Οικονομικό Πανεπιστήμιο, Ειδικά Θέματα Αλγορίθμων
1η-2η Σειρά Ασκήσεων, Χειμερινό Εξάμηνο 2021-2022
Διδάσκων: Γεώργιος Ζώης
Δημοσίευση: 16 Δεκεμβρίου 2021
Παράδοση: 28 Ιανουαρίου 2021

Η ημερομηνία παράδοσης είναι ανελαστική.
[Η παράδοση μπορεί να γίνει μόνο ηλεκτρονικά, με email στη διεύθυνση georzois@aueb.gr]

Άσκηση 1

Θεωρείστε το ακόλουθο πρόβλημα δημοπρασίας. Δοθέντος ενός συνόλου παικτών $P = 1, 2, \ldots, n$ και ενός συνόλου αντικειμένων M, κάθε παίκτης $p \in P$ πλειοδοτεί για την απόκτηση ενός υποσυνόλου αντικειμένων $M_p \subseteq M$ με μία αξία $v_p \geq 0$. Ο δημοπράτης έχει στόχο να κατανείμει τα αντικείμενα σε ένα υποσύνολο παικτών $W \subseteq P$, έτσι ώστε τα πλειοδοτούμενα υποσύνολα αντικειμένων παικτών του W να είναι ξένα μεταξύ τους, δηλ. $M_i \cap M_j = \emptyset, i \neq j \in W$, και η συνολική αξία των παικτών του W, $\sum_{p \in W} v_p$, να είναι η μέγιστη δυνατή. Να αποδείξετε ότι το παραπάνω πρόβλημα είναι NP- complete.

Άσκηση 2

(*i*). Σας δίνονται οι θετικοί ακέραιοι a, b, c και ο πρώτος αριθμός p. Δώστε έναν αλγόριθμο πολυωνυμικού χρόνου για τον υπολογισμό του $a^{b^c} \mod p$.

(*ii*). Η Αλίκη και τρείς φίλοι της είναι χρήστες του RSA κρυπτοσυστήματος. Οι φίλοι της Αλίκης έχουν public keys $(N_1, e_1), (N_2, e_2)$ και (N_3, e_3) αντίστοιχα, όπου $e_1 = e_2 = e_3 = 3$ και καθένα από τα $N_i, i = 1, 2, 3$, ως συνήθως ισούται με το γινόμενο δύο τυχαία επιλεγμένων πρώτων αριθμών μεγέθους n bits ο καθένας. Υποθέστε ότι η Αλίκη στέλνει το ίδιο μήνυμα M μεγέθους n bits (κρυπτογραφημένο με χρήση του RSA) σε καθέναν από τους τρείς φίλους της. Αποδείξτε ότι αν κάποιος υποκλέψει και τα τρία κρυπτογραφημένα μηνύματα, τότε μπορεί με αποδοτικό τρόπο να αποκαλύψει το M.

Άσκηση 3

Μία εταιρεία βιοτεχνολογίας ασχολείται με το πρόβλημα της συνένωσης ακολουθιών γενετικού υλικού. Πιο συγκεκριμένα, αν X, Y είναι δύο ακολουθίες με σύμβολα ενός αλφαβήτου A, τότε η XY (δηλ. η X ακολουθούμενη από την Y) είναι η ακολουθία γενετικού υλικού που προκύπτει από τη συνένωσή

τους. Η εταιρεία έχει εντοπίσει μία ακολουθία γενετικού υλικού Γ, η οποία αποτελείται από n σύμβολα (από το αλφάβητο A) και έχει στόχο την παραγωγή μίας ακολουθίας που να προσομοιάζει (όσο το δυνατόν περισσότερο) το Γ. Για το σκοπό αυτό θα χρησιμοποιήσει μία βάση δεδομένων B η οποία περιλαμβάνει ένα σύνολο από m μικρότερες ακολουθίες, μεγέθους το πολύ k η καθεμία. Κάνοντας χρήση αντιγράφων των ακολουθιών της βάσης B, οι υπάλληλοι της εταιρείας μπορούν να παράξουν οποιαδήποτε ακολουθία γενετικού υλικού. Ορίζουμε *συνένωση ακολουθιών* της βάσης B, οποιαδήποτε ακολουθία της μορφής $S_1 S_2 \dots S_l$, όπου κάθε $S_i \in B$.[1]

Καλείστε λοιπόν να συμβουλέψετε τους υπεύθυνους της εταιρείας δίνοντας λύση στο πρόβλημα εύρεσης μίας συνένωσης ακολουθιών της βάσης B η οποία να αντιστοιχεί στην ακολουθία γενετικού υλικού Γ, και να έχει το μικρότερο δυνατό κόστος αντιστοίχισης (ή κόστος ευθυγράμμισης). Θεωρείστε ότι το *κόστος αντιστοίχισης* ενός οποιουδήποτε τμήματος $[x, y]$ της Γ (δηλ. της υποακολουθίας συμβόλων της ακολουθίας Γ που αρχίζει με το σύμβολο x και τελειώνει με το y) με μία οποιαδήποτε ακολουθία της βάσης B είναι γνωστό εκ των προτέρων.

Άσκηση 4

Στο πρόβλημα εύρεσης του Ελάχιστου Steiner Δένδρου (Minimum Steiner Tree) η είσοδος αποτελείται από ένα πλήρες γράφημα $G = (V, E)$ με θετικές αποστάσεις $d_{u,v} > 0$, ανάμεσα σε κάθε ζεύγος κορυφών $\{u, v\} \in E$, και ένα προκαθορισμένο σύνολο *τερματικών κορυφών* $V' \subseteq V$. Στόχος είναι η εύρεση ενός δένδρου ελαχίστου κόστους (δηλαδή ελαχίστου αθροίσματος των αποστάσεων των ακμών του) το οποίο να περιέχει όλες τις τερματικές κορυφές του συνόλου V', χωρίς όμως να αποκλείεται η συμμετοχή και κορυφών από το σύνολο $V - V'$. Θεωρείστε ότι οι αποστάσεις των κορυφών είναι μετρικές, δηλαδή ισχύει η τριγωνική ανισότητα: $d(u, k) \leq d(u, v) + d(v, k),\ 1 \leq u, v, k \leq |V|$, και αποδείξτε ότι ο αλγόριθμος που επιστρέφει ένα Ελάχιστο Σκελετικό Δένδρο (Minimum Spanning Tree) πάνω στις τερματικές κορυφές V', είναι ένας 2-προσεγγιστικός αλγόριθμος για το πρόβλημα του Ελάχιστου Steiner Δένδρου.

Άσκηση 5

Θεωρείστε το ακόλουθο πρόβλημα: Δίνεται ένα σύνολο (σύμπαν) U από n στοιχεία, καθένα από τα οποία διαθέτει ένα μη αρνητικό βάρος. Δίνεται ακόμη μία συλλογή $S = \{S_1, S_2, \dots, S_l\}$ από υποσύνολα του U και ένας ακέραιος $k \leq l$, και σας ζητείται να διαλέξετε k σύνολα από τη συλλογή S τέτοια που το συνολικό βάρος των καλυπτόμενων στοιχείων του U να είναι το μέγιστο δυνατό. Έστω ο ακόλουθος άπληστος αλγόριθμος:

[1] Επιτρέπονται οι επαναλήψεις ακολουθιών, δηλ. η S_i και η S_j, όπου $i \neq j$, μπορεί να είναι η ίδια ακολουθία.

Όσο δεν έχουν επιλεγεί k σύνολα από το S, πρόσθεσε στη λύση το καλύτερο σύνολο (best set).[2]

Αποδείξτε ότι ο αλγόριθμος αυτός πετυχαίνει λόγο προσέγγισης

$$1 - (1 - \frac{1}{k})^k > 1 - \frac{1}{e}.$$

Άσκηση 6

Θεωρείστε τις ακόλουθες τέσσερις παραλλαγές του προβλήματος Maximum Flow σε ένα γράφημα $G = (V, E)$ (βλ. ορισμό του προβλήματος στα σετ διαφανειών 11 και 13 του μαθήματος, σελ. 4).
(i) Στον G υπάρχουν πολλές πηγές και πολλοί προορισμοί και επιθυμούμε να μεγιστοποιήσουμε το συνολικό Flow από όλες τις πηγές προς όλους τους προορισμούς.
(ii) Κάθε κορυφή $v \in V$ διαθέτει επίσης μία τιμή-χωρητικότητα που περιορίζει το μέγιστο Flow που μπορεί να διέλθει από την v.
(iii) Κάθε ακμή $e \in E$ διαθέτει (επιπλέον της χωρητικότητας) και ένα κάτω φράγμα ως προς το Flow που μπορεί να διέλθει από την e.
(iv) Το εξερχόμενο Flow, από κάθε κορυφή $v \in V$, δεν είναι ίσο με το εισερχόμενο Flow στην v, αλλά μικρότερο κατά ένα παράγοντα $(1 - \epsilon_v)$, όπου ϵ_v είναι ένας συντελεστής απώλειας που συνοδεύει τον κόμβο v.

Αποδείξτε ότι καθεμία από τις παραπάνω παραλλαγές του Maximum Flow μπορεί να λυθεί βέλτιστα σε πολυωνυμικό χρόνο. Πιο συγκεκριμένα, αποδείξτε ότι οι $(i), (ii)$ ανάγονται στη βασική εκδοχή του προβλήματος Maximum Flow, ενώ οι $(iii), (iv)$ μπορούν να αναχθούν στο πρόβλημα του Γραμμικού Προγραμματισμού (LP).

Άσκηση 7

(i) Διατυπώστε το πρόβλημα MAXIMUM INDEPENDENT SET ως πρόβλημα ακέραιου προγραμματισμού (ILP). Γράψτε επίσης την χαλάρωσή του προβλήματος γραμμικού προγραμματισμού (LP relaxation). Εξηγείστε αναλυτικά τις μεταβλητές, τους περιορισμούς και την αντικειμενική συνάρτηση της διατύπωσής σας. Απαντήστε επίσης στο παρακάτω ερώτημα:

(a) Το χάσμα ακεραιότητας (integrality gap) ενός στιγμιοτύπου του προβλήματος ορίζεται ως ο λόγος της βέλτιστης ακέραιας λύσης δια της βέλτιστης λύσης της γραμμικής χαλάρωσης σε πρόβλημα LP. Δώστε ένα στιγμιότυπο του MAXIMUM INDEPENDENT SET για το οποίο το χάσμα ακεραιότητας για τη χαλάρωση που διατυπώσατε παραπάνω είναι μικρότερο ή ίσο με $2/n$.

[2] Πρόκειται για το σύνολο με τα περισσότερο βεβαρημένα ακάλυπτα στοιχεία.

(ii) Έστω ο εξής αλγόριθμος για την εύρεση του MAXIMUM INDEPENDENT SET, S, ενός γράφου $G = (V, E)$:

RANDOM(G);
Get a permutation π of V uniformly at random;
Find a subset $S(\pi) \subseteq V$ as follows:
For each vertex $u \in V$:
$u \in S(\pi)$ if and only if no neighbor of u precedes u in the permutation π.
Return $S(\pi)$;

Για το υποσύνολο $S(\pi)$ αποδείξτε ότι *(a)* είναι ανεξάρτητο υποσύνολο, και *(b)* η αναμενόμενη τιμή του πληθικού αριθμού του είναι $E[|S(\pi)|] = \sum_{i=1}^{n} \frac{1}{d_i+1}$, όπου d_i ο βαθμός της κορυφής i.

Άσκηση 8

Ένας παραγωγός κινηματογραφικών ταινιών αναζητά ηθοποιούς και επενδυτές για τη νέα του ταινία. Υπάρχουν n διαθέσιμοι ηθοποιοί, και m διαθέσιμοι επενδυτές. Ο κάθε ηθοποιός i χρεώνει v_i δολάρια ενώ ο κάθε επενδυτής j πληρώνει p_j δολάρια υπό την προϋπόθεση να συμπεριληφθούν στο cast της ταινίας όλοι οι ηθοποιοί του συνόλου $L_j \subseteq \{1, 2, \ldots, n\}$. Το όφελος του παραγωγού ισούται με το άθροισμα των πληρωμών των επενδυτών μείον τις χρεώσεις των ηθοποιών, και ο στόχος του παραγωγού είναι να μεγιστοποιήσει το όφελός του. Καλείστε να απαντήσετε στα ακόλουθα δύο ερωτήματα:

1. Διαμορφώστε ένα ακέραιο γραμμικό πρόγραμμα για το πρόβλημα αυτό, στο οποίο οι μεταβλητές παίρνουν τιμές $\{0, 1\}$.

2. Θεωρείστε την χαλάρωσή του προβλήματος γραμμικού προγραμματισμού (LP relaxation) και αποδείξτε ότι η βέλτιστη λύση που επιστρέφει είναι ακέραια λύση (όπως ακριβώς και στην περίπτωση του Maximum Matching σε διμερή γραφήματα (bipartite graphs).

Άσκηση 9

Θεωρείστε το πρόβλημα Coin Changing (βλ. Σετ Διαφανειών 4, σελ. 23).
(i) Διαμορφώστε ένα ακέραιο γραμμικό πρόγραμμα για το πρόβλημα αυτό.
(ii) Μπορούμε να λύσουμε το πρόβλημα αυτό ως ένα LP (μέσω LP relaxation) με τη βεβαιότητα ότι η επιστρεφόμενη λύση θα είναι ακέραια (όπως ακριβώς και στην περίπτωση του Maximum Matching σε διμερή γραφήματα (bipartite graphs». Αν όχι, τότε δώστε ένα αντιπαράδειγμα.

Άσκηση 10

Θεωρείστε τη weighted εκδοχή του προβλήματος MAX-SAT όπου επιπλέον έχουμε ένα θετικό βάρος σε κάθε συνθήκη (clause) και θέλουμε να μεγιστοποιήσουμε το συνολικό βάρος των συνθηκών που ικανοποιούνται. Θεωρείστε

επίσης τον Randomized LP-rounding αλγόριθμο για το MAX-SAT, από το Σετ Διαφανειών 12, σελ. 41.
(*i*) Αποδείξτε ότι ο παραπάνω αλγόριθμος πετυχαίνει λόγο προσέγγισης $1 - \frac{1}{e}$ (in expectation) για το weighted MAX-SAT.
(*ii*) Θεωρείστε τώρα τον πιθανοτικό αλγόριθμο για το MAX-SAT που πετυχαίνει παράγοντα προσέγγισης $\frac{1}{2}$ (in expectation) (βλ. Σετ Διαφανειών 10, σελ. 5). Αποδείξτε ότι ο αλγόριθμος που συνδυάζει τους παραπάνω δύο πιθανοτικούς αλγορίθμους, εκτελώντας είτε τον πρώτο είτε τον δεύτερο με πιθανότητα $\frac{1}{2}$, πετυχαίνει παράγοντα προσέγγισης $\frac{3}{4}$ (in expectation).