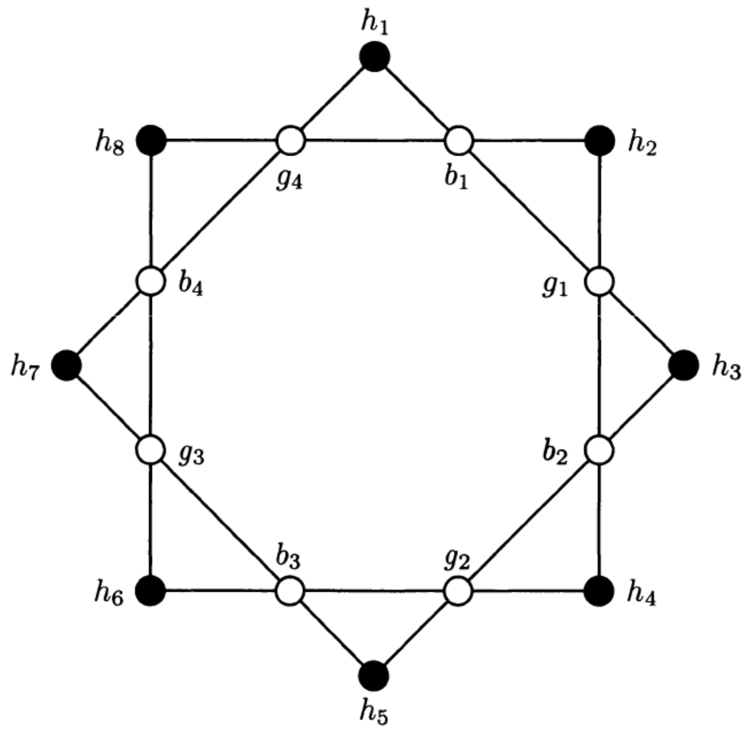


Figure 9-9. The choice-consistency gadget.

TRIPARITE MATCHING είναι NP-πλήρες: Απόρροια

- EXACT COVER BY 3-SETS, SET COVERING και SET PACKING είναι NP-πλήρη
- **SET COVERING**
  - Δίνεται οικογένεια υποσυνόλων $\boldsymbol{F=\{S_1,\ldots,S_n\}}$ ενός πεπερασμένου συνόλου ***U*** και ένα **budget *B***. Υπάρχει σύνολο ***B*** από υποσύνολα του ***F***, των οποίων η ένωση να δίνει *U*;
- **SET PACKING**
  - Παραλλαγή του SET COVERING
    - Δίνεται **οικογένεια** υποσύνολων ***F* του *U*** και ένας **στόχος** (goal) ***K***.
    - **Υπάρχουν *K* ζεύγη**, τα μέλη των οποίων είναι **μη-επικαλυπτόμενα** (disjoint), στην οικογένεια *F*;
- **EXACT COVER BY 3-SETS**
  - Παραλλαγή του SET COVERING
    - Το **πλήθος** των στοιχείων του ***U*** είναι ***3m*** (όπου *m* ακέραιος),
    - Το **πλήθος** των στοιχείων κάθε **υποσυνόλου** $\boldsymbol{S_i}$ είναι **3**
    - Η τιμή του **budget** είναι ***m*** με την επιπλέον προϋπόθεση τα **υποσύνολα** $\boldsymbol{S_i}$ πλήθους *m* να είναι **μη-επικαλυπτόμενα**

INTEGER PROGRAMMING vs. LINEAR PROGRAMMING

- INTEGER PROGRAMMING
  - Ένα σύστημα γραμμικών ανισώσεων *n* μεταβλητών με ακεραίους όρους έχει ακέραιες λύσεις;
- LINEAR PROGRAMMING
  - Ίδιο πρόβλημα με τη μόνη διαφορά ότι επιτρέπονται και μη-ακέραιες λύσεις
- **INTEGER PROGRAMMING** είναι **NP-πλήρες**
  - SET COVERING μπορεί να αναχθεί σε INTEGER PROGRAMMING
    - $Ax \geq \mathbf{1}; \sum_{i=1}^{n} x_i \leq B; 0 \leq x_i \leq 1$
      - $\boldsymbol{X_i}$ δυαδική μεταβλητή που παίρνει την **τιμή 1 ανν** το $\boldsymbol{S_i}$ είναι στο **cover**
      - ***A*** πίνακας του οποίου οι **γραμμές** αποτελούν τα **δυαδικά ανύσματα** των υποσυνόλων
      - ***B*** το budget
- **LINEAR PROGRAMMING** είναι στο **P**

KNAPSACK είναι NP-πλήρες: Επισκόπηση

- Knapsacks
  - Ειδική περίπτωση του INTEGER PROGRAMMING
  - Δίνεται σύνολο **αντικειμένων** ***n***, κάθε ένα από τα οποία έχει **αξία (value)** $\boldsymbol{v_i}$ και **βάρος (weight)** $\boldsymbol{w_i}$ και μπορεί να επιλεγεί μόνο μια φορά. Δίνεται επίσης **όριο βάρους** ***W***
  - Ποια επιλογή αντικειμένων **μεγιστοποιεί την συνολική τους αξία**, **χωρίς υπέρβαση** του ορίου ***W***;
- Πρόβλημα KNAPSACK
  - Για δεδομένο **στόχο** ***K***, υπάρχει υποσύνολο $\boldsymbol{S} \subseteq \boldsymbol{\{1,\ldots,n\}}$ τέτοιο ώστε $\sum_{i \in S} w_i \leq W$ και $\sum_{i \in S} v_i \geq K$;

KNAPSACK είναι NP-πλήρες: Απόδειξη

- **Αναγωγή EXACT COVER BY 3-SETS σε KNAPSACK**
  - Θεωρούμε περίπτωση $\{S_1, S_2, \ldots, S_n\}$ του EXACT COVER BY 3-SETS, όπου $U=\{1,2,\ldots,3m\}$, όπου m ακέραιος
  - Τα υποσύνολα S_i αναπαρίστανται από **δυαδικά ανύσματα** στον χώρο $\{0,1\}^{3m}$, που επίσης μπορούν να ειδωθούν ως ακέραιοι αριθμοί
  - Έτσι η ένωση των συνόλων αναπαρίσταται ως πρόσθεση ακεραίων
    - **ΕΙΔΟΠΟΙΟΣ ΔΙΑΦΟΡΑ**: Στην πρόσθεση ακεραίων πρέπει να λαμβάνουμε υπόψη μας και το κρατούμενο!
    - **ΛΥΣΗ**: Θεωρούμε ότι τα δυαδικά ανύσματα αναπαρίστανται από **ακεραίους αριθμούς ως προς βάση *n* + *1*** και όχι ως προς βάση *2*.

$$
\begin{array}{ccccccccccccc}
\rightarrow & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\rightarrow & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
+ & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}
$$

Figure 9.10. Reduction to KNAPSACK.

Ψευδοπολυωνυμικοί Αλγόριθμοι

- **Κάθε περίπτωση** του **KNAPSACK** μπορεί να λυθεί σε **χρόνο** $\mathcal{O}(nW)$, όπου n είναι ο **αριθμός των αντικειμένων** και W είναι το **όριο βάρους**
- ΠΡΟΣΟΧΗ: Δεν είναι πολυωνυμικός αλγόριθμος γιατί **το χρονικό του όριο nW δεν αποτελεί πολυωνυμική συνάρτηση της εισόδου του**
  - Αντίστοιχη περίπτωση ήταν και ο αλγόριθμος που επέλυε το MAX CUT στην ενότητα 1.2
- Οι ψευδοπολυωνυμικοί αλγόριθμοι έχουν χρησιμότητα στην θεωρία της πολυπλοκότητας
  - Κεφάλαιο 13, προσεγγιστικοί αλγόριθμοι

Ισχυρή NP-πληρότητα

- Αν ένα **πρόβλημα παραμένει NP-πλήρες**, ακόμα και στις περιπτώσεις του εκείνες όπου το **μήκος της εισόδου** του n **περιορίζεται πολυωνυμικά** (είναι δηλαδή της μορφής *p(n)*, όπου *p(n)* πολυώνυμο) τότε ονομάζεται **ισχυρό NP-πλήρες πρόβλημα**
  - **Όλα τα προβλήματα** που εξετάστηκαν στο **Κεφ 9** είναι **ισχυρά NP-πλήρη** προβλήματα
    - **Εξαίρεση το KNAPSACK**
- Τα **ισχυρά NP-πλήρη** προβλήματα **δεν μπορούν** να **εκφραστούν** από **ψευδοπολυωνυμικούς** αλγορίθμους
  - Εκτός και αν κάποτε αποδειχθεί ότι P = NP

BIN PACKING είναι NP-πλήρες: Επισκόπηση Απόδειξης

- BIN PACKING
  - Δίνεται σύνολο ***N* θετικών ακεραίων** $\alpha_1, \alpha_2, \ldots, \alpha_N$ (αντικείμενα), ένας **ακέραιος *C*** (η χωρητικότητα) και ένας **ακέραιος *B*** (ο αριθμός των δοχείων)
  - Μπορούν οι συγκεκριμένοι αριθμοί να χωριστούν σε ***B* υποσύνολα**, κάθε ένα εκ των οποίων έχει **χωρητικότητα το πολύ *C*;**
- **Αναγωγή** του **TRIPARTITE MATCHING** σε **BIN PACKING**
  - Έστω σύνολο αγοριών $B=\{b_1,b_2,\ldots,b_n\}$, κοριτσιών $G=\{g_1,g_2,\ldots,g_n\}$, σπιτιών $H=\{h_1,h_2,\ldots,h_n\}$ και τριπλετών $T=\{t_1,..t_m\} \subseteq B \times G \times H$
  - Υπάρχει **σύνολο *n* τριπλετών** στο *T*, έτσι ώστε **κάθε αγόρι, κορίτσι και σπίτι** να **περιέχεται σε μια** από τις n τριπλέτες;

BIN PACKING είναι NP-πλήρες: Απόδειξη

- ***N = 4m*** αντικείμενα
  - Ένα για κάθε τριπλέτα και ένα για κάθε εμφάνιση αγοριού, κοριτσιού ή σπιτιού σε κάθε τριπλέτα
  - *b_1[1], b_1[2],b_1[N(b_1)]*: εμφανίσεις του *b_1* στις τριπλέτες
  - *N(b_1)*: συνολικός αριθμός εμφανίσεων του *b_1*
- ***M*** πολύ **μεγάλος αριθμός** (πχ *100n*)
- **Μια από τις εμφανίσεις** των αγοριών, των κοριτσιών και των σπιτιών (εδώ επιλέχθηκε αυθαίρετα η πρώτη) **έχει διαφορετικό μέγεθος από τις άλλες**
  - αυτή η εμφάνιση θα συμμετάσχει στο ταίριασμα
- **Χωρητικότητα *C*** κάθε δοχείου: ***$40M^4$ + 15***
  - Χωράει ακριβώς μια τριπλέτα και μια εμφάνιση από κάθε μέλος της με την προϋπόθεση ότι καμία από τις τρείς ή και οι τρεις ανήκουν στην πρώτη εμφάνιση
- Αριθμός **Δοχείων: *m***
- Αριθμός **Τριπλετών: *m***
- ***N* τριπλέτες στα δοχεία** σχηματίζουν **tripartite matching**
- **Αν υπάρχει** tripartite **matching**, **τότε όλα τα** αντικείμενα μπορούν να χωρέσουν στα m δοχεία

Item	Size
first occurrence of a boy $b_i[1]$	$10M^4 + iM + 1$
other occurrences of a boy $b_i[q], q > 1$	$11M^4 + iM + 1$
first occurrence of a girl $g_j[1]$	$10M^4 + jM^2 + 2$
other occurrences of a girl $g_j[q], q > 1$	$11M^4 + jM^2 + 2$
first occurrence of a home $h_k[1]$	$10M^4 + kM^3 + 4$
other occurrences of a home $h_k[q], q > 1$	$8M^4 + kM^3 + 4$
triple $(b_i, g_j, h_k) \in T$	$10M^4 + 8-$ $-iM - jM^2 - kM^3$

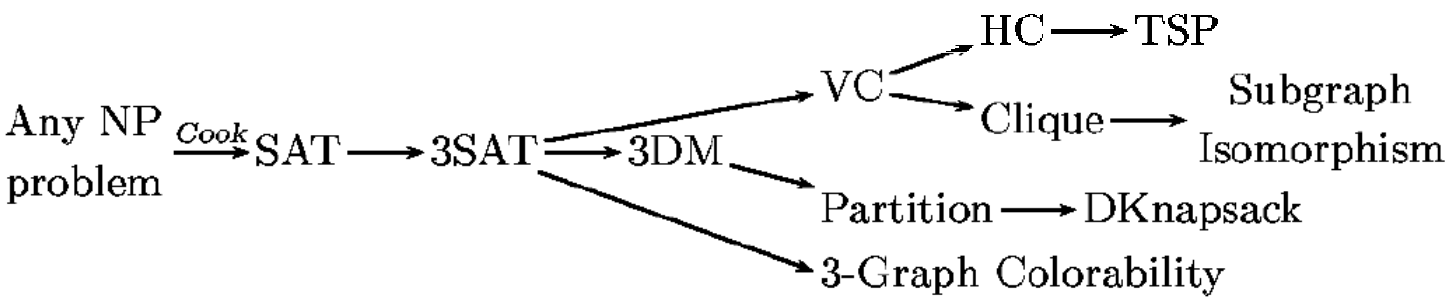
Figure 9.11. The items in BIN PACKING.

BIN PACKING είναι NP-πλήρες: Συμπεράσματα

- Οι αριθμοί που χρησιμοποιήθηκαν στην απόδειξη είναι πολυωνυμικά μεγάλοι, $\mathcal{O}(|x|^4)$
  - x αρχικό instance του TRIPARTITE MATCHING
- Το BIN PACKING είναι ισχυρό NP-πλήρες πρόβλημα και αποτελεί σημείο αφετηρίας για αναγωγές σε προβλήματα στα οποία οι αριθμοί παίζουν κεντρικό ρόλο

Αναγωγές μεταξύ NP-πλήρων προβλημάτων

- Έγιναν με την παρακάτω σειρά που και παρουσιάστηκαν από τον Karp (1972)



Πηγή: Στάθης Ζάχος και Άρης Παγουρτζής, Διαφάνειες Μαθήματος «Αλγόριθμοι και Πολυπλοκότητα»