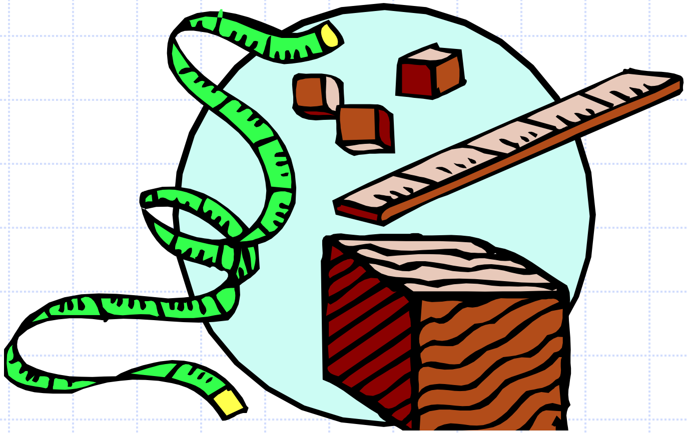
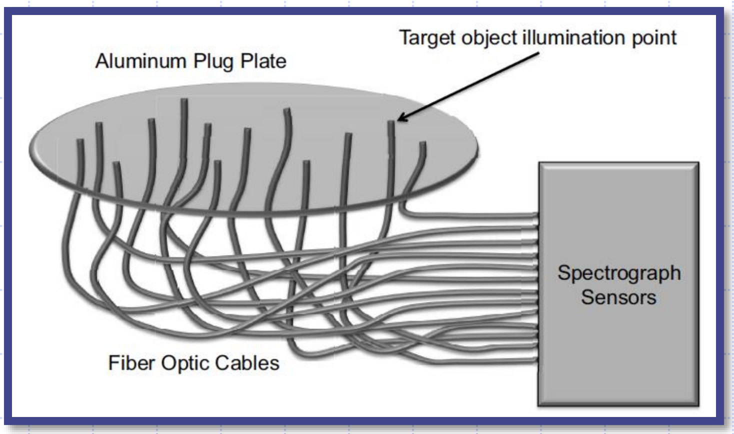
Παρουσίαση για χρήση με το σύγγραμμα, Αλγόριθμοι Σχεδίαση και Εφαρμογές, των M. T. Goodrich and R. Tamassia, Wiley, 2015 (στα ελληνικά από εκδόσεις Μ. Γκιούρδας)

Προσεγγιστικοί αλγόριθμοι



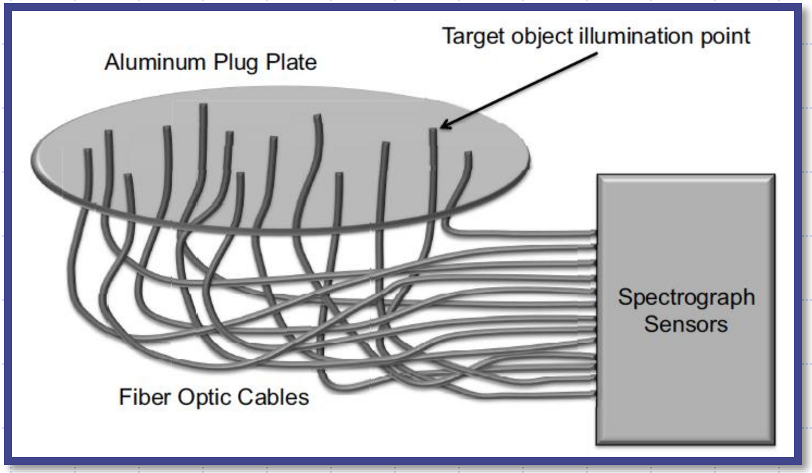
Εφαρμογές

- Ένα από τα πιο χρονοβόρα μέρη της αστρονομίας είναι η συλλογή φωτός από έναν γαλαξία ή ένα άστρο σε μία δεδομένη χρονική περίοδο.
- Για να το κάνουμε αυτό με τηλεσκόπιο, χρησιμοποιούμε ένα μεγάλο δίσκο αλουμινίου στο μέγεθος της διαμέτρου του τηλεσκοπίου.
- Ο δίσκος τοποθετείται στο εστιακό επίπεδο του τηλεσκοπίου, έτσι ώστε το φως από κάθε αστρικό αντικείμενο σε μία παρατήρηση να καταλήγει σε ένα καθορισμένο σημείο στο δίσκο.
- Οι αστρονόμοι γνωρίζουν πού βρίσκονται αυτά τα σημεία και πριν ο δίσκος τοποθετηθεί στο εστιακό επίπεδο χρησιμοποιούν ρομποτικό εξοπλισμό διάνοιξης για να ανοίξουν οπές σε κάθε σημείο που επιθυμούν, εισάγων έναν αγωγό οπτικών ινών σε κάθε τέτοια οπή και τον συνδέουν με ένα φασματογράφο.



Εφαρμογή στο TSP

- Το πρόβλημα της εύρεσης ενός ταχύτατου τρόπου διάνοιξης οπών είναι ένα παράδειγμα του **προβλήματος περιοδεύοντος πωλητή(TSP)**.

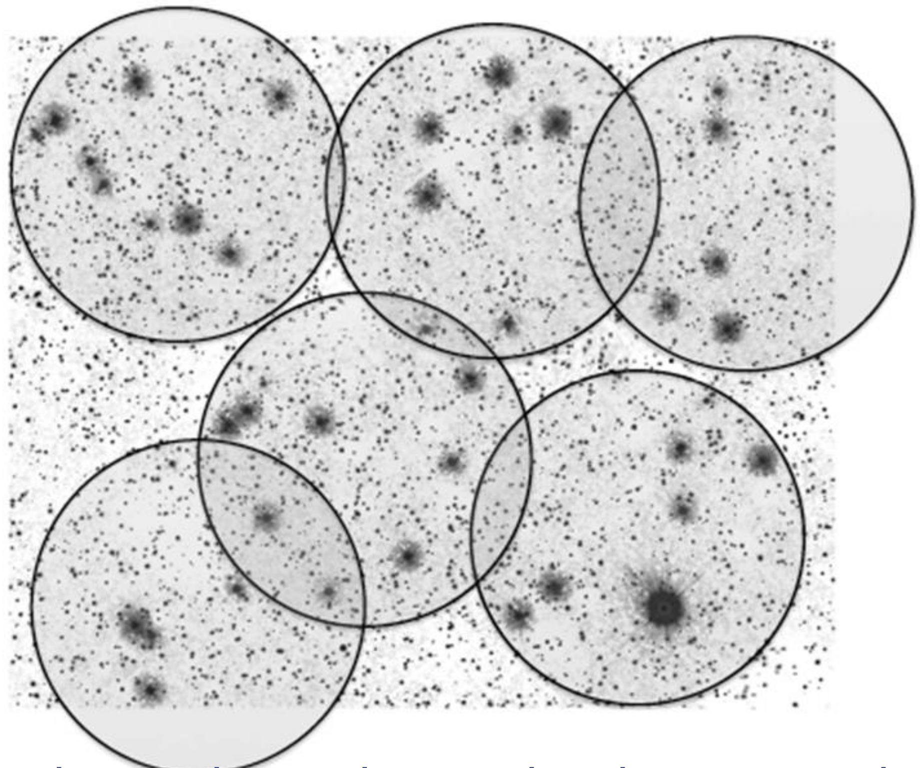


- Σύμφωνα με τη θεωρητική διατύπωση του TSP, κάθε θέση οπής είναι μία “πόλη” και ο χρόνος που απαιτείται, ώστε ένα ρομποτικό τρυπάνι να μετακινηθεί από μία οπή σε άλλη αντιστοιχεί στην απόσταση μεταξύ των “πόλεων” που αντιστοιχούν σ’ αυτές τις δύο οπές.
- Συνεπώς, η διάσχιση ελάχιστης απόστασης των πόλεων, η οποία ξεκινά και τελειώνει στη θέση αναμονής του τρυπανιού είναι εκείνη που ανοίγει τις οπές ταχύτερα.
- Δυστυχώς, το TSP είναι NP-πλήρες.
- Έτσι θα ήταν ιδανικό να μπορέσουμε τουλάχιστον να προσεγγίσουμε το πρόβλημα.

Εφαρμογή στη κάλυψη συνόλου

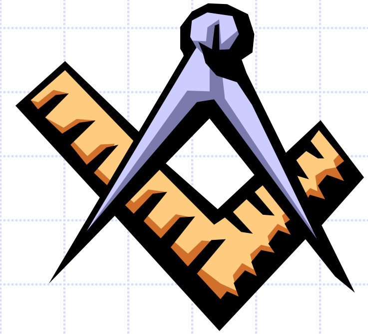
- Ένα ακόμη πρόβλημα βελτιστοποίησης είναι η ελαχιστοποίηση των παρατηρήσεων που απαιτούνται προκειμένου να συλλεχθούν τα φάσματα όλων των αστρικών αντικειμένων που μας ενδιαφέρουν.
- Σ' αυτήν την περίπτωση, οι αστρονόμοι έχουν ένα χάρτη όλων των αστρικών αντικειμένων, που τους ενδιαφέρουν και θέλουν να τον καλύψουν με τον ελάχιστο αριθμό δίσκων ίδιας διαμέτρου με το τηλεσκόπιο.
- Αυτό το πρόβλημα βελτιστοποίησης είναι ένα στιγμιότυπο του **προβλήματος κάλυψης συνόλου**.
- Κάθε ξεχωριστό σύνολο αντικειμένων που μπορεί να συμπεριληφθεί σε μία μόνο παρατήρηση δίνεται ως σύνολο εισόδου και το πρόβλημα βελτιστοποίησης αφορά στην ελαχιστοποίηση του αριθμού των συνόλων των οποίων η ένωση περιλαμβάνει όλα τα αντικείμενα ενδιαφέροντος.
- Αυτό το πρόβλημα είναι επίσης NP-πλήρες, αλλά είναι εάν πρόβλημα για το οποίο αρκεί μια προσέγγιση στο βέλτιστο.

Παράδειγμα κάλυψης συνόλου



Σχήμα 18.2: Ένα παράδειγμα κάλυψης δίσκου για ένα σύνολο σημαντικών αστρικών αντικειμένων (τα μικρότερα αντικείμενα δεν περιλαμβάνονται). Η εικόνα του φόντου είναι μια εικόνα του Ωμέγα Κένταυρου, 2009. Εικόνα της αμερικανικής κυβέρνησης. Πηγή: NASA, ESA και Ομάδα Hubble SM4 ERO team.

Λόγοι προσέγγισης

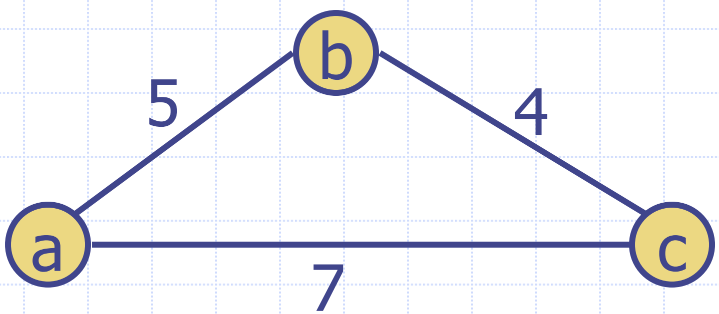


- Προβλήματα βελτιστοποίησης
  - Έχουμε μία περίσταση ενός προβλήματος x που έχει πολλές εφικτές “λύσεις”.
  - Προσπαθούμε να ελαχιστοποιήσουμε (ή να μεγιστοποιήσουμε) κάποια συνάρτηση κόστους c(S) για κάποια “λύση” S του x. Για παράδειγμα,
    - Εύρεση δέντρου επικάλυψης ελάχιστου κόστους
    - Εύρεση του μικρότερου vertex cover ενός γράφου
    - Εύρεση της μικρότερης διαδρομής περιοδεύοντος πωλητή
- Μία προσέγγιση παράγει μία λύση T
  - Η T είναι μία **k-προσέγγιση** στη βέλτιστη λύση OPT εάν $c(T)/c(OPT) \leq k$ (υποθέτοντας ότι είναι πρόβλημα ελαχιστοποίησης.; Η προσέγγιση μεγιστοποίησης θα ήταν το αντίθετο)

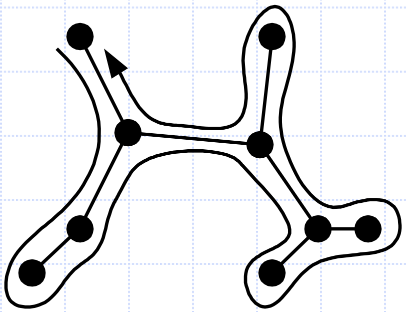
© 2015 Goodrich and Tamassia

Ειδική περίπτωση του προβλήματος περιοδεύοντος πωλητή

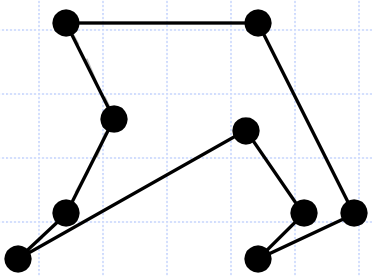
- OPT-TSP: Εύρεση ενός κύκλου ελάχιστου βάρους που επισκέπτεται όλες τις κορυφές σε έναν πλήρη σταθμισμένο γράφο.
  - Το OPT-TSP είναι NP-πλήρες
  - Ειδική περίπτωση: τα βάρη των ακμών ικανοποιούν την τριγωνική ανισότητα (που είναι σύνηθες σε αρκετές εφαρμογές):
    - $w(a,b) + w(b,c) \geq w(a,c)$



Μία 2-Προσέγγιση για το TSP Ειδική περίπτωση



Euler tour P of MST M



Output tour T

Algorithm *TSPApprox*(G)

Input weighted complete graph G, satisfying the triangle inequality

Output a TSP tour T for G

$M \leftarrow$ a minimum spanning tree for G

$P \leftarrow$ an Euler tour traversal of M, starting at some vertex s

$T \leftarrow$ empty list

for each vertex v in P (in traversal order)

if this is v's first appearance in P **then**

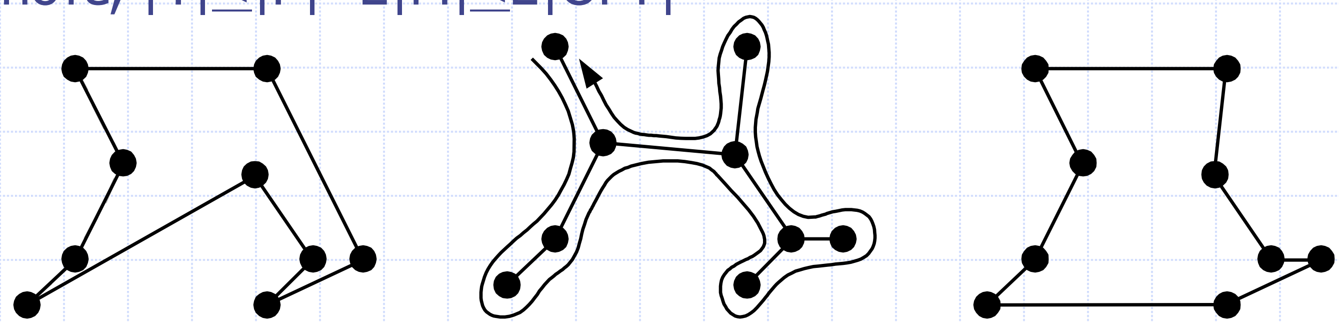
T.insertLast(v)

T.insertLast(s)

return T

Μία 2-Προσέγγιση για το TSP Ειδική περίπτωση - Απόδειξη

- Η ιδανική διαδρομή είναι δέντρο επικάλυψης; Οπότε $|M| \le |OPT|$.
- Η διαδρομή Euler P επισκέπτεται κάθε ακμή του M δύο φορές; Οπότε $|P|=2|M|$
- Κάθε φορά που ενώνουμε μία κορυφή στην διαδρομή Euler δεν θα αυξήσουμε το μήκος της, λόγω της τριγωνικής ανισότητάς ($w(a,b) + w(b,c) \ge w(a,c)$); Οπότε, $|T| \le |P|$.
- Οπότε, $|T| \le |P| = 2|M| \le 2|OPT|$



Output tour *T* (at most the cost of *P*)

Euler tour *P* of MST *M* (twice the cost of *M*)

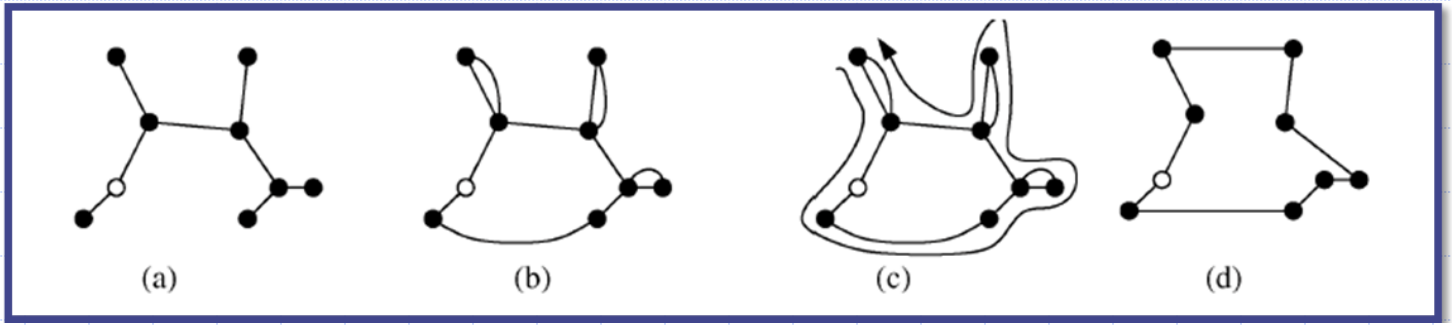
Optimal tour *OPT* (at least the cost of MST *M*)

Ο αλγόριθμος του Χριστοφίδη

1. Κατασκευάζουμε ένα δέντρο επικάλυψης ελάχιστου κόστους, M, για G.
2. Έστω ότι το W είναι το σύνολο των κορυφών του G, που έχουν περιττό βαθμό στο M και H είναι ο υπογράφος του G, που δημιουργείται από τις κορυφές στο W. Με άλλα λόγια, το H είναι ο γράφος, που έχει το W ως κορυφές του και όλες τις ακμές από το C που ενώνουν τέτοιες κορυφές. Με ένα απλό επιχείρημα μπορούμε να αποδείξουμε ότι το πλήθος των κορυφών στο W είναι άρτιο (βλ. Άσκηση Υ-18.12). Υπολογίζουμε μια τέλεια αντιστοίχιση ελάχιστου κόστους, P, στο H.
3. Συνδυάζουμε τους γράφους M και P, ώστε να δημιουργήσουμε έναν γράφο, G', αλλά δεν συνδυάζουμε παράλληλες ακμές σε ενιαίες ακμές. Με άλλα λόγια, αν μια ακμή e βρίσκεται τόσο στο M όσο και στο P, τότε δημιουργούμε δυο αντίγραφα του e στο συνδυασμένο γράφο, G'.
4. Δημιουργούμε ένα κύκλωμα Euler, C, στο G', το οποίο επισκέπτεται κάθε ακμή ακριβώς μία φορά (αντίθετα από τον 2-προσεγγισπκό αλγόριθμο, εδώ οι ακμές του G' είναι μη κατευθυνόμενες).
5. Μετατρέπουμε το C σε μια διάσχιση, T, παρακάμπτοντας τις κορυφές που επισκεφτήκαμε προηγουμένως.

Ο χρόνος εκτέλεσης καθορίζεται από το βήμα 2, το οποίο μπορεί να εκτελεστεί σε χρόνο $O(n^3)$.

Παράδειγμα και αρχή της ανάλυσης



Σχήμα 18.5: Παράδειγμα εκτέλεσης του προσεγγιστικού αλγόριθμου του Χριστοφίδη για το METRIIC-TSP: (a) ένα δέντρο ελάχιστης επικάλυψης, M, για το C, (b) μια τέλεια αντιστοίχιση ελαχίστου κόστους P για τις κορυφές στο W (οι κορυφές στο W παρουσιάζονται συνεχείς, ενώ οι ακμές στο P παρουσιάζονται ως καμπύλα τόξα), (c) ένα κύκλωμα Euler, C, του G\` και (d) προσεγγιστική διάσχιση του TSP, T.

- Για να ξεκινήσουμε την ανάλυση μας για τον προσεγγιστικό αλγόριθμου του Χριστοφίδη έστω ότι το S είναι μια βέλτιστη λύση σ’ αυτό το στιγμιότυπο του METRIC-TSP και T είναι η διάσχιση που παράγεται από τον αλγόριθμο του Χριστοφίδη.
- Επειδή το S περιλαμβάνει ένα δέντρο επικάλυψης και το M είναι ένα δέντρο επικάλυψης ελάχιστου κόστους στο G, c(M) ≤ c(S).

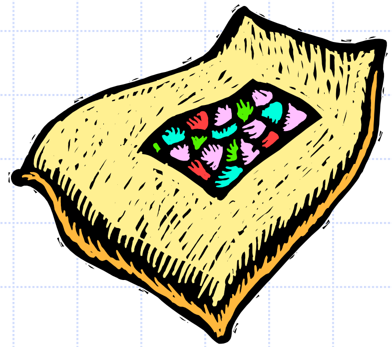
Ανάλυση, συνέχεια

- Επιπλέον, έστω ότι το R συμβολίζει μια λύση στο πρόβλημα περιοδεύοντος πωλητή στο H.
- Επειδή οι ακμές στο G (και, επομένως και στο H) ικανοποιούν την τριγωνική ανισότητα, όλες οι ακμές του H είναι επίσης στο G, $c(R) \leq c(S)$.
- Με άλλα λόγια, η επίσκεψη περισσότερων κορυφών απ' ό,τι στη διάσχιση R δεν μπορεί να μειώσει το συνολικό κόστος της.
- Θα εξετάσουμε τώρα το κόστος μιας τέλειας αντιστοίχισης, P, του H, και τη σχέση της με το R, μιας βέλτιστης διάσχισης περιοδεύοντος πωλητή του H. Απαριθμούμε τις ακμές του R και αγνοούμε την τελευταία ακμή (η οποία επιστρέφει την αρχική κορυφή).
- Σημειώστε ότι το κόστος του συνόλου των περιττών ακμών και το σύνολο των άρτιων ακμών στο Rδίνουν αθροιστικά $c(R)$ και επομένως, ένα απ' αυτά τα δύο σύνολα έχει συνολικό κόστος το πολύ το μισό του R, δηλαδή κόστος το πολύ $c(R)/2$.

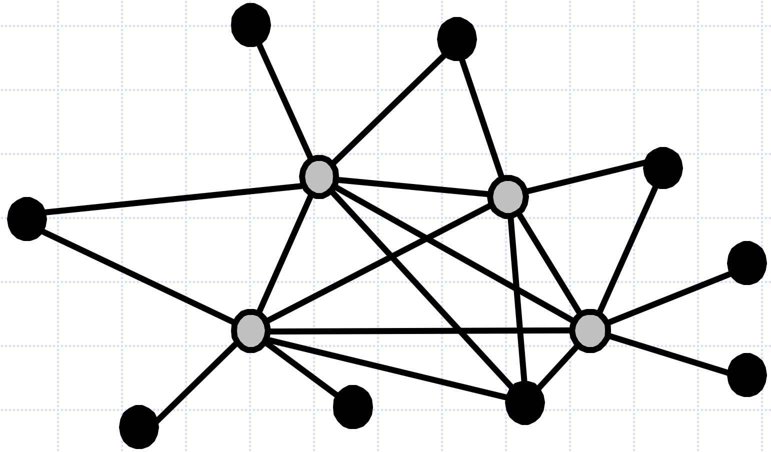
Ανάλυση, ολοκλήρωση

- Επιπλέον, το σύνολο των περιττών ακμών και το σύνολο των άρτιων ακμών στο R είναι αμφότερα τέλειες αντιστοιχίσεις και επομένως το κόστος του P,μιας τέλειας αντιστοίχισης ελάχιστου βάρους στις ακμές του H, θα είναι το πολύ το μικρότερο από τα δύο.
  Δηλαδή, $c(P) \leq c(R)/2$.
- Επομένως, $c(M) + c(P) \leq c(S) + c(R)/2 \leq 3c(S)/2$.
- Επειδή οι ακμές στο G ικανοποιούν την τριγωνική ανισότητα, μπορούμε να βελτιώσουμε το κόστος μιας διάσχισης αν ακολουθήσουμε συντομεύσεις, που αποφεύγουν τις κορυφές που είχαμε επισκεφθεί προηγουμένως. Συνεπώς, $c(T) \leq c(M) + c(P)$, το οποίο σημαίνει ότι $c(T) \leq 3c(S)/2$.
- Με άλλα λόγια ο προσεγγιστικός αλγόριθμος του Χριστοφίδη είναι ένας **(3/2)-προσεγγιστικός αλγόριθμος** για το πρόβλημα βελτιστοποίησης METRIC-TSP, που εκτελείται σε πολυωνυμικό χρόνο.

Κάλυψη κορυφών



- Η **κάλυψη κορυφών** ενός γράφου G=(V,E) είναι ένα υποσύνολο W του V, έτσι ώστε, για κάθε (a,b) στο E, το a είναι στο W ή το b είναι στο W.
- OPT-VERTEX-COVER: Για έναν γράφο G, εύρεση κάλυψης κορυφών του G με το ελάχιστο μέγεθος.
- Το OPT-VERTEX-COVER είναι NP-hard.



Μία 2-Προσέγγιση για το Vertex Cover

- Κάθε επιλεγμένη ακμή έχει και τα δύο άκρα της e στο C
- Αλλά η e πρέπει να καλυφθεί με βέλτιστη κάλυψη; Οπότε μόνο μια άκρη της e θα είναι στην OPT
- Οπότε, υπάρχουν το πολύ διπλάσιες κορυφές στην C σε σχέση με την OPT.
- Συνεπώς, το C είναι 2-προσέγγιση. της OPT
- Χρόνος εκτέλεσης: O(n+m)

Algorithm VertexCoverApprox(G):
    Input: A graph G
    Output: A small vertex cover C for G

```
C ← ∅
while G still has edges do
    select an edge e = (v, w) of G
    add vertices v and w to C
    for each edge f incident to v or w do
        remove f from G
return C
```

Set Cover (Άπληστος αλγόριθμος)

- OPT-SET-COVER: Βάση μίας συλλογής m συνόλων, εύρεση του μικρότερου αριθμού αυτών η ένωση των οποίων είναι η ίδια η συλλογή των m συνόλων?
  - Το OPT-SET-COVER είναι NP-hard
- Η άπληστη προσέγγιση δημιουργεί έναν O(log n)-προσεγγιστικό αλγόριθμο.

Algorithm SetCoverApprox(S):

Input: A collection S of sets $S_1, S_2, \ldots, S_m$ whose union is U

Output: A small set cover C for S

$C \leftarrow \emptyset$ // The set cover built so far

$E \leftarrow \emptyset$ // The elements from U currently covered by C

while $E \neq U$ **do**

  select a set S_i that has the maximum number of uncovered elements

  add S_i to C

  $E \leftarrow E \cup S_i$

Return C.

Ανάλυση άπληστου Set Cover

- Ας σκεφτούμε τη στιγμή στον αλγόριθμο μας, όπου ένα σύνολο S_j προστίθεται στο C και έστω ότι k είναι το πλήθος των προηγουμένως ακάλυπτων στοιχείων στο S_j.
- Πρέπει να καταβάλλουμε συνολική χρέωση 1 για να προσθέσουμε αυτό το σύνολο στο C, γι’ αυτό χρεώνουμε κάθε προηγουμένως ακάλυπτο στοιχείο i του S_j με $c(i) = 1/k$.
- Συνεπώς, το συνολικό μέγεθος της κάλυψης μας ισούται με τις συνολικές αλλαγές που γίνονται.
- Για να αποδείξουμε ένα όριο προσέγγισης, θα μελετήσουμε τις χρεώσεις που γίνονται στα στοιχεία κάθε υποσυνόλου S_j , που ανήκει σε μια βέλτιστη κάλυψη, C'. Γι' αυτό υποθέτουμε ότι το S_j ανήκει στο C ′.
- Γράφουμε $S_j = \{x_1, x_2, . . . , x_{nj}\}$ έτσι ώστε τα στοιχεία του S_j να εμφανίζονται με τη σειρά με την οποία καλύπτονται από τον αλγόριθμο μας .

Ανάλυση άπληστου Set Cover, συνέχεια

- Στη συνέχεια εξετάζουμε την επανάληψη, στην οποία καλύπτεται το χ_1. Τη στιγμή εκείνη, το S_j δεν έχει ακόμα επιλεχθεί και επομένως, όποιο σύνολο κι αν επιλεχθεί, πρέπει να έχει τουλάχιστον n_j ακάλυπτα στοιχεία. Συνεπώς, το χ_1, χρεώνεται το πολύ $1/\eta_j$. Γι' αυτό θα εξετάσουμε τότε τη στιγμή που ο αλγόριθμός μας χρεώνει ένα στοιχείο x_l του S_j. Στην χείριστη περίπτωση, δεν θα έχουμε επιλέξει το S_j (ο αλγόριθμός μας μάλιστα είναι πιθανό να μην επιλέξει ποτέ αυτό το S_j). Όποιο σύνολο κι αν επιλεχθεί σ' αυτήν την επανάληψη, στην χείριστη περίπτωση, θα υπάρχουν τουλάχιστον $n_j – l + 1$ ακάλυπτα στοιχεία και επομένως, το x_l θα χρεώνεται το πολύ $1 /(n_j – l + 1)$. Επομένως, το συνολικό ποσό που χρεώνεται σε όλα τα στοιχεία του S_j είναι το πολύ

$$\sum_{l=1}^{n_j} \frac{1}{n_l - l + 1} = \sum_{l=1}^{n_j} \frac{1}{l},$$

Ανάλυση άπληστου Set Cover, συνέχεια

- το οποίο είναι ο γνωστός **αρμονικός αριθμός**, H_{ni}. Είναι γνωστό (για παράδειγμα, ανατρέξτε στο Παράρτημα) ότι το H_{nj} είναι $O(\log n_j)$. Έστω ότι το $c(S_j)$ συμβολίζει τις συνολικές χρεώσεις που δίνονται σε όλα τα στοιχεία ενός συνόλου S_j που ανήκει στη βέλτιστη κάλυψη C'. Το σύστημα κάλυψης που εφαρμόζουμε συνεπάγεται ότι το $c(S_j)$ είναι $O(\log n_j)$. Συνεπώς, αν αθροίσουμε τα σύνολα του C, παίρνουμε

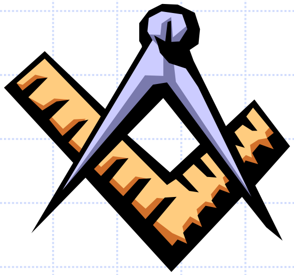
$$\begin{aligned} \sum_{S_j \in C'} c(S_j) &\le \sum_{S_j \in C'} b \log n_j \\ &\le b|C'| \log n, \end{aligned}$$

- για κάποια σταθερά b > 1. Αλλά επειδή το C' είναι μια κάλυψη συνόλων,

$$\sum_{i \in U} c(i) \le \sum_{S_j \in C'} c(S_j).$$

- Επομένως: $|C| \le b|C'| \log n$.

Τεχνικές προσέγγισης πολυωνυμικού χρόνου



- Ένα πρόβλημα L έχει μία **προσεγγιστική μέθοδο πολυωνυμικού χρόνου (PTAS)** ένα έχει έναν πολυωνυμικού χρόνου (1+ε)-προσεγγιστικό αλγόριθμο, για οποιαδήποτε σταθερή τιμή ε >0 (η τιμή αυτή μπορεί να εμφανίζεται και στον χρόνο εκτέλεση).
- Το 0/1 Knapsack έχει PTAS, με χρόνο εκτέλεσης $O(n^3/\varepsilon)$.