Προηγμένα Θέματα Αλγορίθμων (ΣΗΜΜΥ ΕΜΠ)
Αλγόριθμοι Δικτύων και Πολυπλοκότητα (ΕΜΠ - ΑΛΜΑ)

Προσεγγιστικοί Αλγόριθμοι

Άρης Παγουρτζής

Άνοιξη 2018

Προσεγγιστικοί Αλγόριθμοι

- Αφορούν κυρίως σε προβλήματα βελτιστοποίησης: σε κάθε στιγμιότυπο αντιστοιχούν εφικτές (αποδεκτές / έγκυρες) λύσεις (feasible solutions) που η κάθε μια έχει μια τιμή μέσω μιας αντικειμενικής συνάρτησης (objective function) (συνήθως εκφράζει: κόστος, μήκος, βάρος, κ.λπ.). Ζητείται βέλτιστη λύση, δηλαδή εφικτή λύση με βέλτιστη τιμή.
- Προβλήματα ελαχιστοποίησης (minimization) / προβλήματα μεγιστοποίησης (maximization)
- Ενδιαφερόμαστε κυρίως για **NP**-optimization προβλήματα. Κλάση **NPO**: το αντίστοιχο πρόβλημα απόφασης ανήκει στην κλάση **NP**.

Συμβολισμοί

- Π: πρόβλημα βελτιστοποίησης
- I: στιγμιότυπο (είσοδος) του προβλήματος
- $SOL_A(\Pi, I)$: η τιμή της λύσης που επιστρέφει ο αλγόριθμος A για το στιγμιότυπο I του προβλήματος Π.
- $OPT(\Pi, I)$: η τιμή της βέλτιστης λύσης για το στιγμιότυπο I του προβλήματος Π.

Σημείωση: *Συχνά* Π, *A και I παραλείπονται.*

Προσεγγισιμότητα: προβλήματα ελαχιστοποίησης

Ενας αλγόριθμος A για ένα πρόβλημα ελαχιστοποίησης Π επιτυγχάνει λόγο προσέγγισης (approximation ratio) ρ, και λέγεται ρ-προσεγγιστικός, αν για κάθε έγκυρο στιγμιότυπο I:

$$\frac{SOL_A(I)}{OPT(I)} \leq \rho$$

- Λόγος προσέγγισης αλγορίθμου: Το ελάχιστο ρ (τυπικά infimum) που ικανοποιεί την παραπάνω σχέση για κάθε στιγμιότυπο I λέγεται λόγος (ή παράγοντας) προσέγγισης του αλγορίθμου A.
- Λόγος προσέγγισης προβλήματος: Αν για πρόβλημα Π υπάρχει ρ-προσεγγιστικός αλγόριθμος, λέμε ότι το Π προσεγγίζεται με λόγο (ή παράγοντα) ρ. Μας ενδιαφέρει το ελάχιστο ρ μεταξύ όλων των δυνατών προσεγγιστικών αλγορίθμων για το Π.

Προσοχή: Συνήθως ο όρος προσεγγιστικός αλγόριθμος αναφέρεται σε αλγορίθμους πολυωνυμικού χρόνου ως προς το μέγεθος εισόδου $|I|$.

Προσεγγισιμότητα: προβλήματα μεγιστοποίησης

- Προβλήματα μεγιστοποίησης: ο αλγόριθμος A λέγεται ρ-προσεγγιστικός για το Π αν για κάθε I:

$$\frac{SOL_A(I)}{OPT(I)} \geq \rho$$

(δηλαδή $\rho \leq 1$ για προβλήματα μεγιστοποίησης)

- Λόγος προσέγγισης αλγορίθμου: το μέγιστο ρ (τυπικά supremum) που ικανοποιεί την παραπάνω σχέση (για κάθε I).
- Λόγος προσέγγισης προβλήματος: το μέγιστο ρ μεταξύ όλων των αλγορίθμων που το επιλύουν.

Προσεγγισιμότητα: εναλλακτικοί ορισμοί

- Εναλλακτικός ορισμός του λόγου προσέγγισης, κοινός για προβλήματα ελαχιστοποίησης και μεγιστοποίησης:
  Ενας αλγόριθμος A για ένα πρόβλημα ελαχιστοποίησης/μεγιστοποίησης Π επιτυγχάνει λόγο προσέγγισης (approximation ratio) ρ, και λέγεται ρ-προσεγγιστικός, αν για κάθε έγκυρο στιγμιότυπο I:

  $$\max\{\frac{SOL_A(I)}{OPT(I)}, \frac{OPT(I)}{SOL_A(I)}\} \leq \rho$$

  Με βάση τον ορισμό αυτό ισχύει πάντοτε $\rho \geq 1$. Ακολουθείται σε ορισμένα βιβλία, εμείς θα χρησιμοποιήσουμε τους προηγούμενους ορισμούς.
- Παλαιότερος ορισμός εξετάζει το *σχετικό σφάλμα*. Ένας αλγόριθμος A έχει σχετικό σφάλμα προσέγγισης ϵ αν $\forall I$:

  $$\frac{|SOL_A(I) - OPT(I)|}{OPT(I)} \leq \epsilon$$

Προσεγγισιμότητα: άλλες έννοιες

- Ο λόγος προσέγγισης στην γενική περίπτωση είναι *συνάρτηση* του μεγέθους (μήκους) της εισόδου:

$$\forall I : \quad \frac{SOL_A(I)}{OPT(I)} \leq \rho(|I|) \quad (\geq \quad \text{για max})$$

- Ασυμπτωτικός λόγος προσέγγισης: η ανισότητα ισχύει $\forall || \geq n_0$.
- **APX**: κλάση προβλημάτων που επιλύονται με σταθερό λόγο προσέγγισης σε πολυωνυμικό χρόνο. Υποκλάση της **NPO**.
- Πιθανοτικοί προσεγγιστικοί αλγόριθμοι: ο λόγος προσέγγισης επιτυγχάνεται *με μεγάλη πιθανότητα* (τυπικοί ορισμοί αργότερα).

Κλάσεις προσεγγισιμότητας

- **PTAS**: κλάση προβλημάτων που επιλύονται με οποιοδήποτε σταθερό λόγο προσέγγισης $1+\varepsilon$ (μεγιστ/σης: $1-\varepsilon$) σε πολυωνυμικό χρόνο ως προς $|I|$. Υποκλάση της **APX**.
- **FPTAS**: κλάση προβλημάτων που επιλύονται με οποιοδήποτε σταθερό λόγο προσέγγισης $1+\varepsilon$ (μεγιστ/σης: $1-\varepsilon$) σε πολυωνυμικό χρόνο ως προς $|I|$ και $\frac{1}{\varepsilon}$. Υποκλάση της **PTAS**.
- $\log -$**APX**: κλάση προβλημάτων που επιλύονται με λογαριθμικό (ως προς $|I|$) λόγο προσέγγισης σε πολυωνυμικό χρόνο. Υπερκλάση της **APX**.
- **poly − APX**: κλάση προβλημάτων που επιλύονται με πολυωνυμικό (ως προς $|I|$) λόγο προσέγγισης σε πολυωνυμικό χρόνο. Υπερκλάση της $\log -$**APX**.

FP $\subseteq$ **FPTAS** $\subseteq$ **PTAS** $\subseteq$ **APX** $\subseteq$ $\log -$**APX** $\subseteq$ **poly − APX** $\subseteq$ **NPO**

Το πρόβλημα Vertex Cover (VC)

Δίνεται: γράφος $G(V, E)$
Ζητείται: Κάλυψη κορυφών (vertex cover) ελάχιστης πληθικότητας, δηλαδή ελάχιστο σύνολο κόμβων V' έτσι ώστε κάθε ακμή να έχει τουλάχιστον έναν από τους κόμβους της στο V'.

Weighted Vertex Cover (WVC): οι κόμβοι έχουν και βάρος και το ζητούμενο είναι το σύνολο V' να είναι ελαχίστου βάρους.

Σημείωση: *Συχνά ο όρος* Vertex Cover *χρησιμοποιείται για την weighted εκδοχή του προβλήματος. Τότε για το πρώτο πρόβλημα χρησιμοποιούνται οι όροι "Unweighted" ή "Cardinality".*

Προσεγγιστικός αλγόριθμος για το Vertex Cover

```
Βρες maximal matching MM με οποιαδήποτε
μέθοδο.
Επίστρεψε το σύνολο V' των κόμβων που
προσπίπτουν στις ακμές του MM.
```

Θεώρημα

Ο παραπάνω αλγόριθμος είναι 2-προσεγγιστικός.

Απόδειξη.

1. Το V' είναι εφικτή λύση (γιατί;)
2. $|MM| \leq OPT$
3. $SOL = 2 \cdot |MM| \leq 2 \cdot OPT$ $\square$

Ανελαστική Ανάλυση (Tight Analysis)

Η παραπάνω ανάλυση είναι ανελαστική (tight), δηλαδή αποδεικνύεται ότι ο αλγόριθμος δεν μπορεί να πετύχει καλύτερο λόγο.

Η απόδειξη συνήθως συνίσταται στη εύρεση ανελαστικού παραδείγματος (tight example): άπειρη οικογένεια στιγμιοτύπων για τα οποία ο αλγόριθμος δεν μπορεί να πετύχει καλύτερο λόγο.

Για τον προηγούμενο αλγόριθμο για το VC ένα tight example είναι οι πλήρεις διμερείς γράφοι $K_{n,n}$.

Ανελαστικότητα του κάτω φράγματος ως προς το OPT

Ένα διαφορετικό είδος ανελαστικότητας (tightness): αφορά στο λόγο της αντικειμενικής τιμής της βέλτιστης λύσης προς το κάτω φράγμα που χρησιμοποιούμε για το OPT.

Παράδειγμα: για το Vertex Cover, εξετάζοντας πλήρεις γράφους K_n αποδεικνύεται ότι το μέγεθος του maximal matching ως κάτω φράγμα για το OPT δεν μπορεί να δώσει καλύτερο λόγο προσέγγισης από 2.

Ενδεικτικά, δεν μπορεί να υπάρχει αλγόριθμος που να βρίσκει, σε κάθε γράφο, vertex cover μεγέθους $\leq \frac{3}{2}|MM|$.

Φαινομενικά παράδοξο: ο 2-προσεγγιστικός αλγόριθμος επιτυγχάνει βέλτιστη (ή σχεδόν βέλτιστη) λύση για κάθε γράφο K_n. Η ανελαστικότητα αυτή μας δίνει μια αξιολόγηση του κάτω φράγματος και *όχι κάποιου αλγορίθμου*.

Προσεγγίζοντας το Weighted Vertex Cover (WVC)

Ορισμός

Degree-weighted συνάρτηση $w : V \rightarrow \mathbf{Q}$ *(σε γράφο* $G(V, E)$*):*
$\exists c > 0, \forall v \in V, w(v) = c \cdot deg(v)$, *όπου* $deg(v)$ *είναι ο βαθμός του κόμβου* v.

Λήμμα

Αν η συνάρτηση βαρών w *σε γράφο* $G(V, E)$ *είναι degree-weighted τότε:* $w(V) \leq 2 \cdot OPT_{WVC}$.

Σημείωση: $w(V) = \sum_{v \in V} w(v)$.

Απόδειξη.

Αν U είναι vertex cover τότε $deg(U)(=\sum_{u \in U} deg(u)) \geq |E|$.
Οπότε $deg(V)(=\sum_{v \in V} deg(v)) = 2|E| \leq 2deg(U)$.
Επομένως, αν w είναι degree-weighted: $w(V) \leq 2w(U)$.
Αυτό ισχύει και για vertex cover U^* ελαχίστου βάρους. Αρα:

$$w(V) \leq 2w(U^*) = 2OPT_{WVC}$$

□

Ερώτηση: τι συμπεραίνουμε από το παραπάνω λήμμα;

Προσεγγιστικός αλγόριθμος για το Weighted Vertex Cover (WVC)

Βασίζεται σε κατάλληλη διάσπαση (decomposition) της δοθείσης συνάρτησης βάρους σε degree-weighted συναρτήσεις.

```
Επανάλαβε όσο υπάρχουν κόμβοι στο γράφο
   αφαίρεσε κόμβους μηδενικού βαθμού
   βρές μέγιστο c τ.ώ. ∀v ∈ Vc · deg(v) ≤ w(v)
      (όπου w η τρέχουσα συνάρτηση βάρους)
   ∀v ∈ V θέσε w(v) ← w(v) − c · deg(v)
   πρόσθεσε κόμβους βάρους 0 στην κάλυψη
      και αφαίρεσέ τους από το γράφο.
```

Θεώρημα

Ο παραπάνω αλγόριθμος είναι 2-προσεγγιστικός για το πρόβλημα WVC.

Το πρόβλημα Set Cover

Δίνεται: *σύνολο* U *με* n *στοιχεία και συλλογή υποσυνόλων του* U,
$\mathcal{S} = \{S_1, \ldots, S_k\}, \forall i, S_i \subseteq U$
Ζητείται: Μία ελάχιστης πληθικότητας *συλλογή* $\mathcal{S}' \subseteq \mathcal{S}$ *τ.ώ. κάθε στοιχείο του* U *να ανήκει σε τουλάχιστον ένα σύνολο της* $\mathcal{S}'$*:*
$\bigcup_{S \in \mathcal{S}'} S = U$.

Weighted Set Cover: τα υποσύνολα έχουν βάρος (κόστος) και το ζητούμενο είναι η $\mathcal{S}'$ να είναι ελαχίστου βάρους.

Σημείωση: *Συχνά ο όρος* Set Cover *χρησιμοποιείται για την weighted εκδοχή του προβλήματος. Τότε για το πρώτο πρόβλημα χρησιμοποιούνται οι όροι "Unweighted" ή "Cardinality".*

Ο άπληστος αλγόριθμος για το Set Cover

Αλγόριθμος SC-Greedy

$C := \emptyset$

Επανάλαβε όσο $C \neq U$:

> Βρές S_i με μέγιστο πλήθος στοιχείων

> $C := C \cup S_i$

Θεώρημα

Ο άπληστος αλγόριθμος SC-Greedy για το Set Cover *είναι* $\Theta(\log n)$*-προσεγγιστικός.*

Απόδειξη: σε κάθε $k = OPT$ επαναλήψεις καλύπτονται τουλάχιστον τα μισά (ακάλυπτα) στοιχεία του U.

Tight example για τον Greedy αλγόριθμο για το πρόβλημα Set Cover

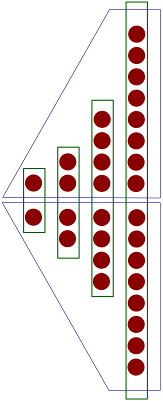
$U = \{a_1, \ldots, a_m\} \cup \{a'_1, \ldots, a'_m\}, \quad m = 2^k - 1$

$S_i = \{a_{2^{i-1}}, \ldots a_{2^i - 1}\} \cup \{a'_{2^{i-1}}, \ldots a'_{2^i - 1}\}, \; i = 1, \ldots, k$

$S_{k+1} = \{a_1, \ldots a_m\}$

$S_{k+2} = \{a'_1, \ldots a'_m\}$

$SOL = k \approx \log n, \quad OPT = 2$



Ο άπληστος αλγόριθμος για το πρόβλημα Weighted Set Cover

Αλγόριθμος Weighted SC-Greedy

$C := \emptyset$

Επανάλαβε όσο $C \neq U$:

> Βρές S_i με ελάχιστο $\alpha_i = cost(S_i)/|S_i \setminus C|$

> $\forall e \in S_i \setminus C$ θέσε $price(e) \leftarrow \alpha_i$

> $C := C \cup S_i$

Σημείωση: *$price(e_k)$ είναι η τιμή που "πληρώσαμε" για να καλυφθεί το στοιχείο e_k.*

Συνολικό κόστος κάλυψης: $SOL = \sum_{k=1}^{n} price(e_k)$.

Ανάλυση του Weighted SC-Greedy

Θεώρημα

Ο αλγόριθμος Weighted SC-Greedy είναι H_n-προσεγγιστικός, όπου $H_n = 1 + \frac{1}{2} + \ldots \frac{1}{n} < \ln n + 1$.

Απόδειξη.

Υποθέτουμε ότι τα στοιχεία του U αριθμούνται με τη σειρά που καλύπτονται από τον αλγόριθμο. Παρατηρούμε ότι:

- Σε οποιαδήποτε επανάληψη του αλγορίθμου μπορούμε να καλύψουμε όλα τα εναπομείναντα στοιχεία με κόστος το πολύ OPT (γιατί;).
- Επομένως υπάρχει πάντα σύνολο με κόστος/νέο στοιχείο το πολύ $OPT/|U - C|$ (C: η τρέχουσα κάλυψη).
- Πριν καλυφθεί το στοιχείο e_k για πρώτη φορά ισχύει $|U - C| \geq n - k + 1$. Άρα $price(e_k) \leq \frac{OPT}{n-k+1}$.

Συνολικό κόστος το πολύ: $\Sigma_{k=1}^{n} \frac{OPT}{n-k+1} = H_n \cdot OPT$ $\square$

Tight example

$U = \{a_1, \ldots, a_n\}$

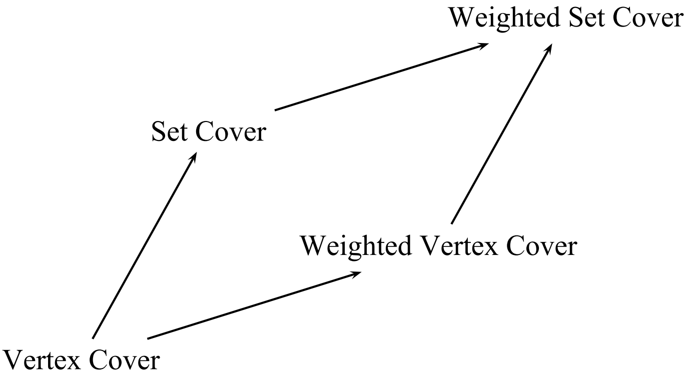
$\forall i \in \{1, \ldots, n\}, S_i = \{a_i\}, cost(S_i) = \frac{1}{i}$

$S_{n+1} = U, cost(S_{n+1}) = 1 + \varepsilon$, για σταθερά $\varepsilon > 0$

$SOL = H_n, \quad OPT = 1 + \varepsilon$, επομένως $\rho(n) = \Omega(H(n)) = \Omega(\log n)$.

Σημείωση: *Αποδεικνύεται ότι ο αλγόριθμος έχει το ίδιο κάτω φράγμα λόγου προσέγγισης και για το* (Cardinality) Set Cover *και για το* Vertex Cover*!*

Ιεράρχηση Προβλημάτων



Πρόβλημα μεγιστοποίησης: Maximum Coverage

Δίνεται: σύνολο U με n στοιχεία, συλλογή $\mathcal{S} = \{S_1, \ldots, S_l\}$ υποσυνόλων του U, και ακέραιος k.
Ζητείται: Μία συλλογή $\mathcal{S}' \subseteq \mathcal{S}$ αποτελούμενη από k σύνολα τ.ώ. το πλήθος των στοιχείων που καλύπτει η $\mathcal{S}'$ να είναι μέγιστο.

Θεώρημα

Ο άπληστος (greedy) αλγόριθμος που κάθε φορά επιλέγει το μεγαλύτερο, ως προς πλήθος νέων στοιχείων, σύνολο επιτυγχάνει (με k επαναλήψεις) λόγο προσέγγισης

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

για το πρόβλημα Maximum Coverage.

Σημείωση: *Το πρόβλημα ανήκει στην κλάση* **APX**, *παρ'ότι το* Set Cover *δεν ανήκει.*

Απόδειξη (i)

Έστω $\mathcal{S}^*$ μια βέλτιστη λύση με πλήθος στοιχείων n_{OPT}. Αφού το πλήθος των συνόλων της $\mathcal{S}^*$ είναι k θα πρέπει να υπάρχει σύνολο στην $\mathcal{S}$ με πληθικότητα $\geq \frac{n_{OPT}}{k}$. Επομένως, ο άπληστος αλγόριθμος στο 1ο βήμα θα βρεί και θα επιλέξει σύνολο τουλάχιστον τόσων στοιχείων.
Με άλλα λόγια, στο 1ο βήμα, καλύπτεται πλήθος στοιχείων που αντιπροσωπεύουν τουλάχιστον το $\frac{1}{k}$ του n_{OPT} ή, ισοδύναμα, μένει ακάλυπτο το πολύ ένα μέρος $1 - \frac{1}{k}$ των στοιχείων της βέλτιστης λύσης (ένωση των συνόλων της $\mathcal{S}^*$).

Σημείωση: *Κάθε φορά που ο άπληστος αλγόριθμος καλύπτει στοιχεία που δεν ανήκουν στη βέλτιστη λύση μπορούμε, για τις ανάγκες της απόδειξης και μόνο, να "σβήνουμε" ισάριθμο πλήθος στοιχείων από τη βέλτιστη λύση, θεωρώντας ότι το αντίστοιχο μέρος έχει καλυφθεί.*

Απόδειξη (ii)

Με ανάλογα επιχειρήματα μπορούμε να δείξουμε ότι στο i-οστό βήμα, το ακάλυπτο μέρος της βέλτιστης λύσης μειώνεται τουλάχιστον κατά $\frac{1}{k}$ (αφού τα k σύνολα της $\mathcal{S}^*$ αρκούν για να το καλύψουν πλήρως) και επομένως απομένει ακάλυπτο το πολύ ένα μέρος $(1-\frac{1}{k})^i$ της βέλτιστης λύσης.
Τελικά, σε k βήματα έχει καλυφθεί τουλάχιστον το $1-(1-\frac{1}{k})^k$ της βέλτιστης λύσης. Επομένως:

$$SOL \geq (1-(1-\frac{1}{k})^k)n_{OPT} > (1-\frac{1}{e})n_{OPT}$$

Γενικεύσεις

- Το αποτέλεσμα ισχύει και στην περίπτωση που τα στοιχεία έχουν βάρος και ζητείται λύση που καλύπτει μέγιστο βάρος στοιχείων.
- Το αποτέλεσμα ισχύει ακόμη και σε περιπτώσεις όπου δεν δίνονται τα σύνολα της $\mathcal{S}$ αναλυτικά (μπορεί να είναι και εκθετικά πολλά), αλλά μπορούμε να βρίσκουμε το "καλύτερο" σύνολο σε πολυωνυμικό χρόνο.
- Εάν μπορούμε να βρίσκουμε το "καλύτερο" σύνολο σε πολυωνυμικό χρόνο με προσέγγιση ρ (≤ 1) τότε ο άπληστος αλγόριθμος επιτυγχάνει λόγο προσέγγισης
$$1-\left(1-\frac{\rho}{k}\right)^k > 1-\frac{1}{e^\rho}$$
για το πρόβλημα Maximum Coverage.

Το Πρόβλημα του Πλανόδιου Πωλητή
Traveling Salesman Problem (TSP)

Ορισμός του προβλήματος

Δίνεται: *πλήρης γράφος* $G(V, E)$ *με μη αρνητικά κόστη (βάρη) στις ακμές του.*
Ζητείται: *κύκλος ελαχίστου κόστους που να επισκέπτεται κάθε κόμβο ακριβώς μία φορά (Hamilton Cycle).*

Μη-προσεγγισιμότητα του TSP

Θεώρημα

Το πρόβλημα TSP δεν μπορεί να προσεγγιστεί με παράγοντα $\alpha(n)$, όπου $n = |V|$, για οποιαδήποτε πολυωνυμικά υπολογιστή συνάρτηση $\alpha : \mathbb{N} \rightarrow \mathbb{R}$, εκτός εάν $\mathbf{P} = \mathbf{NP}$.

Απόδειξη.

Αναγωγή από το Hamilton Cycle: συμπληρώνουμε τον αρχικό γράφο G με ακμές ώστε να κατασκευάσουμε πλήρη γράφο G'. Στις αρχικές ακμές δίνουμε βάρος 1, στις υπόλοιπες δίνουμε βάρος $\alpha(n) \cdot n$. Ισχύει ότι:

- Αν ο G είναι Hamilton τότε υπάρχει κύκλος TSP κόστους n στον G', ενώ
- Αν ο G δεν είναι Hamilton τότε ο βέλτιστος κύκλος TSP στον G' έχει κόστος $> \alpha(n) \cdot n$.

$\square$

Αναγωγές Εισαγωγής Χάσματος (i)
(Gap Introducing Reductions)

Μια αναγωγή h από το πρόβλημα απόφασης Π στο πρόβλημα ελαχιστοποίησης Π' (που απεικονίζει κάθε στιγμιότυπο I του Π σε κάποιο στιγμιότυπο $I' = h(I)$ του Π') λέγεται *αναγωγή εισαγωγής χάσματος* όταν υπάρχουν συναρτήσεις f, α ώστε:

- Αν το I είναι 'yes'-instance του Π τότε $OPT_{\Pi'}(I') \leq f(I')$, ενώ
- Αν το I είναι 'no'-instance του Π τότε $OPT_{\Pi'}(I') > \alpha(|I'|) \cdot f(I')$.

Θεώρημα

Αν το πρόβλημα Π είναι **NP***-complete και υπάρχει αναγωγή εισαγωγής χάσματος με παραμέτρους f, α από το Π στο πρόβλημα Π' τότε το Π' δεν προσεγγίζεται με παράγοντα $\alpha(n)$, εφ'όσον* $\mathbf{P} \neq \mathbf{NP}$ *(όπου n το μήκος της αναπαράστασης της εισόδου του Π').*

Αναγωγές Εισαγωγής Χάσματος (ii) (Gap Introducing Reductions)

Μια αναγωγή h από το πρόβλημα απόφασης Π στο πρόβλημα μεγιστοποίησης Π' (που απεικονίζει κάθε στιγμιότυπο I του Π σε κάποιο στιγμιότυπο $I' = h(I)$ του Π') λέγεται *αναγωγή εισαγωγής χάσματος* όταν υπάρχουν συναρτήσεις f, α ώστε:

- Αν το I είναι 'yes'-instance του Π τότε $OPT_{\Pi'}(I') \geq f(I')$, ενώ
- Αν το I είναι 'no'-instance του Π τότε $OPT_{\Pi'}(I') < \alpha(|I'|) \cdot f(I')$.

Το πρόβλημα Metric TSP

Επιπλέον υπόθεση: τα δοθέντα βάρη ικανοποιούν την τριγωνική ανισότητα.
Το πρόβλημα παραμένει **NP**-complete (γιατί;)

2-προσεγγιστικός αλγόριθμος για το Metric TSP

> Βρες ελάχιστο συνδετικό δένδρο T στον γράφο G.
> Διπλασίασε τις ακμές του T.
> Βρες κύκλο Euler C στο διπλασιασμένο T.
> Δώσε σαν έξοδο τον κύκλο που επισκέπτεται τους κόμβους με τη σειρά εμφάνισής τους στον C (short-cutting).

Ο παραπάνω αλγόριθμος είναι 2-προσεγγιστικός:
$cost(C) \leq cost(C_T) \leq 2cost(T) \leq 2OPT$

Καλύτερη προσέγγιση για το Metric TSP (Christofides' algorithm)

$\frac{3}{2}$-προσεγγιστικός αλγόριθμος για το Metric TSP

Βασίζεται στην εύρεση "φτηνότερου" κύκλου Euler.
> Βρες Eulerian completion του δέντρου T, μέσω minimum cost perfect matching M πάνω στους κόμβους περιττού βαθμού του T.
> Συνέχισε όπως στον Αλγόριθμο Χριστοφίδη.

Ο παραπάνω αλγόριθμος είναι $\frac{3}{2}$-προσεγγιστικός:
$cost(M) \leq \frac{1}{2}OPT$: με short-cutting στον βέλτιστο κύκλο, παίρνουμε κύκλο στους κόμβους περιττού βαθμού του T.

$$cost(C) \leq cost(C_{T,M}) = cost(T) + cost(M) \leq \frac{3}{2}OPT$$

Το πρόβλημα Metric $\text{TSP}_{(s,t)-path}$

Ορισμός του προβλήματος

Δίνονται: *γράφος με βάρη (όπως για το* Metric TSP*) και επιπλέον 2 κόμβοι* s, t.
Ζητείται: *Hamilton path ελαχίστου κόστους από* s *σε* t.

Αλγόριθμος: αποτελείται από 2 ανεξάρτητους αλγορίθμους, ο καθένας εκ των οποίων δημιουργεί γράφο με μονοπάτι Euler με διαφορετικό τρόπο (βλ. επόμενη διαφάνεια). Η επιλογή της καλύτερης από τις 2 λύσεις δίνει $\frac{5}{3}$-προσεγγιστικό αλγόριθμο:

$$\min\{SOL_1, SOL_2\} \leq \frac{5}{3} OPT_{s,t}$$

Το πρόβλημα Metric $\text{TSP}_{(s,t)-path}$ (συν.)

$\frac{5}{3}$-προσεγγιστικός αλγόριθμος για το Metric $\text{TSP}_{(s,t)-path}$

Εκτέλεσε ανεξάρτητα (1) και (2) και επίστρεψε την καλύτερη λύση:

1. Βρες ελάχιστο συνδετικό δέντρο T στον γράφο G. Διπλασίασε τις ακμές του T. Αφαίρεσε ελάχιστο (s,t)-path από το διπλασιασμένο δένδρο. Βρες (s,t)-Euler path $P_{s,t}$, εκτέλεσε short-cutting για να βρείς (s,t)-Hamilton path κόστους:

$$SOL_1 \leq 2OPT_{s,t} - c_{s,t}$$

2. Με μικρή τροποποίηση του αλγόριθμου του Χριστοφίδη (Eulerian completion ώστε να προκύπτει (s,t)-Euler path με short-cutting), βρες (s,t)-Hamilton path με κόστος

$$SOL_2 \leq (3OPT_{s,t} + c_{s,t})/2$$

Το Πρόβλημα Steiner Tree

Ορισμός του προβλήματος

Δίνεται: *γράφος* $G(V, E)$ *με μη αρνητικά βάρη στις ακμές του οι κόμβοι του οποίου χωρίζονται σε δύο σύνολα:* απαραίτητοι *και* Steiner.
Ζητείται: *δέντρο ελαχίστου κόστους που να περιέχει όλους τους απαραίτητους κόμβους.*

Επιπλέον υπόθεση: ο γράφος είναι πλήρης και τα δοθέντα βάρη ικανοποιούν την τριγωνική ανισότητα.

Ισοδυναμία προσεγγισιμότητας Steiner Tree και Metric Steiner Tree

Θεώρημα

Δοθέντος ρ-προσεγγιστικού αλγορίθμου για το Metric Steiner Tree *μπορούμε να κατασκευάσουμε ρ-προσεγγιστικό αλγόριθμο για το* Steiner Tree.

Απόδειξη: με αναγωγή διατήρησης του παράγοντα προσέγγισης από το Steiner Tree στο Metric Steiner Tree.

Steiner Tree $\leq_{\text{apx.pres.}}$ Metric Steiner Tree

Συμπλήρωση του αρχικού γράφου G σε πλήρη γράφο G'. Οι ακμές του G' έχουν το βάρος των συντομότερων μονοπατιών στον G (*metric closure*). Οι απαραίτητοι κόμβοι είναι ίδιοι.
$OPT(I') \leq OPT(I)$ (γιατί;)
Κάθε λύση του I' με κόστος $SOL(I')$ δίνει λύση του I με κόστος $SOL(I) \leq SOL(I')$: αντικατάσταση κάθε ακμής της λύσης με το αντίστοιχο μονοπάτι. Επομένως:

$$SOL(I) \leq SOL(I') \leq \rho OPT(I') \leq \rho OPT(I)$$

Σημείωση: *Ισχύει επιπλέον ότι* $OPT(I) = OPT(I')$, *αλλά δεν το χρειαζόμαστε.*

Αναγωγές Διατήρησης Παράγοντα Προσέγγισης (Approximation Factor Preserving Reductions)

Μια *αναγωγή διατήρησης παράγοντα προσέγγισης* από το πρόβλημα ελαχιστοποίησης Π στο πρόβλημα ελαχιστοποίησης Π' είναι ένα ζεύγος συναρτήσεων πολυωνυμικού χρόνου h, g, όπου η h απεικονίζει κάθε στιγμιότυπο I του Π σε κάποιο στιγμιότυπο $I' = h(I)$ του Π' και η g απεικονίζει λύσεις του I' σε λύσεις του I, ώστε:

- $OPT(I') \leq OPT(I)$
- για κάθε λύση S' του I' με κόστος $SOL(I', S')$ η $S = g(S')$ είναι λύση του I με κόστος $SOL(I, S) \leq SOL(I', S')$.

Θεώρημα

Μια αναγωγή διατήρησης παράγοντα προσέγγισης από το πρόβλημα Π στο πρόβλημα Π' μαζί με έναν ρ-προσεγγιστικό αλγόριθμο για το Π' δίνουν έναν ρ-προσεγγιστικό αλγόριθμο για το Π.

2-προσεγγιστικός αλγόριθμος για το Metric Steiner Tree (και το Steiner Tree)

Αλγόριθμος: βρες και επίστρεψε ελάχιστο συνδετικό δένδρο στον παραγόμενο (node-induced) υπογράφο των απαιτούμενων κόμβων (R).

Η απόδειξη βασίζεται στο ότι ένα τέτοιο δένδρο είναι εφικτή λύση για το Metric Steiner Tree και χρησιμοποιεί παρόμοια τεχνική με τον αλγόριθμο για το TSP: από μια οποιαδήποτε λύση για το Metric Steiner Tree μπορούμε να κατασκευάσουμε συνδετικό δέντρο στους κόμβους του R μόνο, διπλάσιου το πολύ κόστους (short-cutting στον αντίστοιχο κύκλο Euler). Επομένως και στη βέλτιστη λύση αντιστοιχεί ένα συνδετικό δέντρο T_R^*, με κόστος το πολύ $2 \cdot OPT$. Αρα:

$$cost(\mathrm{MST}_R) \leq cost(T_R^*) \leq 2 \cdot OPT$$

Multiway Cut / Minimum k-Cut

Είναι γενικεύσεις του προβλήματος min-cut (θυμηθείτε το θεώρημα max-flow / min-cut):

Στο πρόβλημα Multiway Cut δίνεται γράφος G και k κόμβοι (terminals). Ζητείται η ελάχιστη τομή τέτοια ώστε κάθε terminal να βρίσκεται σε διαφορετική συνεκτική συνιστώσα (connected component). **NP**-hard, ακόμη και για fixed $k \geq 3$. Για $k = 2$;

Στο πρόβλημα Minimum k-Cut δίνεται γράφος G και ζητείται η ελάχιστη τομή τέτοια ώστε ο γράφος να διασπάται σε k συνεκτικές συνιστώσες. Πολυωνυμικά επιλύσιμο για fixed k, **NP**-hard για k που είναι μέρος της εισόδου.

Αλγόριθμος για το Multiway Cut

`Βρες isolating cut` C_i `για κάθε terminal` s_i.
`Δώσε σαν λύση την ένωση όλων αυτών των cuts, εκτός από την βαρύτερη.`

Ορθότητα και λόγος προσέγγισης.

Έστω A η βέλτιστη λύση, χωρίζοντας τον γράφο σε συνιστώσες $V_1, \ldots, V_k$. Αν A_i είναι η τομή-υποσύνολο της A που χωρίζει το V_i από τις υπόλοιπες συνιστώσες, τότε είναι isolating cut για το s_i.
'ρα, $w(C_i) \leq w(A_i)$.
Κάθε edge της A, περιλαμβάνεται σε δύο τομές A_i, A_j, οπότε

$$\sum_{i=1}^{k} w(A_i) = 2w(A) = 2OPT$$

Επομένως:

$$w(C) \leq (1 - 1/k) \sum_{i=1}^{k} w(C_i) \leq (1 - 1/k) \sum_{i=1}^{k} w(A_i) = 2(1 - 1/k)OPT$$

Αλγόριθμος για το Minimum k-Cut

```
1. Compute Gomory-Hu tree T.
2. Output the union of cuts corresponding to
lightest k−1 edges of T.
```

Κατασκευή δένδρου Gomory-Hu (σημαντικές ιδιότητες: κάθε ακμή του αντιστοιχεί σε μια ελάχιστη τομή στον γράφο, και για κάθε ζεύγος (u,v) υπάρχει μια ακμή του δένδρου που αντιστοιχεί σε μια ελάχιστη (u,v)-cut)

1. construct tree T with unique node the set $S_0 = V$
2. while there is S_i s.t. $|S_i| \geq 2$ do
    - choose two vertices in S_i, say x, y
    - compute minimum $x - y$ cut in G'
        - /* $G' = G$ with subtrees of S_i in T collapsed */
    - split S_i accordingly to S_i^x, S_i^y, with an edge between them
        - with weight equal to that of the cut
    - stick each subtree of S_i in T to S_i^x or S_i^y according to the cut

Με απόδειξη παρόμοια (κάπως πιο δύσκολη) αυτής για το multiway cut αποδεικνύεται ότι η λύση είναι $\leq 2(1 - 1/k)OPT$.