Δέντρα με ρίζα.

Ένα κατευθυνόμενο γράφημα D ονομάζεται κατευθυνόμενο δέντρο, εάν από την αντικατάσταση των τόξων του D με ακμές προκύπτει ένα μη-κατευθυνόμενο γράφημα που είναι δέντρο.

π.χ.

Ένα κατευθυνόμενο δέντρο D λέμε ότι έχει ρίζα, εάν (i) υπάρχει σ' αυτό ακριβώς μια κορυφή u_0 με $d_D^-(u_0)=0$ και

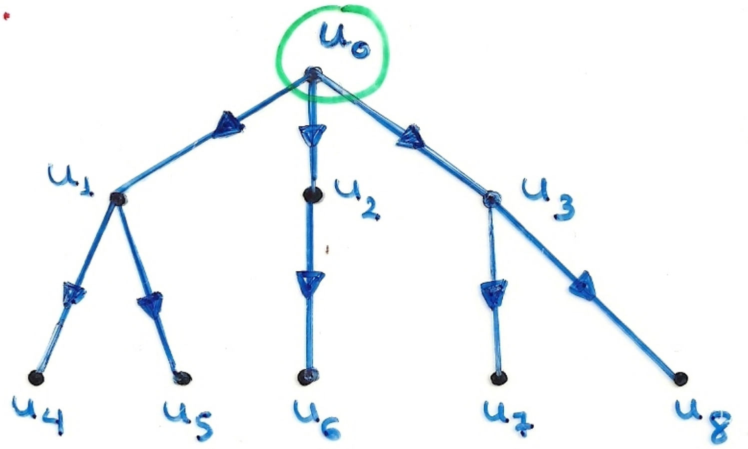
(ii) για όλες τις άλλες κορυφές u_i ισχύει ότι $d_D^-(u_i)=1$

u_0: Ρίζα του δέντρου D.

Φύλλα του D: κορυφές u_j του δέντρου D με $d_D^+(u_j)=0$.

Εσωτερικές κορυφές του D: Όλες οι κορυφές του D που δεν είναι φύλλα.

π.χ.



Δέντρο με ρίζα

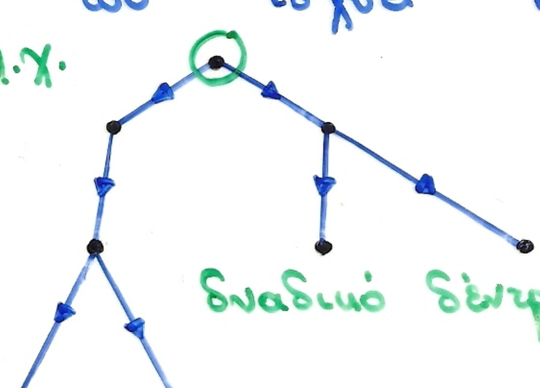
Παιδί της u: Μια κορυφή v για την οποία υπάρχει τόξο από την u στη v. Επίσης η u ονομάζεται πατέρας της v.

Αδέλφια: Κορυφές που είναι παιδιά της ίδιας κορυφής.

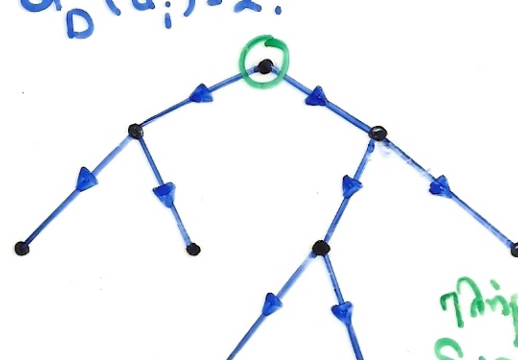
Απόγονος της u: Μια κορυφή v για την οποία υπάρχει κατευθυνόμενο μονοπάτι από την u στη v. Επίσης η u ονομάζεται πρόγονος της v.

Ένα κατευθυνόμενο δέντρο D με ρίζα ονομάζεται δυαδικό εάν $d^+_D(u_i) \le 2$ $\forall u_i \in V(D)$. Το D ονομάζεται πλήρες δυαδικό εάν για κάθε κορυφή u_i που δεν είναι φύλλο του ισχύει ότι $d^+_D(u_i) = 2$.

π.χ.



δυαδικό δέντρο



πλήρες δυαδικό δέντρο.

Παρατήρηση: Σ' ένα κατευθυνόμενο δέντρο με ρίζα, τα τόξα του μπορούν να αντικατασταθούν με ακμές, διότι η επιλογή μιας κορυφής ως ρίζας ουσιαστικά προσδιορίζει τις "κατευθύνσεις" των τόξων.

Θεώρημα: Σ' ένα πλήρες δυαδικό δέντρο T, ο αριθμός των φύλλων L σε σχέση με τον αριθμό των εσωτερικών κορυφών I ικανοποιούν την σχέση $L=I+1$.

Απόδειξη:

Τα φύλλα του T έχουν βαθμό 1.
Οι εσωτερικές κορυφές του T, εκτός της ρίζας, έχουν βαθμό 3, ενώ η ρίζα έχει βαθμό 2. Άρα

$$2|E(T)| = \sum_{x\in V(T)} d_T(x) = L + 3(I-1) + 2. \qquad -(*)$$

Όμως το T είναι δέντρο. Άρα $|V(T)| = |E(T)| + 1$

και επειδή $|V(T)| = L + I$ έχουμε από την

$(*)$, $L = I + 1$.

Ύψος κορυφής στο D: Το μήκος του μοναδικού μονοπατιού που συνδέει αυτή την κορυφή με την ρίζα του D.

Ύψος δέντρου: Μέγιστη τιμή ύψους κορυφής σ' αυτό το δέντρο.

Θεώρημα: Σ' ένα δυαδικό δέντρο T ύψους h υπάρχουν το πολύ 2^h φύλλα.

Απόδειξη:

Θα το αποδείξουμε με επαγωγή ως προς h. Εάν $h=1$ τότε προφανώς η πρόταση ισχύει διότι σ' αυτή την περίπτωση το T θα έχει 1 ή 2 φύλλα.

Έστω τώρα δυαδικό δέντρο T ύψους h.

Διαγράφουμε από το T τις ακμές που έχουν για άκρο τους, την ρίζα του T. Επειδή το T είναι δυαδικό, αυτές θα είναι το πολύ 2. Από την διαγραφή αυτή θα προκύψουν

(i) μια κορυφή βαθμού 0 (η ρίζα)

(ii) ένα ή δύο δυαδικά δέντρα ύψους το πολύ $h-1$.

Από την υπόθεση της επαγωγής καθένα από αυτά τα δυαδικά δέντρα ύψους το πολύ $h-1$, θα έχει το πολύ 2^{h-1} φύλλα. Η ένωση των συνόλων των φύλλων αυτών των δυαδικών δέντρων ισούται με το σύνολο των φύλλων του T. Άρα το T έχει το πολύ $2 \cdot 2^{h-1} = 2^h$ φύλλα.

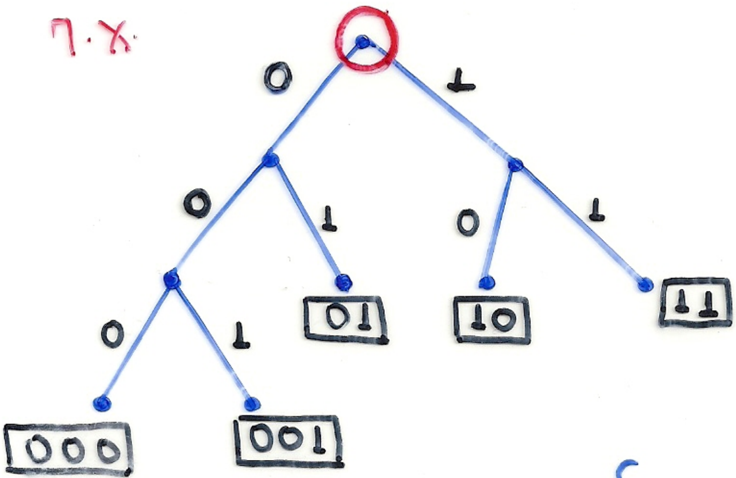
Κώδικες προθέματος. Αλγόριθμος Huffman.

Κώδικας προθέματος: Σύνολο ακολουθιών στο οποίο καμία ακολουθία δεν αποτελεί πρόθεμα κάποιας άλλης.

π.χ. $\{11, 100, 101, 00, 01\}$.

Από ένα πλήρες δυαδικό δέντρο μπορεί να προκύψει ένας κώδικας προθέματος.

π.χ.



$\{11, 10, 01, 001, 000\}$.

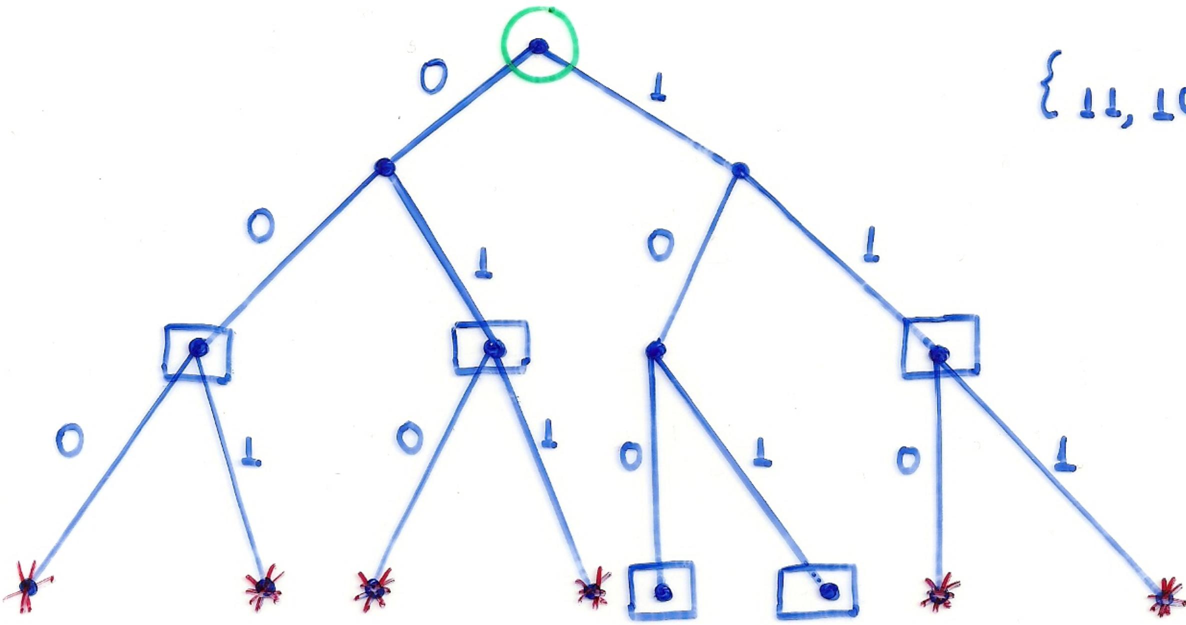
Αντίστροφα εάν έχουμε κάποιο κώδικα προθέματος αυτός μπορεί να προκύψει από ένα δυαδικό δέντρο

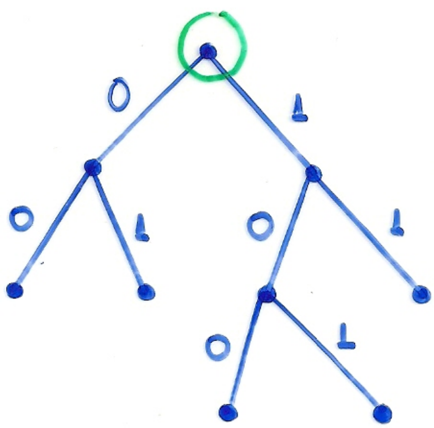
π.χ. $\{11, 100, 101, 00, 01\}$

Κατασκευάζουμε πλήρες δυαδικό δέντρο, στο οποίο κάθε ~~κορυφή~~ φύλλο έχει ύψος h, όπου h το μέγιστο μήκος ακολουθίας που αποτελεί στοιχείο του κώδικα προθέματος.

$h = 3$

$\{11, 100, 101, 00, 01\}$





Έστω ότι η πιθανότητα εμφάνισης του γράμματος a_i ισούται με p_i και έστω ότι με l_i συμβολίζουμε το μήκος της δυαδικής λέξης που θα χρησιμοποιήσουμε για να αναπαραστήσουμε το a_i.

Θέλουμε να βρούμε ένα κώδικα προθέματος για τον οποίο το

$$\sum_{i=1}^{n} l_i p_i$$

να είναι ελάχιστο.

Δηλαδή θέλουμε να βρούμε ένα πλήρες δυαδικό δένδρο για το οποίο το $\sum_{i=1}^{n} l_i p_i$ να είναι ελάχιστο, όπου l_i το μήκος του μονοπατιού που συνδέει την ρίζα με το φύλλο, στο οποίο έχει αντιστοιχηθεί η δυαδική ακολουθία που περιγράφει το a_i.

Ένα τέτοιο δένδρο ονομάζεται βέλτιστο.

p_i: Βάρος του φύλλου

Μπορούμε να υποθέσουμε ότι

$p_1 \leq p_2 \leq \ldots \leq p_n$.

Θεώρημα (Huffman 1952)

Ένα βέλτιστο δέντρο T με βάρη φύλλων $p_1, p_2, \ldots, p_n$ μπορεί να προκύψει από ένα βέλτιστο δέντρο T^* με βάρη φύλλων $p_1+p_2, p_3, \ldots, p_n$.

Αποδ. Θα αποδείξουμε πρώτα ότι υπάρχει βέλτιστο δέντρο T στο οποίο τα φύλλα που έχουν βάρη p_1, p_2 είναι αδέλφια. Έστω βέλτιστο δέντρο T με φύλλα $u_1, u_2, \ldots, u_n$ και αντίστοιχα βάρη $p_1, p_2, \ldots, p_n$. Έστω w εσωτερική κορυφή του T, της οποίας η απόσταση από την ρίζα u_0 είναι μέγιστη. Ας υποθέσουμε επίσης ότι u_x, u_y είναι τα δύο παιδιά της w. Προφανώς $d_T(u_0, u_x) \geq d_T(u_0, u_1)$ δηλ. $l_x \geq l_1$.

Εναλλάσσουμε τους ρόλους των φύλλων u_1, u_x και έστω T' το πλήρες δυαδικό δέντρο που προκύπτει απ' αυτή την εναλλαγή. Έχουμε

$$(\text{βάρος του } T') = (\text{βάρος του } T) - p_x l_x - p_1 l_1 + p_x l_1 + p_1 l_x =$$
$$= (\text{βάρος του } T) + (p_x - p_1)(l_1 - l_x).$$

Επειδή το T είναι βέλτιστο $l_1 \geq l_x$ ή $p_x = p_1$ $\quad (p_x \leq p_1)$

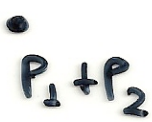
Έστω ότι $l_1 \geq l_x$. Όμως από προηγουμένως έχουμε $l_x \geq l_1$. Άρα $l_x = l_1$ και επομένως το T' είναι επίσης βέλτιστο.

Εάν $p_x = p_1$, πάλι το ίδιο συμπέρασμα προκύπτει. Χρησιμοποιώντας τα ίδια επιχειρήματα για το T' και εναλλάσσοντας τους ρόλους των φύλλων u_2, u_y παίρνουμε ένα βέλτιστο δέντρο T'', στο οποίο τα φύλλα που έχουν βάρη p_1, p_2 είναι αδέλφια.

Έστω T^* βέλτιστο δέντρο με βάρη φύλλων $p_1, p_2, \dots, p_n$ στο οποίο τα φύλλα που έχουν βάρη p_1, p_2 είναι αδέλφια.

(T^*)

Διαγράφουμε αυτά τα δύο φύλλα από το T^* καθώς επίσης και το πατέρα τους και τα αντικαθιστούμε μ'ένα φύλλο που έχει βάρος p_1+p_2.



(T^{**})

Έστω T^{**} το δέντρο που προκύπτει.

Το T^* θα έχει βάρη φύλλων $p_1+p_2, p_3, \dots, p_n$. Θα έχουμε

(βάρος T^*) = (βάρος T^{**}) + p_1+p_2

Έστω T βέλτιστο δέντρο με βάρη φύλλων $p_1+p_2, p_3, \dots, p_n$. Διαγράφουμε από το T το φύλλο που έχει βάρος p_1+p_2 και το αντικαθιστούμε με δύο φύλλα που έχουν βάρη p_1, p_2.

(T)

(T')

Έστω T' το γράφημα που θα προκύψει.

Θα έχουμε:

(βάρος T') = (βάρος T) + p_1+p_2.

Τώρα εάν

(βάρος T') > (βάρος T^*),

αυτό σημαίνει

(βάρος T) > (βάρος T^{**})

L ΑΤΟΠΟ
διότι το T είναι βέλτιστο

Άρα (βάρος T') ≤ (βάρος T^*)

δηλαδή το T' είναι βέλτιστο δέντρο

με βάρη φύλλων $p_1, p_2, \ldots, p_n$

Παράδειγμα: Να κατασκευασθεί βέλτιστο δέντρο με βάρη φύλλων $\boxed{4,5,8,15,16,25}$

$\boxed{9,8,15,16,25}$

8,9,15,16,25

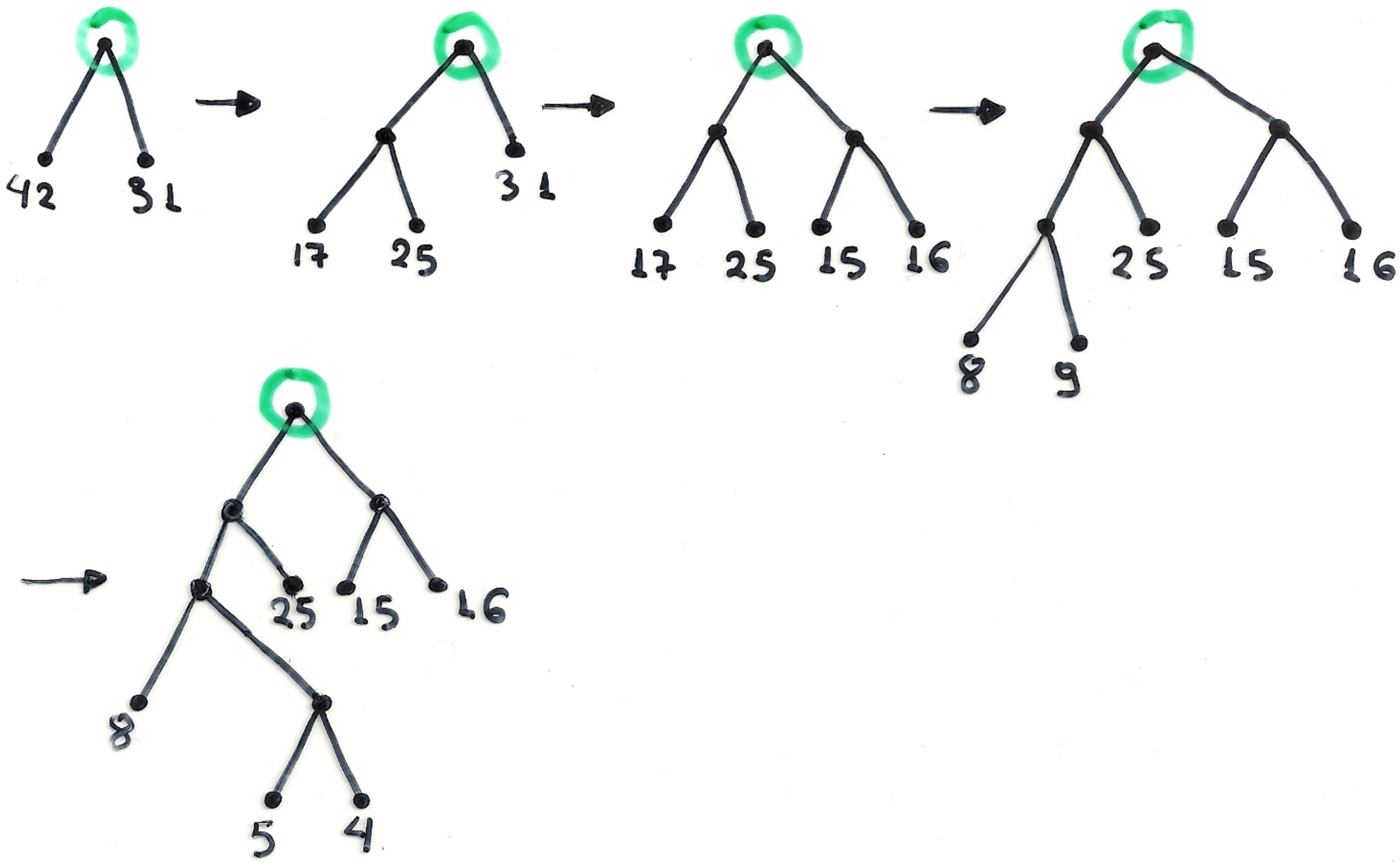
$\boxed{17,15,16,25}$

15,16,17,25

$\boxed{31,17,25}$

17,25,31

$\boxed{42,31}$



42 31
17 25 31
17 25 15 16
8 9 25 15 16
8 5 4 25 15 16

Μονοπάτια και αποστάσεις σε γραφήματα.

$\{u,v\}$: ακμή που έχει για άκρα της, τις κορυφές u και v.

Έστω γράφημα G και έστω συνάρτηση $w: E(G) \to \mathbb{R}^+$.

$w(u,v)$: η τιμή που αντιστοιχεί στην ακμή e, που έχει για άκρα της, τις κορυφές u και v, μέσω της w. (βάρος ή μήκος της e)

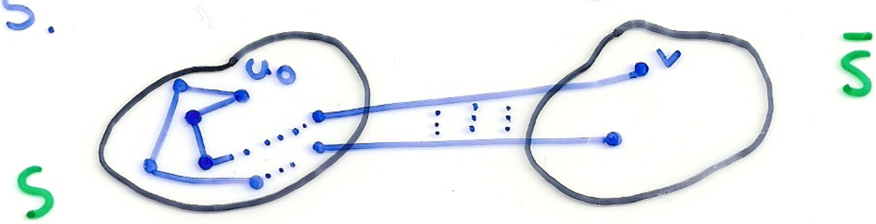
Μήκος μονοπατιού: άθροισμα βαρών (μηκών) ακμών του.

$d_G(u,v)$: μήκος συντομότερου (u,v)-μονοπατιού. (απόσταση μεταξύ των κορυφών u και v).

Πρόβλημα: Έστω γράφημα G, συνάρτηση $w: E(G) \to \mathbb{R}^+$ και έστω $u_0 \in V(G)$. Θέλουμε να προσδιορίσουμε τις αποστάσεις της u_0 από οποιαδήποτε άλλη κορυφή του G.

Η αλγοριθμική διαδικασία που μας λύνει το προηγούμενο πρόβλημα στηρίζεται στις εξής παρατηρήσεις:

α) Έστω $S \subseteq V(G)$ και έστω $u_0 \in S$. Ορίζουμε $\bar{S} = V(G) - S$. Για κάθε κορυφή $v \in \bar{S}$, θεωρούμε όλα τα (u_0, v)-μονοπάτια που (εκτός της v) δεν περιέχουν άλλη κορυφή του $\bar{S}$.

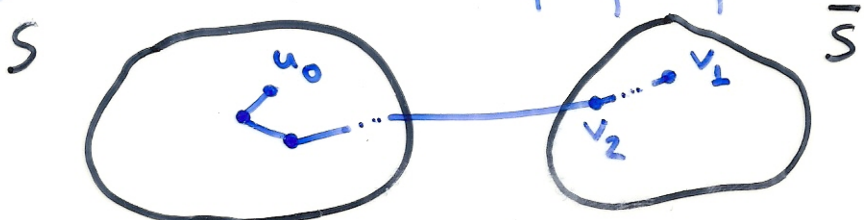


Με $\ell(v)$ θα συμβολίζουμε το μήκος ενός συντομότερου τέτοιου (u_0, v)-μονοπατιού. Εάν δεν υπάρχει τέτοιο μονοπάτι τότε θέτουμε όπου $\ell(v)$ το ∞.

$\ell(v)$: δείκτης της κορυφής v ως προς το σύνολο S.

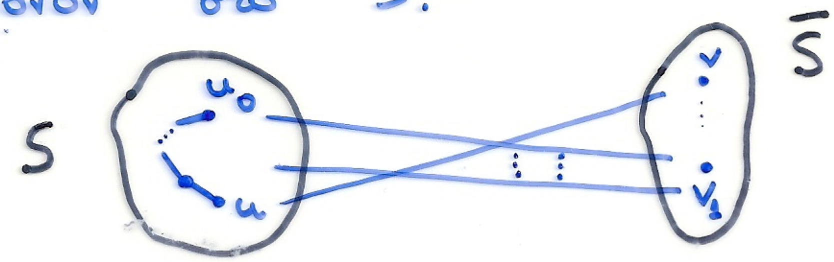
Έστω v_1 κορυφή με τον μικρότερο δείκτη. Σ' αυτή την περίπτωση ο $\ell(v_1)$ είναι ίσος με το μήκος του συντομότερου (u_0, v_1)-μονοπατιού.

Απόδειξη: Έστω ότι το παραπάνω δεν ισχύει και έστω ότι υπάρχει (u_0, v_1)-μονοπάτι με μήκος μικρότερο από $\ell(v_1)$.



Ένα τέτοιο μονοπάτι θα περιέχει μια ή περισσότερες κορυφές που ανήκουν στο σύνολο $\bar{S}-\{v_1\}$.
Έστω v_2 η πρώτη απ' αυτές τις κορυφές που συναντάμε σ' αυτό το μονοπάτι καθώς πηγαίνουμε από το u_0 στο v_1.
Προφανώς $\ell(v_2)<\ell(v_1)$ - ΑΤΟΠΟ διότι η v_1 είναι μια κορυφή με τον μικρότερο δείκτη.

b) Έστω $S \subseteq V(G)$ και $u_0 \in S$. Έστω επίσης ότι για κάθε κορυφή u που ανήκει στο S, υπάρχει συντομότερο (u_0,u)-μονοπάτι στο G που περιέχει κορυφές που ανήκουν μόνον στο S.

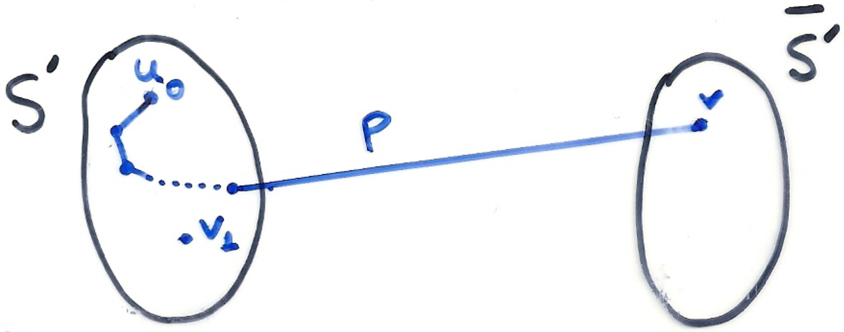


Υποθέτουμε ότι έχουμε προσδιορίσει τον δείκτες $\ell(v)$ ως προς S, για κάθε κορυφή v του $\bar{S}$.
Έστω $v_1 \in \bar{S}$. Ορίζουμε $S' = S \cup \{v_1\}$ και $\bar{S}' = \bar{S}-\{v_1\}$. Εάν με $\ell'(v)$ συμβολίζουμε τον δείκτη κάθε κορυφής v του $\bar{S}'$ ως προς S', θα έχουμε ότι:

$$\boxed{\ell'(v)=\min[\ell(v), \ell(v_1)+w(v,v_1)]^{*}} \quad -(1)$$

* Εάν δεν υπάρχει ακμή $\{v,v_1\}$ τότε θεωρούμε ότι $w(v,v_1)=\infty$.

Απόδειξη της 1:

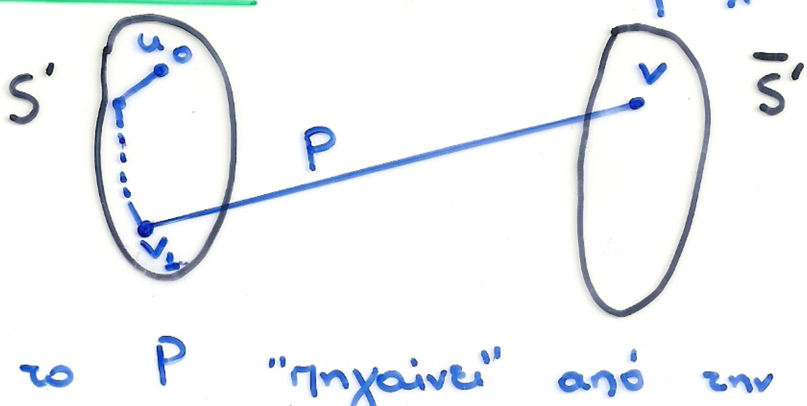


Θεωρούμε ένα (u_0, v)-μονοπάτι, το οποίο εκτός από την v δεν περιέχει άλλη κορυφή του $\bar{S}'$. Ας ονομάσουμε αυτό το μονοπάτι P και ας υποθέσουμε ότι είναι ελάχιστου μήκους ως προς την παραπάνω ιδιότητα.

<u>Περίπτωση 1</u>: Το P <u>δεν</u> περιέχει την v_1. Σ' αυτή την περίπτωση προφανώς

$$\ell'(v) = \ell(v).$$

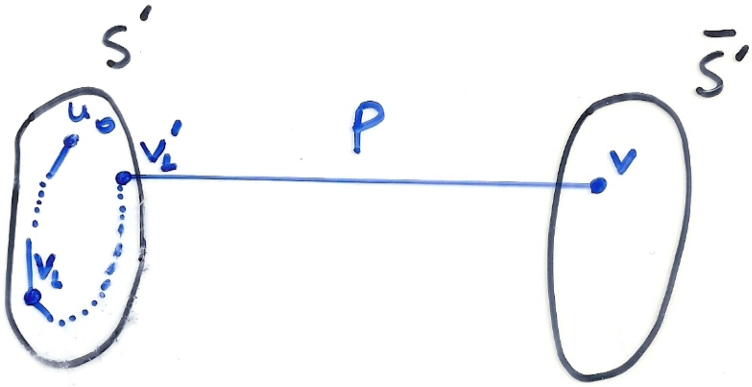
<u>Περίπτωση 2</u>: Το P περιέχει την v_1.



Εάν το P "πηγαίνει" από την u_0 στην v_1 ακολουθούμενο από την ακμή $\{v_1, v\}$, τότε θα έχουμε

$$\ell'(v) = \ell(v_1) + w(v_1, v)$$

Έστω τώρα ότι το P "πηγαίνει" από την u_0 στην v_1, μετά σε κάποια κορυφή v_1' του S και μετά στην v.



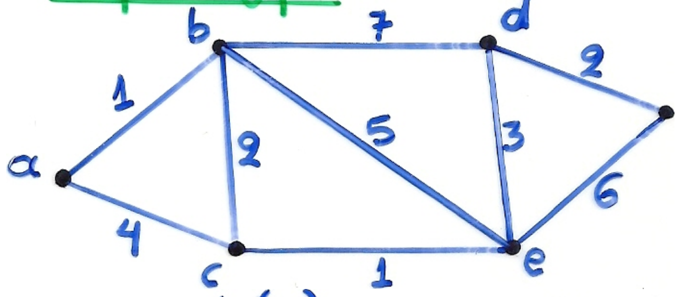
Από τον τρόπο κατασκευής του S,

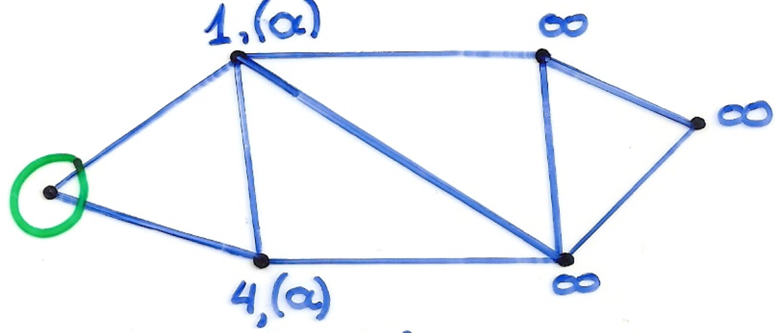
∃ συντομότερο (u_0, v_1')-μονοπάτι, του οποίου όλες οι κορυφές ανήκουν στο S. Το τελευταίο αυτό μονοπάτι ακολουθούμενο από την ακμή $\{v_1', v\}$ μπορεί να αντικαταστήσει το P, οπότε αυτή η δεύτερη υποπερίπτωση ανάγεται στην πρώτη.

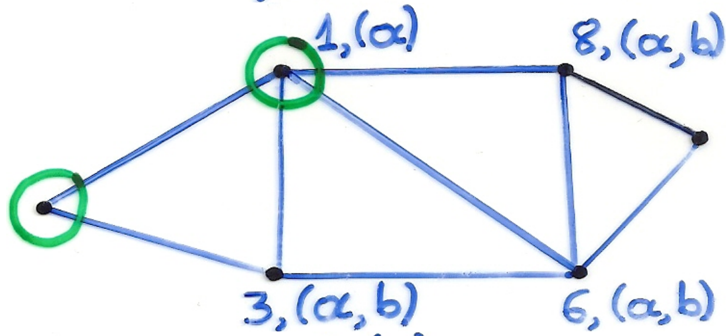
Αλγόριθμος του Dijkstra [1959] Whiting, Hiller [1960].

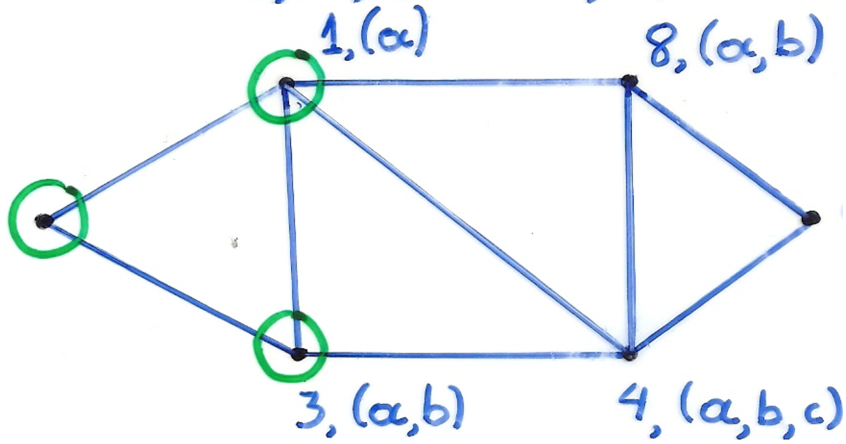
1. Θέτουμε $\ell(u_0)=0$, $\ell(v)=\infty$ για $v \neq u_0$, $S_0=\{u_0\}$ και $i=0$.
2. Για κάθε $v \in \bar{S}_i$, αντικαθιστούμε το $\ell(v)$ με το $\min\{\ell(v), \ell(u_i)+w(u_i,v)\}$. Υπολογίζουμε το $\min\limits_{v\in \bar{S}_i}\{\ell(v)\}$ και έστω u_{i+1} η κορυφή στην οποία επιτυγχάνεται αυτό το minimum. Θέτουμε $S_{i+1}=S_i \cup \{u_{i+1}\}$.
3. Εάν $i=|V(G)|-1$, ο αλγόριθμος διακόπτεται. Εάν $i<|V(G)|-1$, αντικαθιστούμε το i με το $i+1$ και πηγαίνουμε στο Βήμα 2.

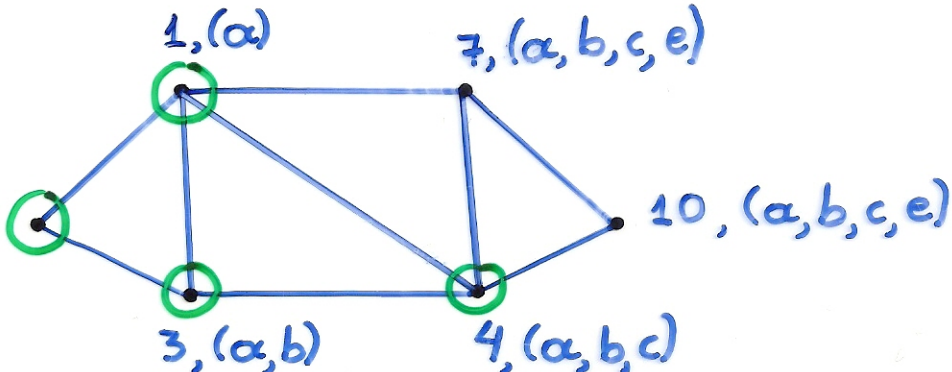
Παράδειγμα:

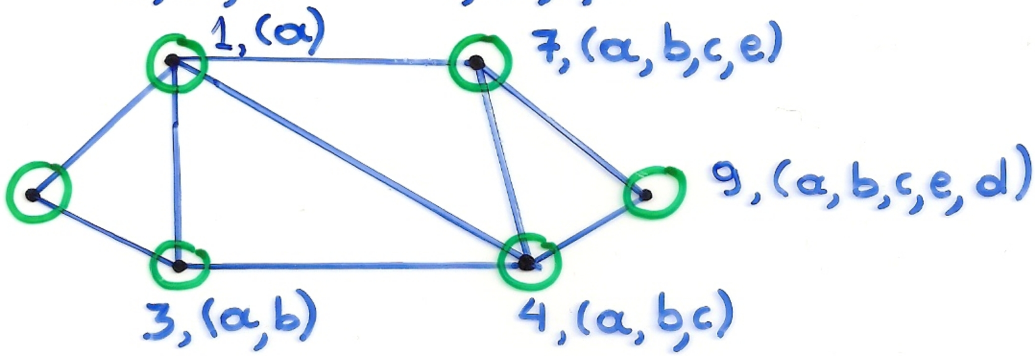




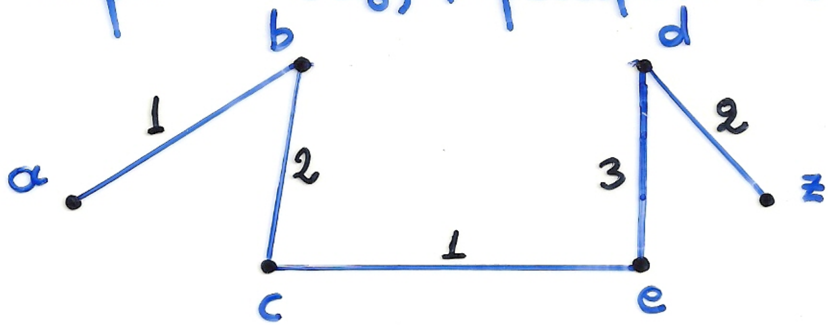








<u>Παρατήρηση</u>: Η ένωση των παραπάνω συντομότερων μονοπατιών αποτελεί ένα επικαλυπτικό δέντρο T για το γράφημα G, το οποίο έχει την εξής ιδιότητα: για κάθε $v \in V(T)$, το ένα και μοναδικό (u_0, v)-μονοπάτι στο T είναι ένα συντομότερο (u_0, v)-μονοπάτι στο G.



Εκκεντρικότητα κορυφών και κέντρο γραφήματος.

Έστω συνεκτικό γράφημα G και έστω συνάρτηση $w: E(G) \longrightarrow \mathbb{R}^+$

Εάν $u \in V(G)$, η εκκεντρικότητά της συμβολίζεται με $e(u)$ και ορίζεται ως εξής:

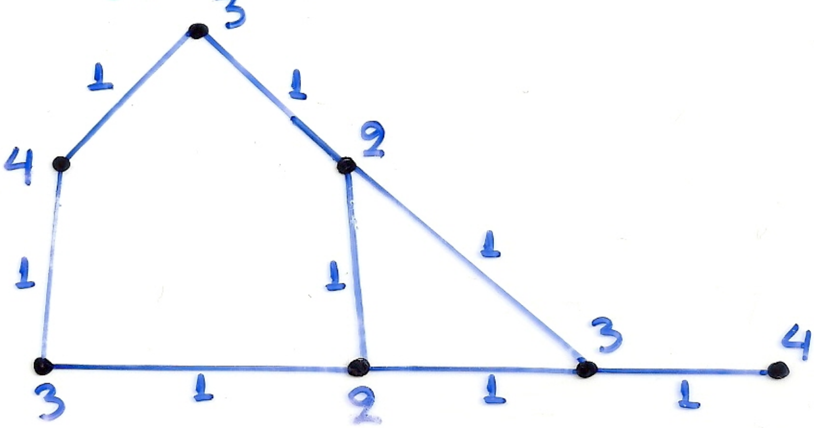
$$e(u) = \max\{ d_G(u,v) \mid v \in V(G)\}.$$

Η ακτίνα και διάμετρος του G συμβολίζονται με $\mathrm{rad}G$, $\mathrm{diam}G$ αντίστοιχα και ορίζονται ως εξής:

$$\mathrm{rad}G = \min\{ e(v) \mid v \in V(G)\}$$
$$\mathrm{diam}G = \max\{ e(v) \mid v \in V(G)\}.$$

<u>Παρατήρηση</u>: Η εκκεντρικότητα μιας κορυφής είναι ουσιαστικά η απόστασή της από την πλέον απομακρυσμένη από αυτήν κορυφή στο γράφημα. Επίσης την διάμετρο ενός γραφήματος G μπορούμε να την θεωρήσουμε ως την μεγαλύτερη απόσταση μεταξύ ζεύγους κορυφών στο G.

Παράδειγμα:



$$radG = 2 \qquad diamG = 4$$

Θεώρημα: Για κάθε συνεκτικό γράφημα G,

$$radG \le diamG \le 2\,radG.$$

Αποδ. Έστω $u, v \in V(G)$ όπου

$$d_G(u,v) = diamG.$$

Έστω επίσης κορυφή t του G, με

$$e(t) = radG.$$

Έχουμε ότι

$$diamG = d_G(u,v) \le d_G(u,t) + d_G(t,v) \le 2\,radG.$$

Το κέντρο ενός συνεκτικού γραφήματος G συμβολίζεται με $C(G)$ και ορίζεται ως το υπογράφημα του G που παράγεται από όλες εκείνες τις κορυφές που έχουν εκκεντρικότητα ίση με radG.

Για το γράφημα G, του προηγούμενου παραδείγματος έχουμε ότι

$$C(G) \simeq K_2$$

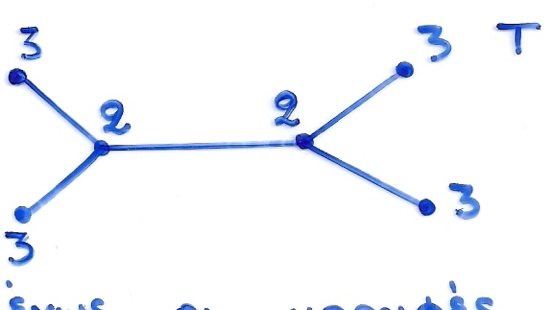
Θεώρημα (Jordan 1869, Sylvester 1873):

Έστω δένδρο T, στο οποίο κάθε ακμή έχει βάρος 1. Τότε είτε $C(T) \simeq K_1$ ή $C(T) \simeq K_2$.

Αποδ. Εάν $T \simeq K_1$ ή $T \simeq K_2$, τότε προφανώς το θεώρημα ισχύει.

Έστω S το σύνολο των κορυφών του T που έχουν βαθμό 1 και έστω $T' = T - S$. Προφανώς το T' είναι επίσης δένδρο. Εάν u είναι μια οποιαδήποτε κορυφή στο αρχικό δένδρο T, η μέγιστη απόστασή της από οποιαδήποτε άλλη κορυφή v, θα επιτυγχάνεται μόνον όταν η v έχει βαθμό 1. Άρα οι εκκεντρικότητες των κορυφών στο T' θα είναι κατά μία μικρότερες απ' αυτές που είχαμε στο T.

π.χ.



Επομένως οι κορυφές που είχαν ελάχιστη εκκεντρικότητα στο T είναι ακριβώς οι ίδιες κορυφές που έχουν ελάχιστη εκκεντρικότητα τώρα στο T'.

Δηλαδή τα T και T' έχουν το ίδιο κέντρο.

Εάν συνεχισθεί αυτή η διαδικασία διαγραφής κορυφών που έχουν βαθμό 1, στο τέλος θα προκύψει είτε το δένδρο K_1 ή το δένδρο K_2, το οποίο θα έχει το ίδιο κέντρο με όλα τα προηγούμενα και επομένως και με το T.

Η προηγούμενη απόδειξη μας δίνει τον παρακάτω αλγόριθμο για την εύρεση του κέντρου ενός δένδρου T.

1. Θέτουμε $T' = T$.
2. Εάν $T' \simeq K_1$ ή $T' \simeq K_2$, τότε $C(T) = T'$, διαφορετικά εκτελούμε το Βήμα 3.
3. Διαγράφουμε από το T' τις κορυφές που έχουν βαθμό 1. Έστω T'' το δένδρο που προκύπτει. Επαναλαμβάνουμε το Βήμα 2, χρησιμοποιώντας ως T' το T''.

Άσκηση.

Να βρεθεί το κέντρο του παρακάτω δέντρου, χρησιμοποιώντας τον προηγούμενο αλγόριθμο.

Απ.