

Εξόρυξη από Γραφήματα

Οι διαφάνειες στηρίζονται στο P.-N. Tan, M. Steinbach, V. Kumar, «Introduction to Data Mining», Addison Wesley, 2006



Εισαγωγή

Γράφημα
 $G(V, E)$
 $u_i \in V, (u_i, u_j) \in E$

Ετικέτα $l(u_i), l(u_i, u_j)$

Επέκταση της εξόρυξης κανόνων συσχέτισης για τον εντοπισμό συχνών υπο-γραφημάτων

Γράφημα με ετικέτες

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 2

Εισαγωγή

Εφαρμογές:

Web Mining:
 Γράφος (web browsing patterns) Κόμβοι (web pages) Ακμές (hyperlinks)
 Υπολογιστική Χημεία
 Γράφος (structure of chemical compounds) Κόμβοι (atoms, ions)
 Ακμές (bonds)

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 3

Εισαγωγή

Εφαρμογές:

Δίκτυα Υπολογιστών:
 Γράφος (computer network) Κόμβοι (computers, servers) Ακμές (interconnections)

Semantic Web
 Γράφος (collections of XML documents) Κόμβοι (XML elements) Ακμές (parent-child)

Bioinformatics
 Γράφος (protein structures) Κόμβοι (amino acids) Ακμές (contact residue)

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 4

Εισαγωγή

Aspirin

Yeast protein interaction network

Internet

Co-author network

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 5

Γραφήματα: Ορισμοί

Υπο-γράφημα
 $G'(V', E')$ υπο-γράφημα του $G(V, E)$:
 $V' \subseteq V$
 $E' \subseteq E$
 $G' \subseteq G$

Γράφημα

Υπογράφημα

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 6

Γραφήματα: Ορισμοί

Έστω μια συλλογή από γραφήματα SG , η υποστήριξη ενός υπογραφήματος g είναι το ποσοστό όλων των γραφημάτων του SG που περιέχουν το g ως υπογράφημά τους

$$s(g) = \frac{|\{G_i \mid g \subseteq G_i, G_i \in SG\}|}{|SG|}$$

Figure 7.10. Computing the support of a subgraph from a set of graphs.

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 7

Εξόρυξη Συχνών Υπογραφημάτων

Εξόρυξη Συχνών Υπογραφημάτων

Έστω ένα σύνολο γραφημάτων SG
και ένα κατώφλι υποστήριξης minsup

Βρες όλα τα υπογραφήματα g τέτοια ώστε $s(g) \geq \text{minsup}$

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 8

Εξόρυξη Συχνών Υπογραφημάτων

Θα εστιάσουμε σε μη κατευθυνόμενα, συνδεδεμένα γραφήματα

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 9

Εξόρυξη Συχνών Υπογραφημάτων

Μπορούμε να ακολουθήσουμε τη brute-force μέθοδο,
Δηλαδή παράγουμε όλα τα πιθανά υπογραφήματα
Υπολογίσει την υποστήριξη του καθενός
Κράτησε μόνο όσους έχουν υποστήριξη $\geq \text{minsup}$

Μέγεθος του search space:

$$\sum_{i=1}^d \binom{d}{i} 2^{i(i-1)/2}$$

- Επιλογή i από d κόμβους
- Διαφορετικές συνδέσεις, μέγιστος αριθμός για μη κατευθυνόμενο είναι $i(i-1)/2$ (πλήρως συνδεδεμένος)

Πολύ μεγάλος αριθμός, πχ για $d = 7$, αριθμός στοιχειοσυνόλων, $2^d = 128$, αριθμός γραφημάτων 2,350.602!
Κάποια βέβαια, μη συνδεδεμένα άρα (σχετικά) μικρότερος αριθμός

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 10

Εξόρυξη Συχνών Υπογραφημάτων

Παράδειγμα

Ετικέτες για τις ακμές $\{p, q\}$

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 11

Εξόρυξη Συχνών Υπογραφημάτων

Ένα στοιχείο μόνο μια φορά σε ένα στοιχειοσύνολο, ενώ η ετικέτα ενός κόμβου πολλές φορές σε ένα γράφημα

Για το ίδιο ζευγάρι κόμβων, έχουμε διαφορετικές επιλογές για την ετικέτα των ακμών τους

Εξόρυξη Δεδομένων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 12

Αλγόριθμος βασισμένος στον αρτιοί

Μπορούμε να χρησιμοποιήσουμε έναν αλγόριθμο βασισμένο στον αρτιοί όπως στα συχνά στοιχειοσύνολα:

Η αρχή αρτιοί ισχύει: γράφημα συχνό => υπογραφήματα του συχνά

Γραφήματα ως δοσοληψίες, πως;

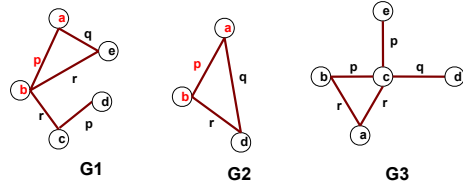
Στοιχείο: τριάδα (κόμβος1, κόμβος2, ακμή) $l(u_1), l(u_2), l(e)$

Πλάτος δοσοληψίας: αριθμός των ακμών

Δουλεύει μόνο αν οι τριάδες είναι μοναδικές

Αλγόριθμος βασισμένος στον αρτιοί

Παράδειγμα: Δοσοληψίες ως Γραφήματα



	(a,b,p)	(a,b,q)	(a,b,r)	(b,c,p)	(b,c,q)	(b,c,r)	...	(d,e,r)
G1	1	0	0	0	0	1	...	0
G2	1	0	0	0	0	0	...	0
G3	0	0	1	1	0	0	...	0
G3

Αλγόριθμος βασισμένος στον αρτιοί

Βήμα 1:

Βρες τα συχνά 1-υπογραφήματα

Βήμα 2:

Επανάλαβε μέχρι να μην παράγονται νέα συχνά υπογραφήματα

- **Δημιουργία Υποψηφίων - Candidate Generation:**
 - Χρησιμοποίησε τα k-1 υπογραφήματα για τη δημιουργία k-υπογραφημάτων
- **Ψαλίδισμα Υποψηφίων - Candidate Pruning:**
 - Ψαλίδισε τα k-υπογραφήματα που περιέχουν μη συχνά (k-1)-υπογραφήματα
- **Υπολογισμός Υποστήριξης - Support Counting:**
 - Υπολόγισε την υποστήριξη των k-υπογραφημάτων που απομένουν
- **Υπολογισμός Υποψηφίων - Candidate Elimination:**
 - Διώξε τα υποψήφια k-υπογραφήματα που δεν είναι συχνά

Στην πραγματικότητα δεν είναι τόσο απλό, πολλά περισσότερα θέματα

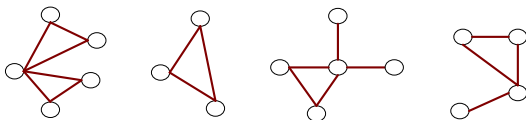
Αλγόριθμος βασισμένος στον αρτιοί

Επίσης,

- Additional constraints imposed by pattern structure
- Support and confidence are not the only constraints
- Assumption: frequent subgraphs must be connected

Αλγόριθμος βασισμένος στον αρτιοί

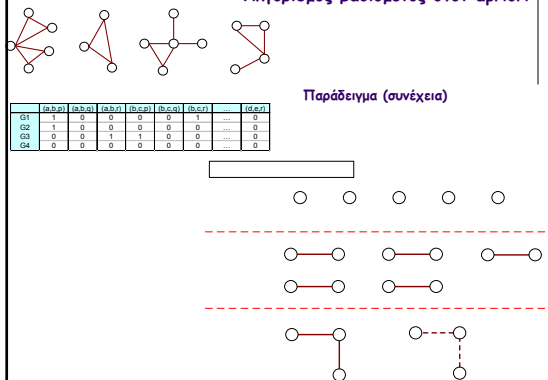
Παράδειγμα: Σύνολο Γραφημάτων



	(a,b,p)	(a,b,q)	(a,b,r)	(b,c,p)	(b,c,q)	(b,c,r)	...	(d,e,r)
G1	1	0	0	0	0	1	...	0
G2	1	0	0	0	0	0	...	0
G3	0	0	1	1	0	0	...	0
G4	0	0	0	0	0	0	...	0

Αλγόριθμος βασισμένος στον αρτιοί

Παράδειγμα (συνέχεια)





Δημιουργία Υποψηφίων Υπο-γραφημάτων

Από k-1 υπογραφήματα σε k-υπογραφήματα

Τι είναι το k:

- Αριθμός κόμβων; **Ανάπτυξη Κόμβων (Vertex Growing)**
- Αριθμός Ακμών; **Ανάπτυξη Ακμών (Edge Growing)**



Δημιουργία Υποψηφίων Υπο-γραφημάτων

Αποφυγή δημιουργίας του ίδιου υπο-γραφήματος

Τα δύο k-1 υπογραφήματα που συγχωνεύουμε

- πρέπει να έχουν ένα κοινό k-2 υπογράφημα

Το κοινό υπο-γράφημα ονομάζεται ο **πυρήνας** τους (*core*)



- Στον Αρριόρι:

Η συγχώνευση δύο συχνών (k-1)-στοιχειοσυνόλων δημιουργεί ένα υποψήφιο k-στοιχειοσύνολο

- Στην εξόρθξη συχνών υπογραφημάτων (frequent subgraph mining):

Η συγχώνευση δύο συχνών (k-1)-υπογραφημάτων μπορεί να δημιουργήσει παραπάνω από ένα υποψήφιο k-υπογράφημα



Ανάπτυξη Κόμβων

Συγχώνευση δύο υπογραφημάτων με **k-1 κόμβους**

Οι οποίοι έχουν ένα *κοινό υπο-γράφημα* (πυρήνα) με k-2 κόμβους

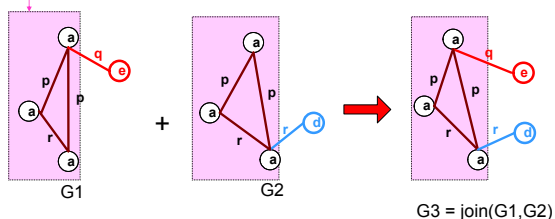
-> νέο υπογράφημα με **k κόμβους**



Ανάπτυξη Κόμβων

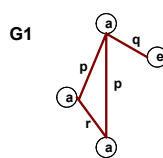
πυρήνας

2 υπογραφήματα με 4 κόμβους και ένα κοινό 3-υπογράφημα



Αναπαράσταση Γραφημάτων

Πίνακας Γειτνίασης (Adjacency Matrix)



$$M_{G1} = \begin{pmatrix} 0 & p & p & q \\ p & 0 & r & 0 \\ p & r & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix} \begin{matrix} a \\ a \\ a \\ e \end{matrix}$$

Δημιουργία Υποψηφίων

Ανάπτυξη Κόμβων

Ο πίνακας γειτνίασης M_1 συγχωνεύεται με τον πίνακα γειτνίασης M_2

αν οι υπο-πίνακες που προκύπτουν σβήνοντας τη τελευταία γραμμή και στήλη του M_1 και M_2 είναι ίδιοι

$$M_{G_1} = \begin{pmatrix} 0 & p & p & q \\ p & 0 & r & 0 \\ p & r & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix} \quad M_{G_2} = \begin{pmatrix} 0 & p & p & 0 \\ p & 0 & r & 0 \\ p & r & 0 & r \\ 0 & 0 & r & 0 \end{pmatrix}$$

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ 25

Δημιουργία Υποψηφίων

Ανάπτυξη Κόμβων

$G_3 = \text{join}(G_1, G_2)$

$$M_{G_3} = \begin{pmatrix} 0 & p & p & 0 & q \\ p & 0 & r & 0 & 0 \\ p & r & 0 & r & 0 \\ 0 & 0 & r & 0 & 0 \\ q & 0 & 0 & 0 & 0 \end{pmatrix}$$

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ 26

Δημιουργία Υποψηφίων

Ανάπτυξη Κόμβων

Ο πίνακας γειτνίασης M_1 συγχωνεύεται με τον πίνακα γειτνίασης M_2 **αν και μόνο αν** οι υπο-πίνακες που προκύπτουν σβήνοντας τη τελευταία γραμμή και στήλη του M_1 και M_2 είναι ίδιοι

Το **αποτέλεσμα** είναι ο πίνακας γειτνίασης M_3 που έχει τον ίδιο υπο-πίνακα επεκταμένος με τις τελευταίες γραμμές των M_1 και M_2

Αρκεί;

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ 27

Δημιουργία Υποψηφίων

Ανάπτυξη Κόμβων: Πολλαπλοί Υποψηφίοι

$$M_{G_1} = \begin{pmatrix} 0 & p & p & q \\ p & 0 & r & 0 \\ p & r & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix} \quad M_{G_2} = \begin{pmatrix} 0 & p & p & 0 \\ p & 0 & r & 0 \\ p & r & 0 & r \\ 0 & 0 & r & 0 \end{pmatrix} \quad M_{G_3} = \begin{pmatrix} 0 & p & p & 0 & q \\ p & 0 & r & 0 & 0 \\ p & r & 0 & r & 0 \\ 0 & 0 & r & 0 & ? \\ q & 0 & 0 & ? & 0 \end{pmatrix}$$

Πρέπει να θεωρήσουμε όλες τις πιθανές ετικέτες για την ακμή (d, e) => παραπάνω από ένα υποψήφιο υπογράφημα

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ 28

Δημιουργία Υποψηφίων

Ανάπτυξη Ακμών

Συγχώνευση δύο υπογραφημάτων με **k-1 ακμές**

Οι οποίοι έχουν ένα **κοινό υπο-γράφημα**

-> νέο υπογράφημα με **k ακμές**

Το νέο υπογράφημα μπορεί να μην έχει περισσότερους κόμβους από τα αρχικά

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ 29

Δημιουργία Υποψηφίων

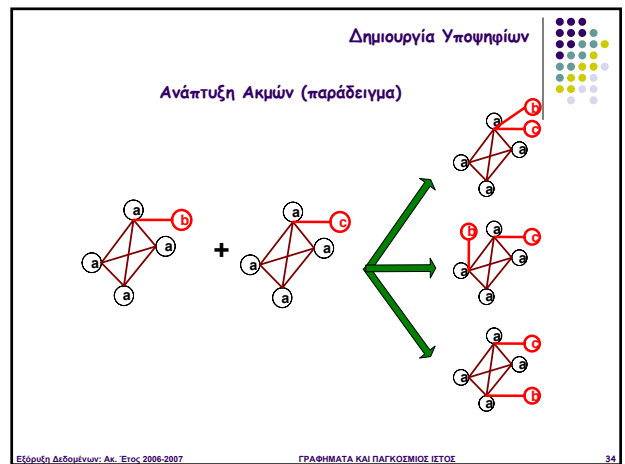
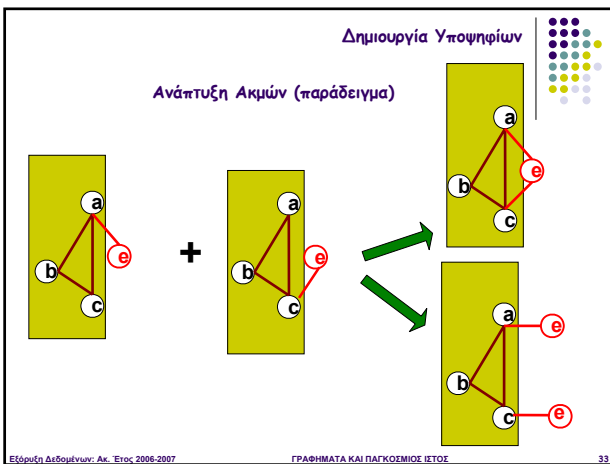
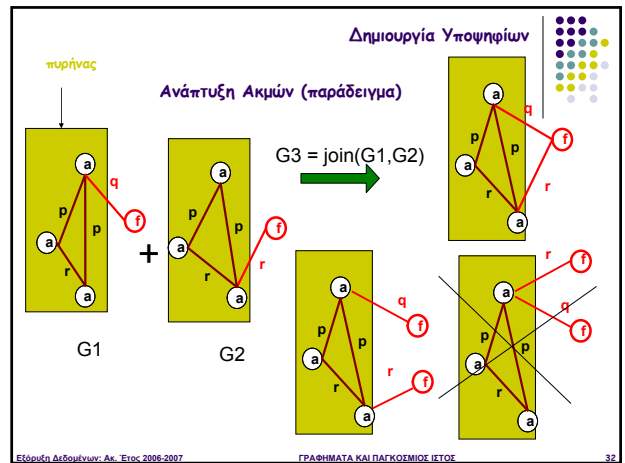
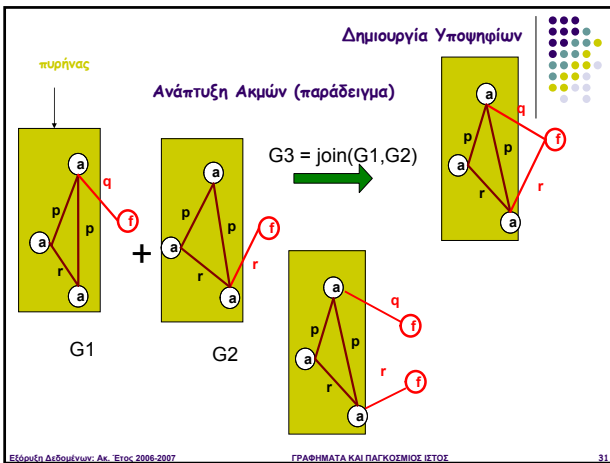
Ανάπτυξη Ακμών

Ενας συχνό k-1 υπογράφημα g_1 συγχωνεύεται με ένα συχνό k-1 υπογράφημα g_2 αν και μόνο όταν αφαιρέσουμε μια ακμή από το g_1 προκύπτει ένα k-2 υπογράφημα που είναι **τοπολογικά ισοδύναμο** ή **ισομορφικό (isomorphic)** με ένα k-2 υπογράφημα που προκύπτει αν αφαιρέσουμε μια ακμή από τον g_2

Το **αποτέλεσμα** είναι ένα k υπογράφημα που προκύπτει αν προσθέσουμε την ακμή που αφαιρέσαμε από το g_2 στο g_1

Θα δούμε είναι η «τοπολογική ισοδυναμία»

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΣΤΟΣ 30



Δημιουργία Υποψηφίων

Ανάπτυξη Ακμών

Ενας συχνό k-1 υπογράφημα g_1 συγχωνεύεται με ένα συχνό k-1 υπογράφημα g_2 αν και μόνο όταν αφαιρέσουμε μια ακμή από το g_1 προκύπτει ένα k-2 υπογράφημα που είναι **τοπολογικά ισοδύναμο** ή **ισομορφικό (isomorphic)** με ένα k-2 υπογράφημα που προκύπτει αν αφαιρέσουμε μια ακμή από τον g_2

Το αποτέλεσμα είναι ένα k υπογράφημα που προκύπτει αν προσθέσουμε την ακμή που αφαιρέσαμε από το g_2 στο g_1

Παρατήρηση: πολλά διαφορετικά υπογράφημα όταν υπάρχουν στον πυρήνα «ισοδύναμες» διαφορετικές θέσεις για να τοποθετήσουμε την ακμή - δηλαδή, τοπολογικά ισοδύναμοι κόμβοι

Εξέλιξη Διδασκάλων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 35

Δημιουργία Υποψηφίων

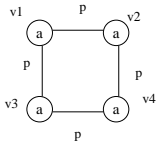
Ανάπτυξη Ακμών

Δυο γράφοι είναι **τοπολογικά ισοδύναμοι** αν υπάρχει μια 1-1 απεικόνιση μεταξύ των κόμβων τους

Εξέλιξη Διδασκάλων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 36



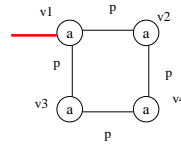
Ανάπτυξη Ακμών



Πρόσθεσε ακμή



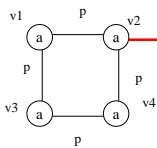
Ανάπτυξη Ακμών



Πρόσθεσε ακμή



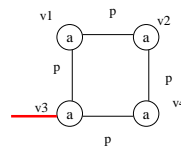
Ανάπτυξη Ακμών



Πρόσθεσε ακμή



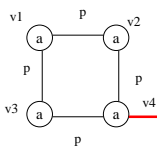
Ανάπτυξη Ακμών



Πρόσθεσε ακμή



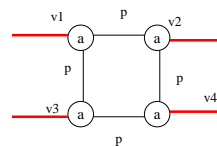
Ανάπτυξη Ακμών



Πρόσθεσε ακμή



Ανάπτυξη Ακμών



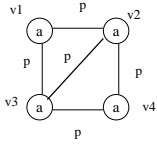
Όπου και να προστεθεί το ίδιο

v1, v2, v3, v4 τοπολογικά ισοδύναμα (4 δυνατές θέσεις)



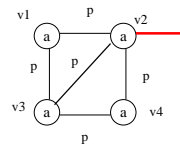
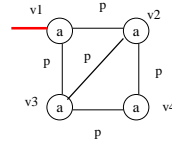
Ανάπτυξη Ακμών

Πρόσθεσε ακμή



Ανάπτυξη Ακμών

Πρόσθεσε ακμή

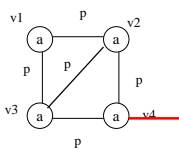
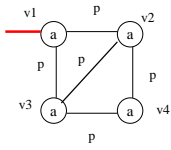


Δεν είναι ισοδύναμα



Ανάπτυξη Ακμών

Πρόσθεσε ακμή

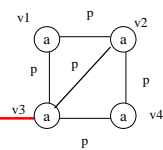
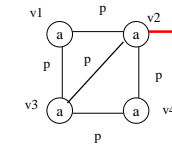


Είναι ισοδύναμα



Ανάπτυξη Ακμών

Πρόσθεσε ακμή

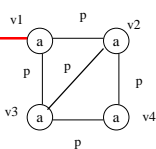
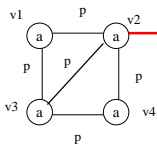


Είναι ισοδύναμα



Ανάπτυξη Ακμών

Πρόσθεσε ακμή

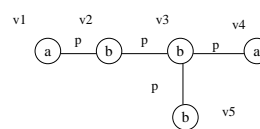


Άρα, δύο δυνατές θέσεις



Ανάπτυξη Ακμών

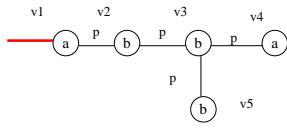
Πρόσθεσε ακμή





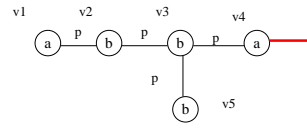
Ανάπτυξη Ακμών

Πρόσθεσε ακμή



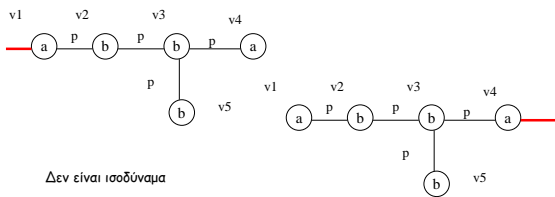
Ανάπτυξη Ακμών

Πρόσθεσε ακμή



Ανάπτυξη Ακμών

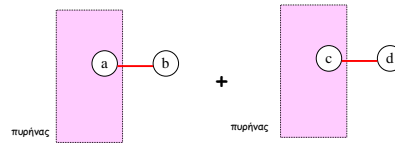
Πρόσθεσε ακμή



Δεν είναι ισοδύναμα



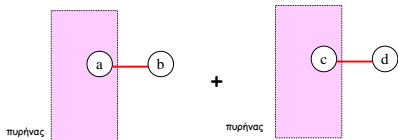
Ανάπτυξη Ακμών - Γενική Περίπτωση



Είναι τα a και c τοπολογικά ισοδύναμα:



Ανάπτυξη Ακμών



Συμβολισμός

a = c, αν τοπολογικά ισοδύναμα

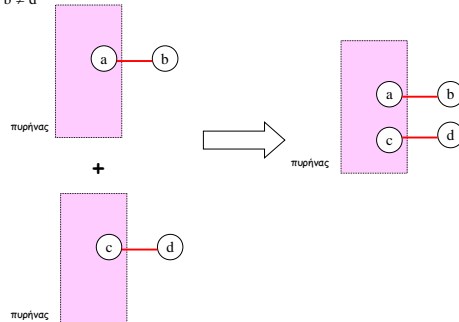
b = d αν ίδιες ετικέτες



Ανάπτυξη Ακμών

Περίπτωση

a ≠ c, b ≠ d



Δημιουργία Υποψηφίων

Περίπτωση
 $a = c, b \neq d$

Ανάπτυξη Ακμών

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 55

Δημιουργία Υποψηφίων

Περίπτωση
 $a \neq c, b = d$

Ανάπτυξη Ακμών

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 56

Δημιουργία Υποψηφίων

Περίπτωση
 $a = c, b = d$

Ανάπτυξη Ακμών

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 57

Δημιουργία Υποψηφίων

Ανάπτυξη Ακμών

Τέλος, μπορεί να έχουμε Πολλαπλούς Πυρήνες

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 58

Δημιουργία Υποψηφίων

Ανάπτυξη Κόμβων και Ακμών

Παρόλου που η ανάπτυξη ακμών παράγει πολλαπλούς υποψηφίους, γενικά τείνει να παράγει λιγότερα υπογραφήματα

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 59

Αλγόριθμος βασιμένος στον αριθμό

Βήμα 1:
Βρες τα συχνά 1-υπογραφήματα

Βήμα 2:
Επανάλαβε μέχρι να μην παράγονται νέα συχνά υπογραφήματα

- **Δημιουργία Υποψηφίων - Candidate Generation:**
 - Χρησιμοποίησε τα k-1 υπογραφήματα για τη δημιουργία k-υπογραφημάτων
- ▪ **Ψαλίδισμα Υποψηφίων - Candidate Pruning:**
 - Ψαλίδισε τα k-υπογραφήματα που περιέχουν μη συχνά (k-1)-γραφήματα
- **Υπολογισμός Υποστήριξης - Support Counting:**
 - Υπολόγισε την υποστήριξη των k-υπογραφημάτων που απομένουν
- **Υπολογισμός Υποψηφίων - Candidate Elimination:**
 - Διώξε τα υποψήφια k-υπογραφήματα που δεν είναι συχνά

Στην πραγματικότητα δεν είναι τόσο απλό, πολλά περισσότερα θέματα

Εξέλιξη Διδασκόντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 60



Ψαλίδισμα Υποψηφίων

Μετά τη δημιουργία των k -υπογραφημάτων, ψαλιδίζονται (pruned) τα υπογραφήματα που έχουν ένα $k-1$ υπογράφημα που δεν είναι συχνό

Πως:

Αφαίρεσε μια ακμή και έλεγξε αν το υπογράφημα που προκύπτει είναι συνδεδεμένο και συχνό

Αν όχι, ψαλίδισε το k -υπογράφημα



Ψαλίδισμα Υποψηφίων

Έλεγξε αν το υπογράφημα που προκύπτει είναι συχνό

Πως:

Ταίριαξε το με τα συχνά $k-1$ υπογραφήματα

Ισότητα μεταξύ γραφημάτων με βάση τοπολογική ισοδυναμία

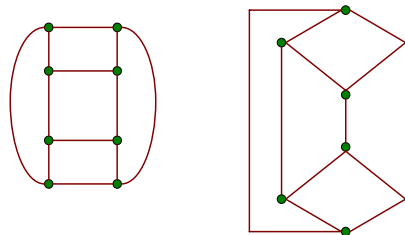


Ανάπτυξη Ακμών

Δύο γράφοι είναι **τοπολογικά ισοδύναμοι** ή **ισομορφικοί** αν υπάρχει μια 1-1 απεικόνιση μεταξύ των κόμβων τους



Παράδειγμα



Έλεγχος για ισομορφισμό μεταξύ γραφημάτων γίνεται:

1. **During candidate generation step**, to determine whether a candidate has been generated
2. **During candidate pruning step**, to check whether its ($k-1$)-subgraphs are frequent
3. **During candidate counting**, to check whether a candidate is contained within another graph



Κανονική Ετικέτα

Ο τυπικός τρόπος αντιμετώπισης του ισομορφισμού είναι η απεικόνιση ενός γραφήματος σε μια μοναδική αναπαράσταση με μια διατεταγμένη συμβολοσειρά που ονομάζεται **κώδικας (code)** ή **κανονική ετικέτα (canonical label)**

Με την ιδιότητα:

Αν δύο γραφήματα είναι ισομορφικά \Rightarrow ίδιοι κώδικες

Ισομορφισμός Γραφημάτων

Κανονική Ετικέτα

Κατασκευή του πίνακα γειτνιάσης

G1

$$M_{G1} = \begin{pmatrix} 0 & p & p & q \\ p & 0 & r & 0 \\ p & r & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix}$$

Ένα γράφημα μπορεί να έχει παραπάνω από μια αναπαράσταση βασισμένη σε πίνακα γειτνιάσης γιατί υπάρχουν διαφορετικοί τρόποι να διαταχθούν οι κόμβοι του (άρα και οι στήλες και οι γραμμές του πίνακα) - Άρα κατασκευή όλων

Εξόφλη Διδασκντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 67

Ισομορφισμός Γραφημάτων

Κανονική Ετικέτα

Οι διαφορετικές διατάξεις αντιστοιχούν στον πολλαπλασιασμό του πίνακα γειτνιάσης με έναν κατάλληλο πίνακα διάταξης

Πχ αλλαγή 1 γραμμής (στήλης) με 3 γραμμή (στήλη)

$$P13 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Μοναδιαίο + αλλαγή 1<->3 γραμμής

$M \times P13 =$ αλλαγή 1 και 3 **στήλης**

Εξόφλη Διδασκντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 68

Ισομορφισμός Γραφημάτων

Κανονική Ετικέτα

Πχ αλλαγή 1 γραμμής (στήλης) με 3 γραμμή (στήλη)

$$P^{T13} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Μοναδιαίο + αλλαγή 1<->3 στήλης

$P^{T13} \times M =$ αλλαγή 1 και 3 **γραμμής**

Εξόφλη Διδασκντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 69

Ισομορφισμός Γραφημάτων

Κανονική Ετικέτα

Στη συνέχεια κατασκευάζουμε την ετικέτα

Εξόφλη Διδασκντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 70

Ισομορφισμός Γραφημάτων

Κανονική Ετικέτα

Στη συνέχεια κατασκευάζουμε την ετικέτα

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

String: 0010001111010110 Canonical: 0111101011001000

Παραγωγή όλων και επιλογή του λεξικογραφικά μεγαλύτερης συμβολοσειράς
Επειδή ο πίνακας είναι συμμετρικός, μπορεί να κρατήσουμε μόνο το πάνω δεξιά τμήμα του

Εξόφλη Διδασκντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 71

Ισομορφισμός Γραφημάτων

Κανονική Ετικέτα

	A(1)	A(2)	A(3)	A(4)	B(5)	B(6)	B(7)	B(8)
A(1)	1	1	1	0	1	0	0	0
A(2)	1	1	0	1	0	1	0	0
A(3)	1	0	1	1	0	0	1	0
A(4)	0	1	1	1	0	0	0	1
B(5)	1	0	0	0	1	1	1	0
B(6)	0	1	0	0	1	1	0	1
B(7)	0	0	1	0	1	0	1	1
B(8)	0	0	0	1	0	1	1	1

	A(1)	A(2)	A(3)	A(4)	B(5)	B(6)	B(7)	B(8)
A(1)	1	1	0	1	0	1	0	0
A(2)	1	1	1	0	0	0	1	0
A(3)	0	1	1	1	1	0	0	0
A(4)	1	0	1	1	0	0	0	1
B(5)	0	0	1	0	1	0	1	1
B(6)	1	0	0	0	0	1	1	1
B(7)	0	1	0	0	1	1	1	0
B(8)	0	0	0	1	1	1	0	1

Εξόφλη Διδασκντων: Ακ. Έτος 2006-2007 ΓΡΑΦΗΜΑΤΑ ΚΑΙ ΠΑΓΚΟΣΜΙΟΣ ΙΕΤΟΣ 72



Πολύ ακριβή

Για κάθε $(k-1)$ υπογράφημα λίστα με τα ids των γραφημάτων στα οποία ανήκουν

Τομή των λιστών, κάθε φορά που δημιουργείται κάποιο k -υπογράφημα