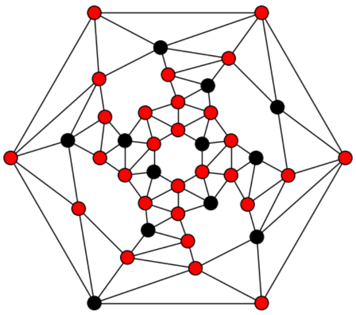
Πανεπιστήμιο Πατρών

Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογιάς Υπολογιστών

Τελική εργασία

Vertex Cover Problem



Δημήτριος Δήμου

June 1, 2016

Contents

List of Figures

Chapter 1

Set cover problems

To set cover problem είναι ένα κλασικό πρόβλημα στον τομέα της συνδυαστικής βελτιστοποίησης και της θεωρίας υπολογιστών η μελέτη του οποίου έχει οδηγήσει στην ανάπτυξη θεμελιωδών τεχνικών στο πεδίο των προσεγγιστικών αλγορίθμων. Λόγω της γενικής του διατύπωσης βρίσκει εφαρμογές σε μια ευρεία γκάμα προβλημάτων.

1.1 Διατύπωση

Το πρόβλημα μπορεί να διατυπωθεί με διάφορους τρόπους, είτε ως πρόβλημα βελτιστοποίησης όπου ζητείται ο ελάχιστος αριθμός υποσυνόλων ή, αν έχει ανατεθεί συνάρτηση κόστους στα υποσύνολα, ζητείται το σύνολο με το ελάχιστο κόστος, είτε ως πρόβλημα απόφασης.

Απλή διατύπωση

Δεδομένου ενός σύμπαντος U αποτελούμενο από n στοιχεία κι ενός συνόλου από υποσύνολα του U, $S = \{S_1, ..., S_k\}$ τέτοια ώστε η ένωσή τους να είναι το σύνολο U, βρες το ελάχιστο υποσύνολο του S που καλύπτει όλα τα στοιχεία του U.

Με συνάρτηση κόστους

Δεδομένου ενός σύμπαντος U αποτελούμενο από n στοιχεία, μιας συλλογής από υποσύνολα του U, $S = S_1, ..., S_k$, και μίας συνάρτησης κόστους $c : S \rightarrow \boldsymbol{Q}^+$, βρες το υποσύνολο του S με το ελάχιστο κόστος που καλύπτει όλα τα στοιχεία του U.

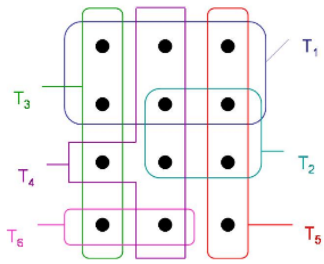
Πρόβλημα απόφασης

Δεδομένου ενός σύμπαντος U αποτελούμενο από n στοιχεία, μιας συλλογής από υποσύνολα του U, $S = S_1, ..., S_k$, και ενός ακεραίου k, αποφάσισε αν υπάρχει υποσύνολο του S με το πολύ k στοιχεία που καλύπτει όλα τα στοιχεία του U.

Παράδειγμα

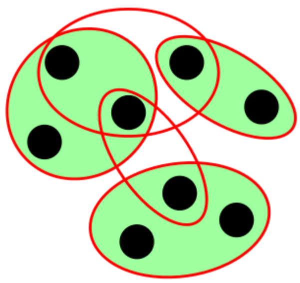
Έστω το σύμπαν $U = \{1, 2, 3, 4, 5\}$ και η συλλογή από υποσύνολα του $S = \{\{1, 2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}\}$. Η ένωση των στοιχείων του S καλύπτει το U. Η ελάχιστη συλλογή υποσυνόλων του S που καλύπτει το U είναι τα : $\{\{1, 2, 3\}, \{4, 5\}\}$.

FIGURE 1.1: Set cover example



Το ελάχιστο set cover είναι το σύνολο $S' = \{T_3, T_4, T_5\}$

FIGURE 1.2: Minimum set cover example



1.2 NP-πληρότητα

Το set cover decision problem είναι ένα από τα 21 NP-πλήρης προβλήματα του Karp που αποδείχθηκε ότι είναι NP-πλήρης το 1972. Αυτό σημαίνει ότι ανήκει στην κλάση NP δηλαδή δεδομένου ενός σύμπαντος U, μιας συλλογής S από υποσύνολα, ενός ακεραίου k και μίας λύσης S' η λύση αυτή μπορεί να επαληθευτεί σε πολυωνυμικό χρόνο όσον αφορά το μέγεθος των στοιχείων της εισόδου. Επίσης ανήκει και στην κλάση NP-hard. Αυτό οδήγησε στην ανάπτυξη προσεγγιστικών αλγορίθμων για την επίλυση του προβλήματος αυτού.

1.3 Λύσεις

1.3.1 Πρόβλημα ακέραιου προγραμματισμού

Το minimum set cover problem μπορεί να διατυπωθεί ως το ακόλουθο πρόβλημα ακέραιου προγραμματισμού

$$min\{\sum_{S \in \mathcal{S}} x_S\}$$

subject to

$$\sum_{S : e \in \mathcal{S}} x_S \geq 1, \quad \forall e \in \mathcal{U}$$

$$x_S \in \{0, 1\}$$

Επειδή το πρόβλημα ακέραιου προγραμματισμού είναι NP-hard χαλαρώνουμε τους περιορισμούς του προβλήματος για το x_S και το ανάγουμε σε πρόβλημα γραμμικού προγραμματισμού το οποίο λύνεται σε πολυωνυμικό χρόνο. Έτσι καταλήγουμε στο πρόβλημα:

$$min\{\sum_{S\in\mathcal{S}} x_S\}$$

subject to

$$\sum_{S:e\in\mathcal{S}} x_S \geq 1, \quad \forall e \in \mathcal{U}$$

$$x_S \in [0,1]$$

Επειδή το integrality gap αυτού του προβλήματος είναι το πολύ $\log n$ η χαλάρωση του δίνει factor-$\log n$ προσεγγιστικό αλγόριθμο. Αν κάθε στοιχείο εμφανίζεται το πολύ σε $\mathcal{F}$ τότε μπορεί να βρεθεί λύση σε πολυωνυμικό χρόνο η οποία προσεγγίζει το βέλτιστο με παράγοντα $\mathcal{F}$ χρησιμοποιώντας το πρόβλημα γραμμικού προγραμματισμού.

1.3.2 Άπληστος αλγόριθμος

Υπάρχει και ένας άπληστος αλγόριθμος για προσέγγιση σε πολυωνυμικό χρόνο ο οποίος διαλέγει τα σύνολα με έναν μόνο κανόνα: σε κάθε στάδιο διαλέγει το σύνολο που περιέχει τον μεγαλύτερο αριθμό από ακάλυπτα στοιχεία.

Αλγόριθμος:

1. $C \leftarrow \emptyset$
2. While $C \neq \mathcal{U}$ do
   (α′) Find the set whose cost effectiveness is smallest, say S_i.
   Let $a = \frac{c(S_i)}{|S_i - C|}$.
   Pick S_i and $\forall e \in S_i - C$, set $price(e) = a$.
   (β′) $C \leftarrow S_i \cup C$
3. Output C

1.4 Hitting set

Το πρόβλημα set cover είναι ισοδύναμο με το πρόβλημα hitting set. Αν σε έναν διμερή γράφο το ένα σύνολο κόμβων $\mathcal{U}$ αντιπροσωπεύει τα υποσύνολα $\mathcal{S}$ του σύμπαντος, το άλλο σύνολο κόμβων $\mathcal{V}$ αντιπροσωπεύει τα στοιχεία του σύμπαντος και οι ακμές αντιπροσωπεύουν την συμπερίληψη ενός στοιχείου σε ένα σύνολο τότε βρίσκουμε τον ελάχιστο αριθμό κόμβων του συνόλου $\mathcal{U}$ που καλύπτει όλους τους κόμβους του συνόλου $\mathcal{V}$.

1.5 Εφαρμογές

IBM finds computer viruses (wikipedia) elements- 5000 known viruses sets- 9000 substrings of 20 or more consecutive bytes from viruses, not found in 'good' code A set cover of 180 was found. It suffices to search for these 180 substrings to verify the existence of known computer viruses. Another example: Consider General Motors needs to buy a certain amount of varied supplies and there are suppliers that offer various deals for different combina tions of materials (Supplier A: 2 tons of steel + 500 tiles for $x; Supplier B: 1 ton of steel + 2000 tiles for $y; etc.). You could use set covering to find the best way to get all the materials while minimizing cost

Chapter 2

Vertex cover problem

Το vertex cover ενός γράφου είναι ένα σύνολο κόμβων τέτοιο ώστε κάθε ακμή του γράφου είναι προσκήμενη σε τουλάχιστον ένα κομβό του συνόλου αυτού. Το πρόβλημα εύρεσης του ελάχιστου vertex cover είναι κλασικό προβλημα στον τομέα της συνδυαστικής βελτιστοποίησης και της θεωρίας υπολογιστών και κλασικό παράδειγμα NP-hard προβλήματος βελτιστοποίησης.

2.1 Διατύπωση

Το vertex cover V' ενός μη κατευθυντικού γράφου $G = (V, E)$ είναι ένα υποσύνολο του V τέτοιο ώστε:

$$\forall uv \in E \Rightarrow u \in V' \vee v \in V'$$

Ένα τέτοιο σύνολο λέμε ότι καλύπτει τις ακμές του G. Το ελάχιστο vertex cover ενός γράφου G είναι το σύνολο V' με τον μικρότερο αριθμό στοιχείων.

Figure 2.1: Vertex cover

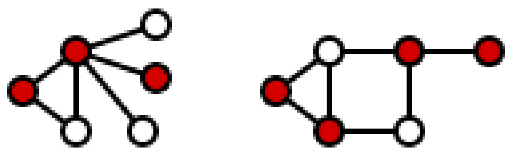
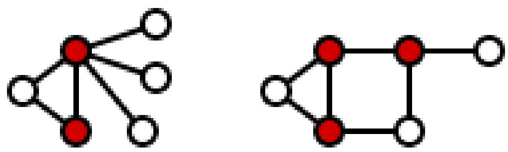


Figure 2.2: Minimum vertex cover



2.2 NP-πληρότητα

Στη γενική περίπτωση το πρόβλημα vertex cover είναι NP-πλήρης οπότε είναι απίθανο να βρεθεί ακριβής αλγοριθμική λύση σε πολυωνυμικό χρόνο εκτός και αν P = NP. Η NP-πληρότητα μπορεί να αποδειχθεί με υπαγωγή από το 3-SAT πρόβλημα ή από το Clique πρόβλημα. Σε ειδικές περιπτώσεις γράφων όμως μπορούν να βρεθούν πολυωνυμικοί αλγόριθμοι. Με εξαντλητική αναζήτηση το πρόβλημα μπορεί να λυθεί σε $2^k n^{O(1)}$ χρόνο, όπου k το μέγεθος του συνόλου και n ο αριθμός των κόμβων, το οποίο κάνει το πρόβλημα fixed-parameter tractable.

Οπότε αν ενδιαφερόμαστε για μικρά k μπορούμε να επιλέξουμε αυτή τη μέθοδο και να έχουμε λύση σε πολυωνυμικό χρόνο.

2.3 Λύσεις

2.3.1 Πρόβλημα ακέραιου προγραμματισμού

Το minimum vertex cover problem μπορεί να διατυπωθεί ως το ακόλουθο πρόβλημα ακέραιου προγραμματισμού

$$min\{\sum_{u\in V} c(v)x_v\}$$

subject to

$$x_u + x_v \geq 1, \quad \forall (u,v) \in E$$

$$x_v \in \{0,1\} \quad \forall v \in V$$

όπου $c : V \rightarrow \boldsymbol{R}^+$ μια συνάρτηση κόστους για τους κόμβους του γράφου.

Επειδή το πρόβλημα ακέραιου προγραμματισμού είναι NP-hard χαλαρώνουμε τους περιορισμούς του προβλήματος για το x_v και το ανάγουμε σε πρόβλημα γραμμικού προγραμματισμού το οποίο λύνεται σε πολυωνυμικό χρόνο. Έτσι καταλήγουμε στο πρόβλημα:

$$min\{\sum_{u\in V} c(v)x_v\}$$

subject to

$$x_u + x_v \geq 1, \quad \forall (u,v) \in E$$

$$x_v \geq 0, \quad \forall v \in V$$

Το integrality gap αυτού του προβλήματος είναι 2 οπότε η χαλάρωση του δίνει έναν factor-2 προσεγγιστικό αλγόριθμο. Επίσης η γραμμική χαλάρωση του προβλήματος είναι half-integral, δηλάδη υπάρχει βέλτιστη λύση στην οποία $x_v \in \{0, \frac{1}{2}, 1\}$

2.3.2 Προσεγγιστικοί αλγόριθμοι

Έχουν αναπτυχθεί πολλές παραλλαγές προσεγγιστικών αλγορίθμων που λύνουν το συγκεκριμένο πρόβλημα. Ο πιο απλός αλγόριθμος είναι factor-2 προσεγγιστικός και αναπτύχθηκε ανεξάρτητα από τους Fanica Gavril και τον Μιχάλη Γιαννακάκη.
Η γενική ιδέα είναι η εξής: σε κάθε επανάληψη διαλέγει μια ακμή και εισάγει και τα δύο άκρα (u, v) της στο vertex cover V', και αφαιρεί από το σύνολο των ακμών κάθε ακμή που είναι προσκείμενη είτε στον κόμβο u είτε στον v μέχρι να μείνει το κενό σύνολο.
Αλγόριθμος:

1. $V' \leftarrow \emptyset$
2. $E' \leftarrow E$
3. While $E' \neq \emptyset$ do

α′) let (u, v) be an arbitrary edge of E'

β′) $V' \leftarrow V' \cup \{u, v\}$

γ′) remove from E' every edge incident on either u or v

4. Output V'

Ο αλγόριθμος αυτός τρέχει σε χρόνο $O(|V|+|E|)$. Όσον αφορά τον παράγοντα προσέγγγισης του αλγορίθμου φαίνεται εύκολα ότι για το σύνολο των ακμών που επιλέγονται στο βήμα α') ισχύει

$$|V^*| \geq |A|$$

αφού το σύνολο δεν περιέχει προσκήμενες ακμές και επειδή το σύνολο V' που επιστρέφει ο αλγόριθμος περιέχει και τις δυο κορυφές των ακμών που επιλέγει έχουμε

$$|V'| = 2|A|$$

οπότε

$$|V'| \leq 2|V^*|$$

Έχουν αναπτυχθεί και άλλοι προσεγγίστικοι αλγόριθμοι με καλύτερο παράγοντα προσέγγισης, όπως $2-\Theta\left(\frac{1}{\sqrt{\log|V|}}\right)$ αλλά δεν έχει βρεθεί καλύτερος αλγόριθμος σταθερού προσεγγιστικού πράγοντα. Το minimum vertex cover πρόβλημα είναι $APX-$πλήρης δηλαδή δεν μπορεί να προσσεγιστεί αυθαίρετα καλά αν δεν ισχύει $P = NP$. Οι Dinur και Safra απέδειξαν ότι το πρόβλημα δε μπορεί να προσεγγιστεί με παράγοντα μικρότερο του 1.3606 για έναν ακρετά μεγάλο γράφο αν δεν ισχύει $P = NP$, επίσης αν ισχυέι η εικάσια unique games τότε το πρόβλημα δεν μπροεί να προσεγγιστεί με σταθερό πράγοντα μικρότερο του 2.

Chapter 3

Ειδικές περιπτώσεις

Το πρόβλημα vertex cover παρόλο που είναι NP-πλήρης στη γενική περίπτωση σε κάποιες ειδικές περιπτώσεις γράφων είναι δυνατό να λυθεί σε πολυωνυμικό χρόνο. Παρακάτω παραθέτονται κάποιες από αυτές τις ειδικές περιπτώσεις και ορισμένες λύσεις και πορίσματα.

3.1 Bigraphs

3.1.1 Εισαγωγικές έννοιες

1. Διμερής γράφοι:
   Οι διμερής γράφοι είναι ειδικές περιπτώσεις γράφων όπου οι κόμβοι τους μπορούν να χωριστούν σε δύο ανεξάρτητα σύνολα $\mathcal{U}$ και $\mathcal{V}$ ετσί ώστε κάθε ακμή να ενώνει ένα κόμβο του $\mathcal{U}$ με έναν κόμβο του $\mathcal{V}$.

2. Matching:
   Δεδομένου ενός γράφου $G = (V, E)$ ονομάζουμε matching στον G ένα σύνολο μη γειτονικών ακμών, δηλάδη ακμές που δεν είναι προσκήμενες στον ίδιο κόμβο. Μέγιστο matching ονομάζουμε το σύνολο με τον μέγιστο αριθμό ακμών.

FIGURE 3.1: Matching example

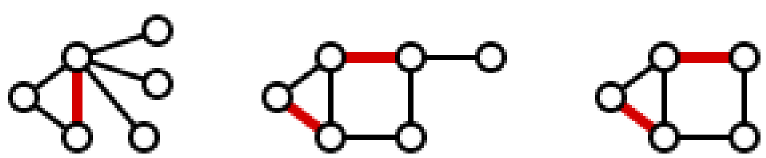
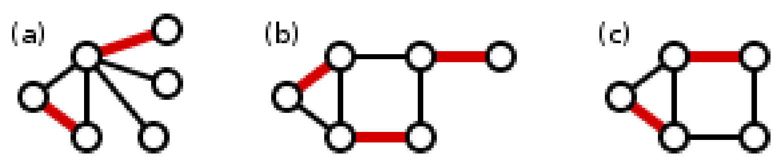


FIGURE 3.2: Maximum matching example



3.1.2 Θεώρημα Konig

Σε αυτό το είδος γράφων βρίσει εφαρμογή το θεώρημα του Konig το οποίο αναφέρει πως για κάθε διμερή γράφο $G = (V, E)$ ισχύει $\nu(G) = \tau(G)$ όπου
$\nu(G) :=$ maximum size of a matching in G,
$\tau(G) :=$ minimum size of a vertex cover in G.

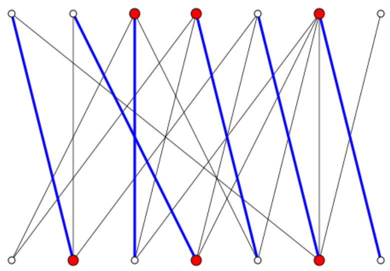
3.1.3 Απόδειξη

Έστω ένας διμερής γράφος $G = (L, R, E)$ και ένα μέγιστο ταίριασμα M. Τότε επειδή κανένας κόμβος ενός vertex cover δεν μπορεί να καλύπτει περισσότερες από δύο ακμές του συνόλου M (διαφορετικά δεν θα ήταν ταίριασμα), ένα vertex cover μεγέθους $|M|$ θα είναι το ελάχιστο vertex cover.
Για να δημιουργήσουμε ένα τέτοιο vertex cover, έστω U το σύνολο των μη ταιριασμένων κόμβων του L, και Z το σύνολο των ακμών που είτε είναι στο U είτε συνδέονται με αυτό μέσω εναλακτικών μονοπατιών (alternating paths). Τότε το σύνολο $V' = (L \setminus Z) \cup (R \cap Z)$ είναι ένα vertex cover.

3.1.4 Συμπεράσματα

Οπότε χρησιμοποιόντας τον αλγόριθμο Hopcroft-Karp, ο οποίος βρίσκει ένα μέγιστο ταίριασμα σε ένα διμερή γράφο σε χρόνο $O(|E|\sqrt{|V|})$, μπορούμε έπειτα να υπολογίσουμε το σύνολο V' αποδοτικά.

FIGURE 3.3: Maximum matching - minimum vertex cover example



3.2 Tree graphs

Για δένδρα υπάρχει ένας άπληστος αλγόριθμος που βρίσκει το ελάχιστο vertex cover σε πολυωνυμικό χρόνο.

3.2.1 Εισαγωγικές έννοιες

Ένα δένδρο είναι ένας μη κατευθυντικός γράφος $G = (V, E)$ που είναι συνεκτικός και δεν έχει κύκλους.

3.2.2 Αλγόριθμος

Η βασική ιδέα του αλγορίθμου είναι η εξής: χρησιμοποιόντας την αναζήτηση πρώτα σε βάθος βρίσκουμε όλα τα φύλλα του δένδρου και έπειτα για κάθε φύλλο επιλέγουμε τον πατέρα του και για κάθε επιλέγουμε κάθε εσωτερικό κόμβο που τα παιδιά του δεν έχουν επιλεχθεί μέχρι να μην υπάρχουν άλλοι κόμβοι να επιλεχθούν.

3.3 Hypergraphs

3.3.1 Εισαγωγικές έννοιες

Υπεργράφος είναι μια γενίκευση του γράφου όπου μια ακμή μπορεί να συνδέσει περισσότερους από έναν κόμβους. Πιο αυστηρά ένας υπεργράφος H είναι ένα ζευγάρι $H = (X, E)$ όπου X είναι ένα σύνολο του οποίου τα στοιχεία είναι οι κόμβοι και E είναι ένα σύνολο αποτελούμενο από μη κενά υποσύνολα του X και ονομάζονται υπερακμές.

3.4 Δυικά προβλήματα

3.4.1 Clique

Ένα clique ενός μη κατευθυντικού γράφου $G = (V, E)$ είναι ένα υποσύνολο των ακμών $C \subseteq V$ τέτοιο ώστε όλοι οι κόμβοι του να είναι γειτονικοί ανά δύο.

Δεδομένου ενός μη κατευθυντικού γράφου $G = (V, E)$ ορίζουμε το συμπλήρωμα του ως $\overline{G} = (V, \overline{E})$ όπου $\overline{E} = \{(u, v) : u, v \in V, u \neq v, (u, v) \notin E\}$, τότε αν υπάρχει ένα clique C στον $\overline{G}$ το $V \setminus C$ είναι ένα vertex cover του G.

3.4.2 Independent set

Chapter 4

Εφαρμογες