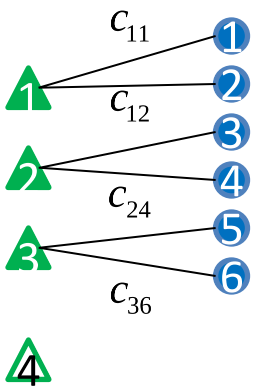
Ισίδωρος Ροδομαγουλάκης
Αλγόριθμοι Δικτύων και Πολυπλοκότητα
2010 - 2011

K-median

Επισκόπηση του κεφαλαίου 25 από το βιβλίο «Approximation algorithms» του V. Vazirani

Metric Uncapacitated Facility Location



- Έστω ο διμερής γράφος G=(F,C)
  - $F = \{f_1, f_2, f_3\}$: το σύνολο των facilities(κέντρων)
  - $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$: το σύνολο των πόλεων

- **Κάθε** πόλη πρέπει να εξυπηρετηθεί/συνδεθεί σε **ένα** από τα |F| κέντρα έτσι ώστε να **ελαχιστοποιείται** το συνολικό κόστος συνδέσεων σε συνδυασμό με το κόστος ανοίγματος των κέντρων $I \subseteq F$ που χρησιμοποιούνται.

$$\sum_{i\in I} c_{ij} + \sum_{i\in I} y_i \qquad \forall j \in C$$

- f_i το κόστος ανοίγματος του κέντρου **i**
- c_{ij} το κόστος σύνδεσης της πόλης **j** με το κέντρο **i**

Metric case : Ισχύει η τριγωνική ανισότητα στα κόστη σύνδεσης

Γραμμικός προγραμματισμός

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{i\in F,\, j\in C} c_{ij}x_{ij} + \sum_{i\in F} f_i y_i & \\
\text{subject to} & \displaystyle\sum_{i\in F} x_{ij} \geq 1, & j \in C \\
& y_i - x_{ij} \geq 0, & i \in F,\ j \in C \\
& x_{ij} \geq 0, & i \in F,\ j \in C \\
& y_i \geq 0, & i \in F
\end{array}
$$

Μεταβλητές Πρωτεύοντος

x_{ij} : αληθές όταν η πόλη **j** είναι συνδεδεμένη με το κέντρο **i**

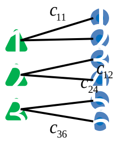
y_i : αληθές όταν το κέντρο i είναι ανοικτό

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{j\in C} \alpha_j & \\
\text{subject to} & \alpha_j - \beta_{ij} \leq c_{ij}, & i \in F,\ j \in C \\
& \displaystyle\sum_{j\in C} \beta_{ij} \leq f_i, & i \in F \\
& \alpha_j \geq 0, & j \in C \\
& \beta_{ij} \geq 0, & i \in F,\ j \in C
\end{array}
$$

Μεταβλητές Δυϊκού

a_i : το συνολικό κόστος σύνδεσης της πόλης i

β_{ij} : η συνεισφορά της πόλης j στο άνοιγμα του κέντρου i



Ερμηνεία Περιορισμών

- $\sum_{i \in F} x_{ij} \geq 1 \quad \forall j \in C$ κάθε πόλη πρέπει να συνδεθεί τουλάχιστον σ’ένα κέντρο
- $y_i - x_{ij} \geq 1 \quad \forall i \in F \quad \forall j \in C$ οι πόλεις συνδέονται μόνο σε ανοικτά κέντρα

- Slackness Conditions

$$\forall i \in F, j \in C : x_{ij} > 0 \Rightarrow \alpha_j - \beta_{ij} = c_{ij}$$

Για να συνδεθεί μία πόλη σ’ένα κέντρο πρέπει συνεισφέρει για το άνοιγμα του και να πληρώσει τη σύνδεσή της

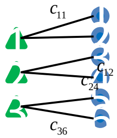
$$\forall i \in F : y_i > 0 \Rightarrow \sum_{j \in C} \beta_{ij} = f_i$$

Για να ανοίξει το κέντρο **i** πρέπει να συνεισφέρουν για την εκπλήρωση του κόστους ανοίγματος όλες οι ενδιαφερόμενες πόλεις

$$\forall j \in C : \alpha_j > 0 \Rightarrow \sum_{i \in F} x_{ij} = 1$$

$$\forall i \in F, j \in C : \beta_{ij} > 0 \Rightarrow y_i = x_{ij}$$

Οι πόλεις συνεισφέρουν μόνο για το κέντρο που τελικά συνδέονται



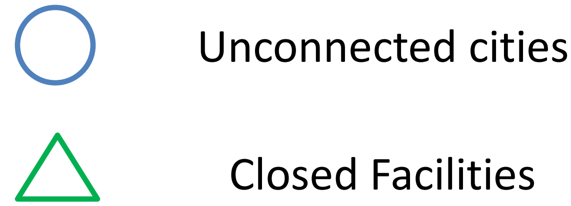
Εκτιμητικός Αλγόριθμος Primal-dual σχήματος

- Βρίσκουμε επαναληπτικά βέλτιστες ακέραιες λύσεις στο πρωτεύον γραμμικό πρόγραμμα φράζοντάς τες με εφικτές λύσεις του δυϊκού του.
- Συνήθως οι προσεγγιστικοί αλγόριθμοι που προκύπτουν είναι πιο γρήγοροι από αυτούς που προκύπτουν από τη στρογγυλοποίηση της κλασματικής λύσης του πρωτεύοντος προβλήματος
- Ειδικά για το facility location, ο χρόνος εκτέλεσης είναι της τάξης του O(**m***log***m**) όπου **m** το πλήθος των ακμών και ο λόγος προσέγγισης που πετυχαίνει είναι <=**3**

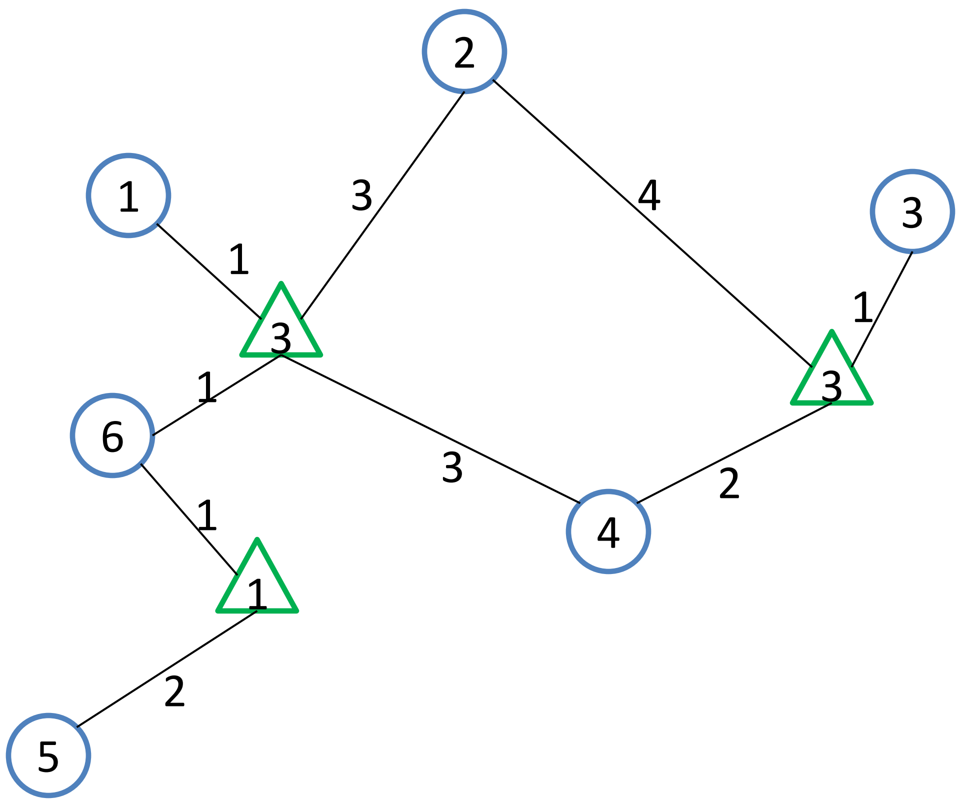
Περιγραφή του Αλγορίθμου

- Κατά τη Φάση 1 βρίσκει μία εφικτή λύση για το δυϊκό πρόβλημα μεγιστοποίησης του $\sum_{j \in C} a_j$.
- Κατά τη Φάση 2 βρίσκει μία ακέραια λύση $(x_{ji} \in \{0,1\}, y_i \in \{0,1\})$ του πρωτεύοντος προβλήματος βάση του κάτω ορίου που έδωσε η Φάση 1
- Οι πόλεις αρχίζουν και πληρώνουν ίσα ποσά ανά χρονικό βήμα ώσπου να συνδεθούν σ’ένα κέντρο και να συνεχίζουν να πληρώνουν ώσπου αυτό να ανοίξει προσωρινά
- Στη Φάση 2 ελέγχεται αν κάποιες πόλεις συνδέονται σε περισσότερα από ένα προσωρινά ανοικτά κέντρα και τότε επιλέγεται μόνο ένα υποσύνολο από αυτά

Phase 1 t = 0



$$G = (F, C)$$
$$F = \{f_1, f_2, f_3\}$$
$$C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$$
$$\mathbf{f}_1 = 3 \quad \mathbf{f}_2 = 3 \quad \mathbf{f}_3 = 4$$



Ο χρόνος t μοντελοποιεί τις σταδιακές αυξήσεις στη διϋκή μεταβλητή a_j των πόρων που προσφέρει κάθε πόλη **j** για να ενωθεί

Phase 1 t = 1

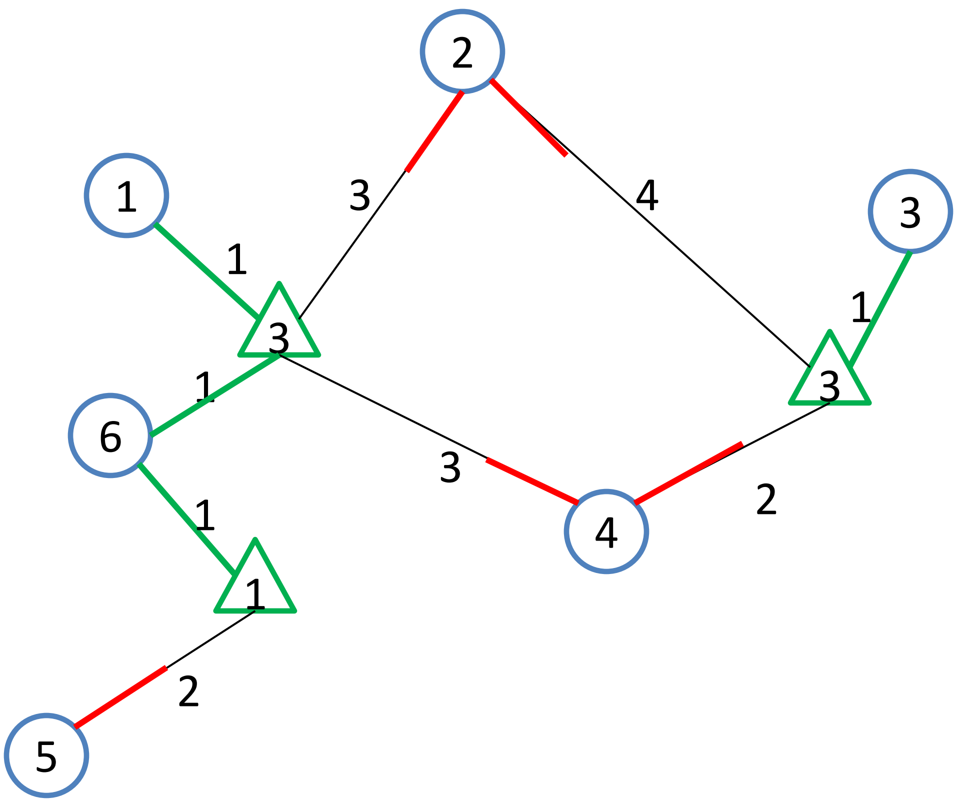
Unconnected cities

Closed Facilities

payment

tight

$$a_j = a_j + 1$$
$$\beta_{ji} = 0$$



Οι πόλεις 3,5 και 6 εκπλήρωσαν το κόστος σύνδεσής τους και θεωρούνται «δεμένες»

Phase 1 $\underline{t = 2}$

Unconnected cities

Connected cities

Closed Facilities

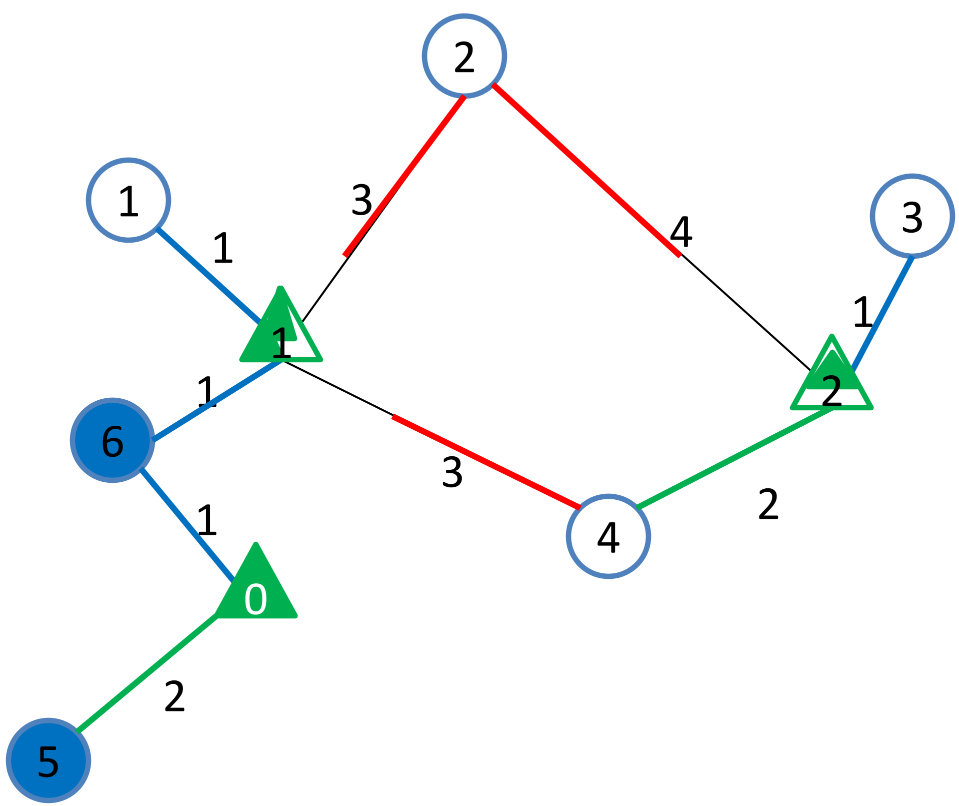
payment

tight

special

$$a_j = a_j + 1$$

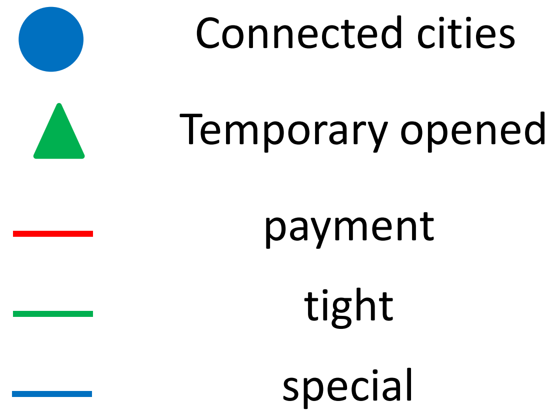
$$\beta_{61} = 1 \quad \beta_{63} = 1 \quad \beta_{32} = 1$$

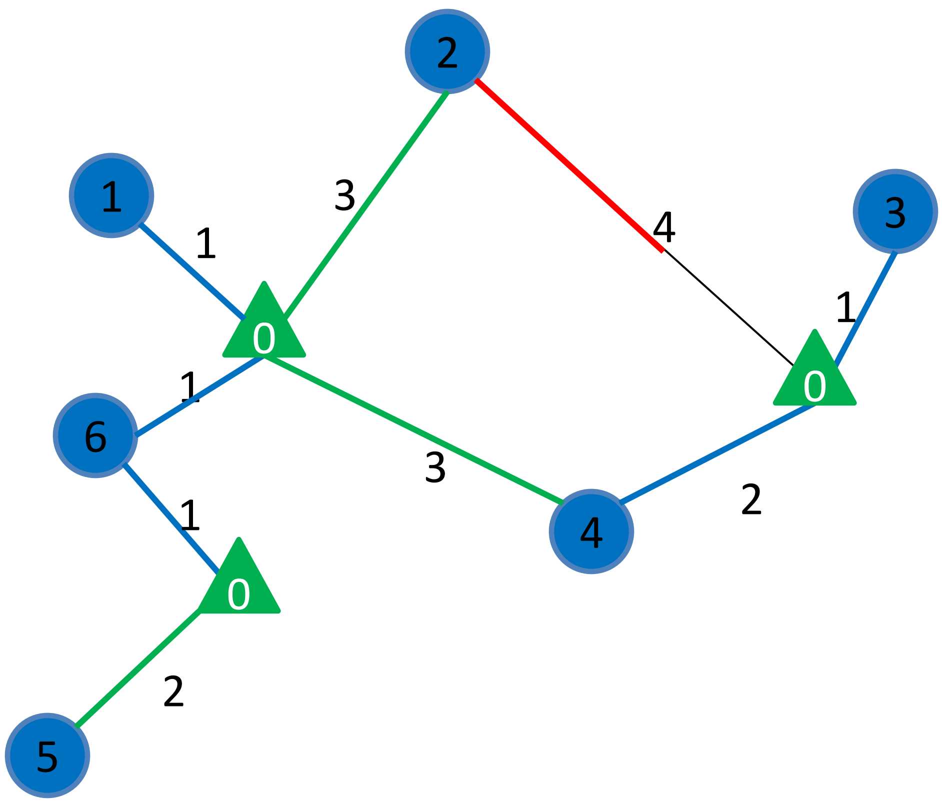


Οι «δεμένες» πόλεις αρχίζουν να προσφέρουν στο άνοιγμα των κέντρων στα οποία συνδέονται

Phase 1 t = 3

$$a_j = a_j + 1$$
$$\beta_{61} = 2 \quad \beta_{32} = 2$$
$$\beta_{11} = 1 \quad \beta_{42} = 1$$





Τα κέντρα που πληρώθηκαν εξολοκλήρου ονομάζονται «προσωρινά ανοιχτά» και οι πόλεις που συνεισέφεραν «συνδεδεμένες»

Phase 2

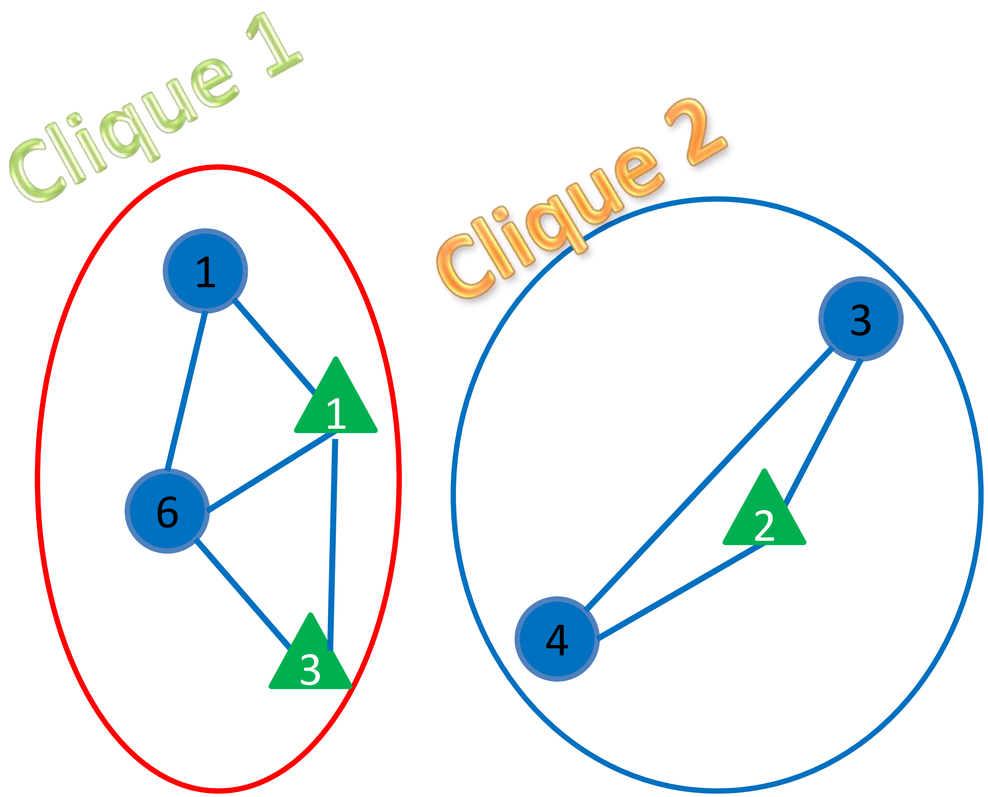
Connected cities

Temporary opened

special

$$T = (G, \mathrm{E}(G) = '\mathit{special}')$$

$$T^2 = (\mathrm{V}(T), \mathrm{E}(T) \cup \{(m,n), E(m,i)=1 \ \& \ E(i,n)=1\}) \quad i \in \mathrm{V}(T)$$



Όταν μία πόλη έχει συνεισφέρει στο άνοιγμα δύο κέντρων τότε το ένα πρέπει να κλείσει. Η επιλογή είναι τυχαία.

Phase 2

Λύση 1/2

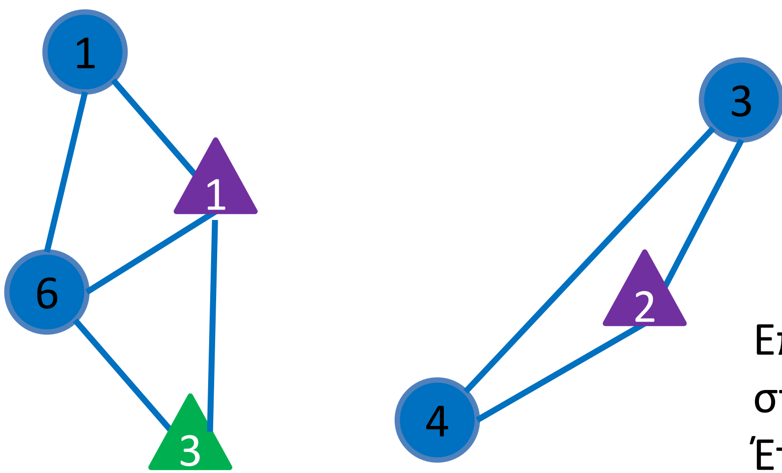
Connected cities

Opened

special

$$H = (T^2, \mathrm{V}(T^2) \in F)$$

$$S = MIS(H) = \{f_1, f_2\} \quad | \quad \{f_2, f_3\}$$



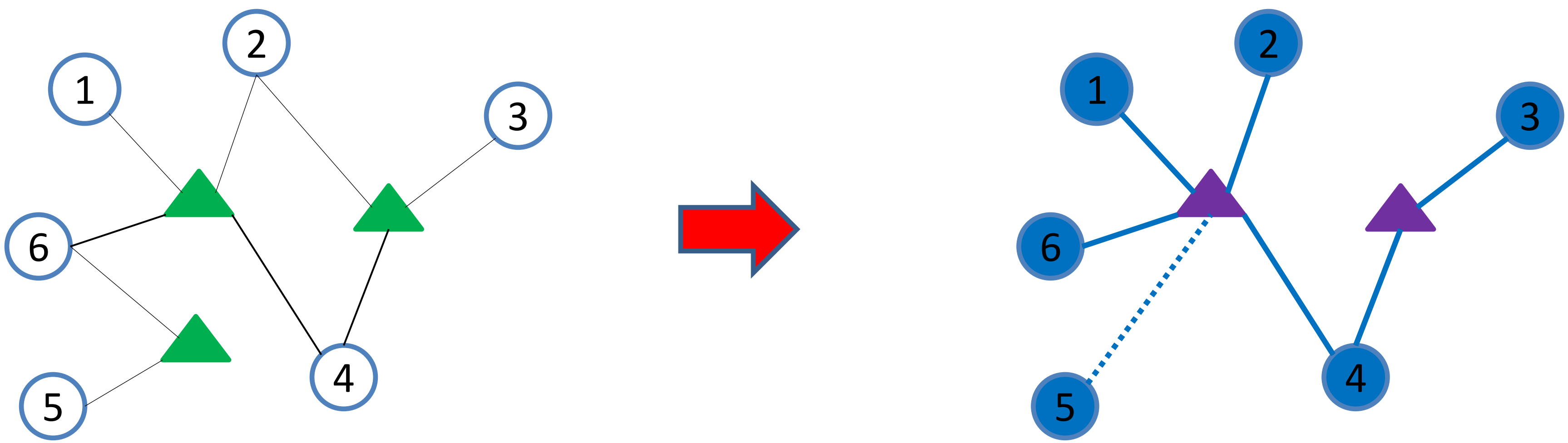
Επίλεξε σαν maximal independent set το $\{f_1, f_2\}$ στον υπογράφο **H** που περιέχει μόνο τα κέντρα. Έτσι, το «προσωρινά ανοικτό» κέντρο f_3 κλείνει και η πόλη **5** συνδέεται «έμμεσα» με το ποιο κοντινό, το f_1

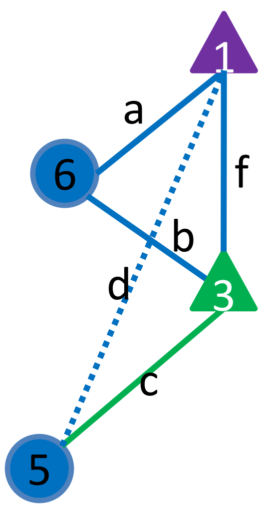
Phase 2

Λύση 1/2

Direct connected

Indirect connected





$$\left.\begin{aligned} a &\le c \\ b &\le c \\ a+b &\le f \\ c &\le a_j^e \end{aligned}\right\} \quad \left.\begin{aligned} & d \le f + c \Rightarrow \\ & d \le a+b+c \Rightarrow \\ & d \le 3*c \Rightarrow \\ & c_{ji} \le 3*a_j^e \end{aligned}\right\} \quad \begin{aligned} & \sum_{i\in F, j\in C} c_{ji}x_{ji} \le 3*\sum_{j\in C} a_j^e \Rightarrow \\ & \sum_{i\in F, j\in C} c_{ji}x_{ji} + 3\sum_{i\in F} f_i y_i \le 3*\sum_{j\in C} a_j \end{aligned}$$

Θεώρημα 24.7

Από το facility location στο

K-MEDIAN

Γραμμικός Προγραμματισμός για το k-median (1)

Πρωτεύων

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{i\in F,\, j\in C} c_{ij}x_{ij} & \\
\text{subject to} & \displaystyle\sum_{i\in F} x_{ij} \geq 1, & j \in C \\
& y_i - x_{ij} \geq 0, & i \in F,\ j \in C \\
& \displaystyle\sum_{i\in F} -y_i \geq -k & \\
& x_{ij} \geq 0, & i \in F,\ j \in C \\
& y_i \geq 0, & i \in F
\end{array}
$$

Δυϊκό

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{j\in C} \alpha_j - zk & \\
\text{subject to} & \alpha_j - \beta_{ij} \leq c_{ij}, & i \in F,\ j \in C \\
& \displaystyle\sum_{j\in C} \beta_{ij} \leq z, & i \in F \\
& \alpha_j \geq 0, & j \in C \\
& \beta_{ij} \geq 0, & i \in F,\ j \in C \\
& z \geq 0 &
\end{array}
$$

Γραμμικός Προγραμματισμός για το k-median (2)

- Οι μεταβλητές (**x,y**) στο πρωτεύον πρόβλημα έχουν την ίδια σημασία με τις αντίστοιχες το *facility location*

- Το *k-median* ανάγεται στο *facility location* αν θέσουμε το κόστος ανοίγματος **fi** κάθε κέντρου ίσο με **z** και βρούμε βέλτιστη κλασματική λύση με **k** κέντρα

- Στη βέλτιστη λύση (**x,y**) του facility location ισχύει

$$\sum_{i\in F,\, j\in C} c_{ij}x_{ij} + \sum_{i\in F} zy_i = \sum_{j\in C} a_j$$

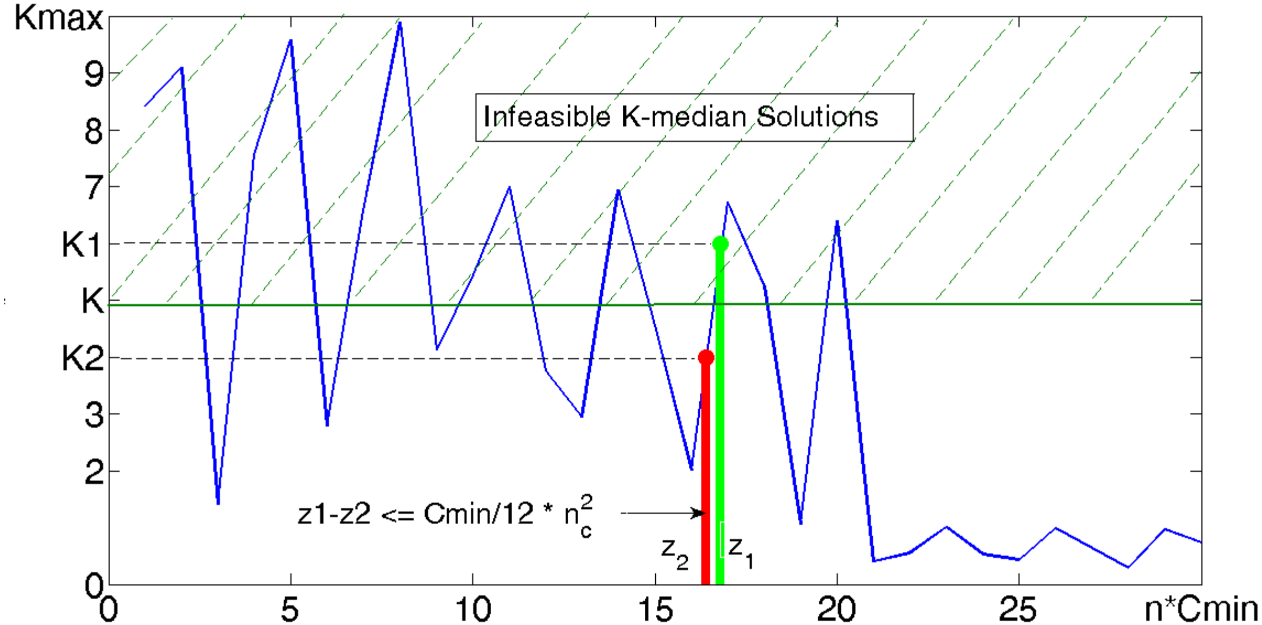
Συνεπώς η λύση (**x,y**) είναι βέλτιστη και για το *k-median* εφόσον ισχύει το κριτήριο της δϋικότητας

$$\sum_{i\in F,\, j\in C} c_{ij}x_{ij} = \sum_{j\in C} a_j - \sum_{i\in F} zy_i$$

- Αρκεί να βρούμε λύση στο *facility location* με $\sum_{i\in F} y_i = k$ και θα έχουμε κλασματική λύση για το *k-median* εώς **τρεις φορές** χειρότερη από τη βέλτιση

Αναζήτηση λύσης με **k** κέντρα

- Αυξάνοντας/μειώνοντας το κόστος **z** ο αλγόριθμος για το facility location χρησιμοποιεί λιγότερα/περισσότερα κέντρα **k**
- Γίνεται Δυαδική Αναζήτηση στο διάστημα $z \in [0, n \cdot c_{\min}]$ για εύρεση λύσεων (x^s, y^s) και (x^l, y^l) στις οποίες χρησιμοποιήθηκαν κέντρα $k_2 \leq k \leq k_1$ με κόστη z_1 z_2 με $z_2 - z_1 \leq (c_{\min}/12n^2)$
- Η λύση $(x, y) = a(x^s, y^s) + b(x^l, y^l)$ είναι νόμιμη για το k-median αν το $k = ak_1 + \beta k_2$ είναι το ζητούμενο είτε μικρότερο



Σφάλμα προσέγγισης

- Η προσέγγιση για το k-median είναι έως $(3+1/n_c)$ φορές χειρότερη από τη βέλτιστη κλασματική λύση όπου n_c το πλήθος των πόλεων

$$\left.\begin{aligned}
\sum_{i\in F,\ j\in C} c_{ij}x_{ij}^{s} &\le 3\Big(\sum_{j\in C}\alpha_j^{s} - z_1k_1\Big),\\
\sum_{i\in F,\ j\in C} c_{ij}x_{ij}^{l} &\le 3\Big(\sum_{j\in C}\alpha_j^{l} - z_2k_2\Big).\\
\textstyle\sum_{i\in F,\ j\in C} c_{ij}x_{ij}^{l} &\ge c_{\min}\\
z_1 &> z_2\\
z_1 - z_2 &\le (c_{\min}/12n_c^2).
\end{aligned}\right\} \quad \sum_{i\in F,\ j\in C} c_{ij}x_{ij} \le \left(3+\frac{1}{n_c}\right)\left(\sum_{j\in C}\alpha_j - z_1k\right)$$

Τυχαία Στρογγυλοποίηση της κλασματικής λύσης (1)

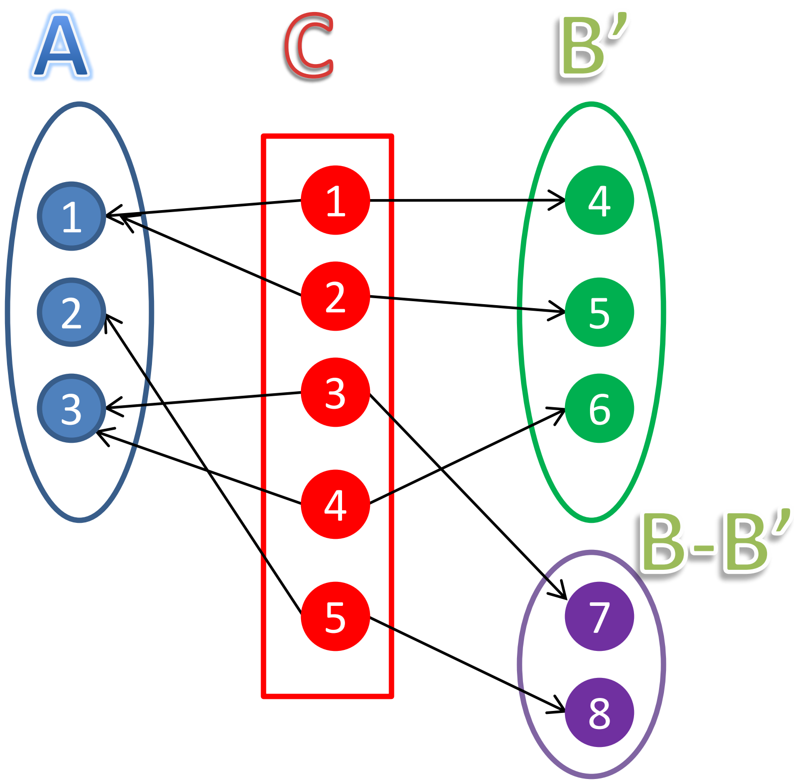
- $(x,y)=a(x^s,y^s)+b(x^l,y^l)$ η προσεγγιστική κλασματική λύση
- ***α,β*** *: το ποσοστό συμμετοχής των δύο λύσεων του facility location με* $k_1 \leq k$ *και* $k_2 \geq k$ *στη λύση του k-median*
- ***A,B*** *τα σύνολα των κέντρων για τις δύο λύσεις αντίστοιχα και* ***B'*** *τα πιο κοντινά κέντρα του* ***B*** *στα κέντρα του* ***A***
- Επιλέγεται τυχαία από τα **A,B**' με πιθανότητες **α** και **β=1-α** αντίστοιχα ένα σύνολο **k1** κέντρων και συμπληρωματικά ένα σύνολο **k-k1** κέντρων από το **B-B'**

$$I=(A'\subseteq A)\cup(B''\subseteq B')\cup(B'''\subseteq B-B')$$

$$|I|=k$$

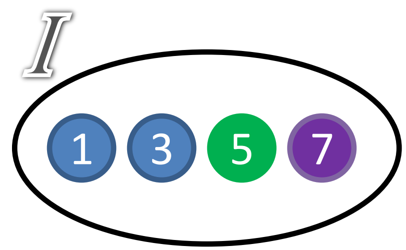
Τυχαία Στρογγυλοποίηση της κλασματικής λύσης (2)

- Μένει να βρούμε την αντιστοίχηση $\phi : C \to I$ κάθε πόλης $c \in C$ σε ένα από τα **k** κέντρα $f \in I$



$$a = 0.6, b = 1 - \alpha = 0.4$$

$$k_1 = 3, k_2 = 5, k = 4$$



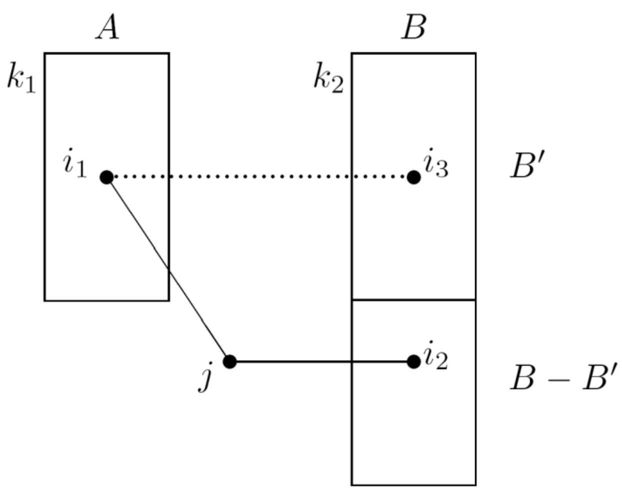
$$\phi(1) = 1, \phi(2) = 5$$

$$\phi(3) = 7, \phi(4) = 3$$

$$\phi(5) = 2$$

Αναμενόμενο Κόστος Ακέραιας Λύσης

- Το κόστος σύνδεσης της πόλης **j** στην κλασματική λύση είναι $ac_{i_1 j} + bc_{i_2 j}$
- Το αναμενόμενο κόστος σύνδεσης στην ακέραια λύση είναι
  - 1ος τρόπος σύνδεσης: $\mathbf{E}[c_{\phi(j)j}] = ac_{i_1 j} + bc_{i_2 j} = \mathrm{cost}(j)$
  - 2ος τρόπος σύνδεσης: $\mathbf{E}[c_{\phi(j)j}] \leq bc_{i_2 j} + a^2 c_{i_1 j} + abc_{i_3 j}$
- Συνολικά $\mathbf{E}[c_{\phi(j)j}] \leq (ac_{i_1 j} + bc_{i_2 j})(1 + \max(a,b))$



Η ακέραια λύση είναι το πολύ κατά **1+max(a,b)** χειρότερη από την κλασματική

$$\mathbf{E}\left[\sum_{i\in F,\ j\in C} c_{ij} x_{ij}^k\right] \leq (1+\max(a,b))\left(\sum_{i\in F,\ j\in C} c_{ij} x_{ij}\right)$$

Αφαίρεση Τυχαιότητας (Derandomization)

- Εύρεση ακέραιας λύσης ελαχίστου κόστους με δοκιμές μέσω του αναμενόμενου κόστους υπο συνθήκη
- Τα βήματα τυχαίας επολογής κέντρων και συνδέσεων αντικαθίστανται από επιλογές που δίνουν το μικρότερο κόστος
  - 1ο βήμα: Ανοιξε τα κέντρα του **A** ή του **B’** ανάλογα με το μικρότερο αναμενόμενο κόστος

$$\min(E[A, D], E[B', D]), \quad D \subseteq B - B', |D| = k - k_1$$

  - 2ο βήμα: Βάση των ήδη επιλεγμένων k_1 κέντρων από το βήμα 1, διάλεξε τα υπόλοιπα $k - k_1$ κέντρα από το **B-B’** δοκιμάζοντας και ελαχιστοποιώντας το κόστος υπο συνθήκη

$$E[D \cup \{i\}, B - (B' \cup D \cup \{i\}] \leq E[D, B - (B' \cup D]$$

$$D \subseteq B - B', |D| \leq k - k_1$$

Ταχύτητα και Συντελεστής Εκτίμησης

- Ο αλγόριθμος είναι 6-προσεγγιστικός
  - $3+\frac{1}{n_c}$ η προσέγγιση που δίνει ο αλγόριθμος του facility location
  - $1+\max(a,b)$ η κλιμάκωση του κόστους ύστερα από την τυχαία στρογγυλοποίηση
    - Το άνω φράγμα για το $\max(a,b)$ είναι $1+\frac{1}{n_c}$
  - Συνολικά, ο συντελεστής προσέγγισης είναι $(3+\frac{1}{n_c})(2+\frac{1}{n_c})<6$

- Ο χρόνος εκτέλεσης είναι της τάξης $O(m\cdot\log m\cdot(L+\log n))$
  - $O(m\cdot\log m)$ για τον αλγόριθμο του facility location
  - $O(\log_2(n^3\cdot c_{\max}/c_{\min}))=O(L+\log n)$ για τη δϋαδική αναζήτηση των k_1 k_2
  - Η τυχαία στρογγυλοποίηση χρειάζεται $O(n)$
  - Το derandomization χρειάζεται $O(m)$

Tight Example

Για την επαλήθευση του 6-προσεγγιστικού αλγόριθμου δεν υπάρχει οικογένεια παραδειγμάτων παρα μόνο για την επαλήθευση της προσέγγισης που δίνει η τυχαία στρογγυλοποίηση