

HDMS 2005, Αθήνα, Αύγουστος 25-26, 2005

Constrained

## Optimalities in Query Personalization

Γεωργία Κούτρικα, Γιάννης Ιωαννίδης  
Πανεπιστήμιο Αθηνών

### Constrained Query Personalization

1. Εισαγωγή στη CQP οικογένεια

2. Αρχιτεκτονική Συστήματος

3. Αναζήτηση στον Χώρο Καταστάσεων

4. Πειράματα

5. Σύνοψη

## 1 Η CQP οικογένεια

### Εξατομίκευση Ερώτησης

Ο δυναμικός εμπλουτισμός μίας **ερώτησης** με σχετικές προτιμήσεις προερχόμενες από ένα **προφίλ χρήστη** με στόχο την παροχή εξατομικευμένων απαντήσεων

## 1 Η CQP οικογένεια

### Η Εξατομίκευση Ερώτησης είναι

- Ένα πρόβλημα βελτιστοποίησης
- **μεγιστοποίηση του ενδιαφέροντος** χρήστη στην απάντηση με βάση το προφίλ του

## 1 Η CQP οικογένεια

Η Εξατομίκευση Ερώτησης είναι

Ένα πρόβλημα βελτιστοποίησης

- **μεγιστοποίηση του ενδιαφέροντος** χρήστη στην απάντηση με βάση το προφίλ του



⚠ “Υπερεξατομικευμένες” ερωτήσεις μπορεί να

- είναι ακριβές
- παράγουν κενές λίστες αποτελεσμάτων

## 1 Η CQP οικογένεια

Η Εξατομίκευση Ερώτησης είναι

Ένα πρόβλημα βελτιστοποίησης **με περιορισμούς**

- **μεγιστοποίηση του ενδιαφέροντος** χρήστη στην απάντηση

**με περιορισμούς στο κόστος εκτέλεσης και/ή στο μέγεθος της απάντησης**

- **ελαχιστοποίηση κόστους** εκτέλεσης ερώτησης **με περιορισμούς στο ενδιαφέρον και/ή στο μέγεθος της απάντησης**

## 1 Η CQP οικογένεια



## 1 Η CQP οικογένεια

### Παράδειγμα

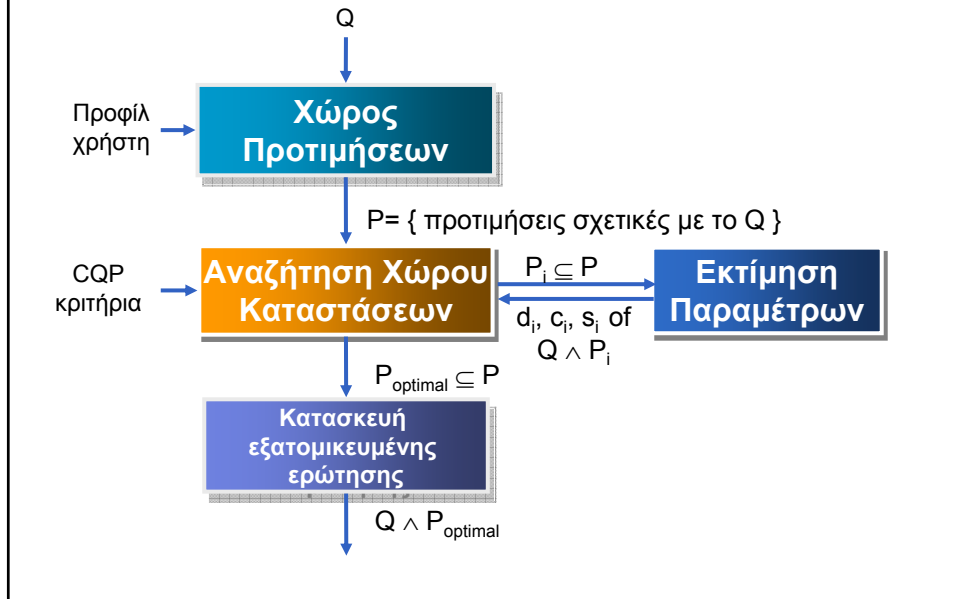


- Ακριβές ερωτήσεις
- Εκτενή αποτελέσματα



- Γρηγορότερη απόκριση
- Λιγότερα αποτελέσματα

## 2 CQP Αρχιτεκτονική Συστήματος



## 2 CQP Αρχιτεκτονική Συστήματος

### Προτιμήσεις

< q, d >

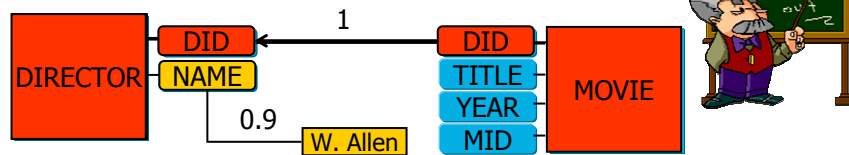
Ατομική συνθήκη q

Βαθμός ενδιαφέροντος  $d \in [0, 1]$



## 2 CQP Αρχιτεκτονική Συστήματος

### Προτιμήσεις



< DIRECTOR.name = 'W.Allen', 0.9 > Αποθηκευμένες  
< MOVIE.did = DIRECTOR.did, 1 > προτιμήσεις

.....  
< MOVIE.did = DIRECTOR.did and  
DIRECTOR.name = 'W.Allen', 0.9 \* 1 > Έμμεσες  
προτιμήσεις

## 2 CQP Αρχιτεκτονική Συστήματος

### Κατασκευή εξατομικευμένης ερώτησης

**Q:** SELECT title FROM MOVIE

**p<sub>1</sub>:** MOVIE.did=DIRECTOR.did and DIRECTOR.name='W. Allen'

**p<sub>2</sub>:** MOVIE.mid=MGENRE.mid and MGENRE.genre='comedy'

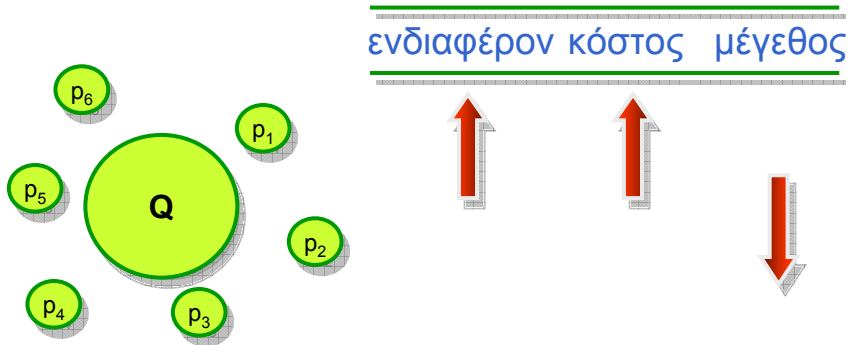
**Q<sub>1</sub>:** SELECT title FROM MOVIE M, DIRECTOR DIR  
WHERE M.did=DIR.did and DIR.name='W. Allen'

**Q<sub>2</sub>:** SELECT title FROM MOVIE M, MGENRE G  
WHERE M.mid=G.mid and G.genre = 'comedy'

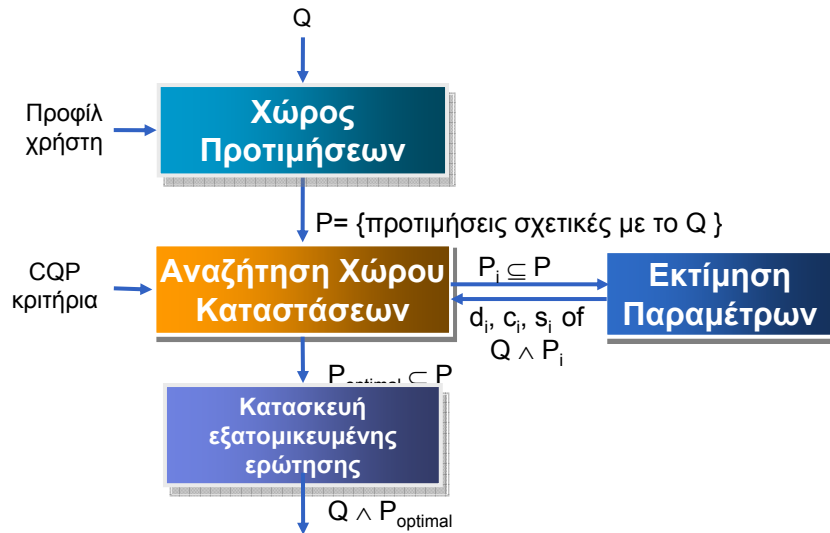
SELECT title FROM Q<sub>1</sub> union all Q<sub>2</sub>  
GROUP BY title HAVING count(\*)>1

## 2 CQP Αρχιτεκτονική Συστήματος

### Εκτίμηση Παραμέτρων



## 2 CQP Αρχιτεκτονική Συστήματος



### 3 CQP Χώρος Καταστάσεων

Ένα CQP πρόβλημα

Μεγιστοποίηση ενδιαφέροντος με κόστος  $\leq c_{upper}$

$Q, P \Rightarrow$  ο αριθμός των δυνατών υποσυνόλων του  $P$  είναι εκθετικός

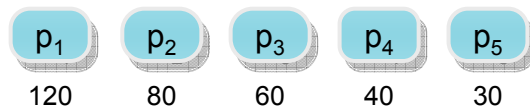
$P_x$  Κατάσταση,  $P_x \subseteq P$

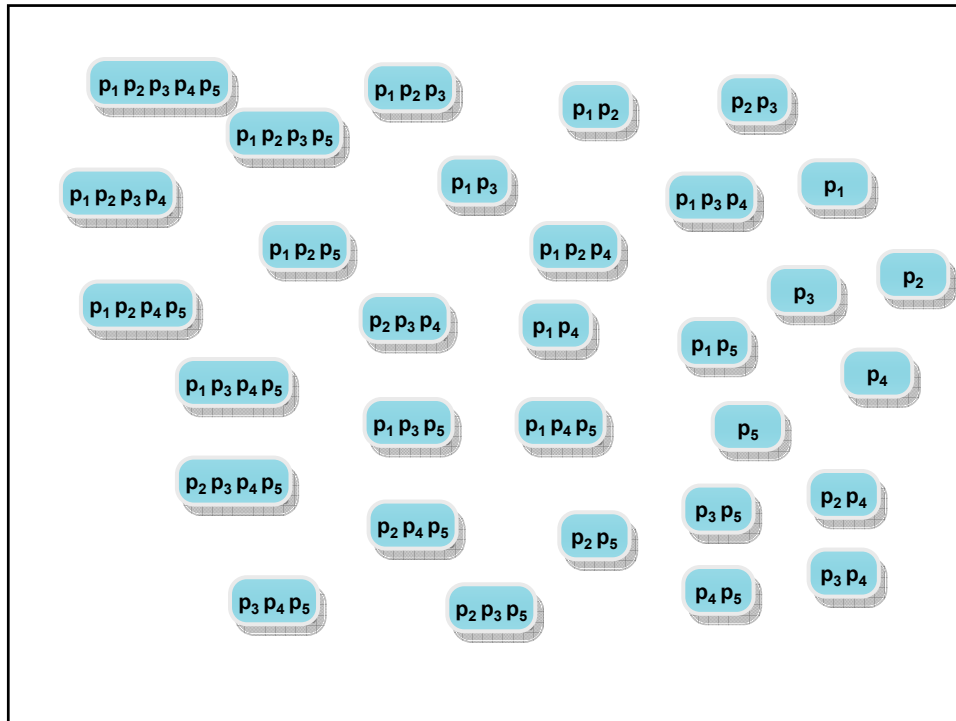
$P_x \rightarrow P_y$  Μετάβαση με βάση:  
το κόστος, το μέγεθος, ή το ενδιαφέρον

### 3 CQP Χώρος Καταστάσεων

Καταστάσεις

Έστω  $P$  διατεταγμένο κατά φθίνουσα σειρά κόστους:  $cost(p_i) \geq cost(p_{i+1})$





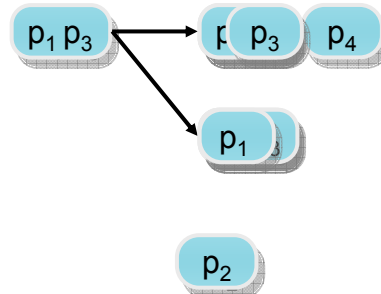
### 3 CQP Χώρος Καταστάσεων

Μεταβάσεις με βάση το κόστος

■ Horizontal( $p_1 p_3$ )



■ Vertical( $p_1 p_3$ )



### 3

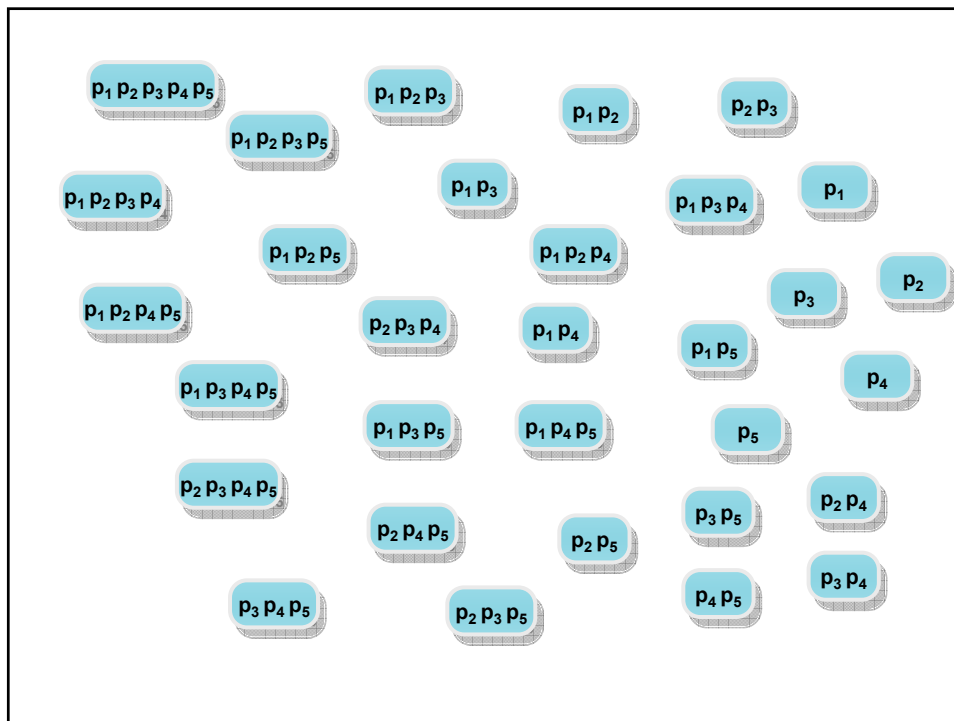
## CQP Χώρος Καταστάσεων

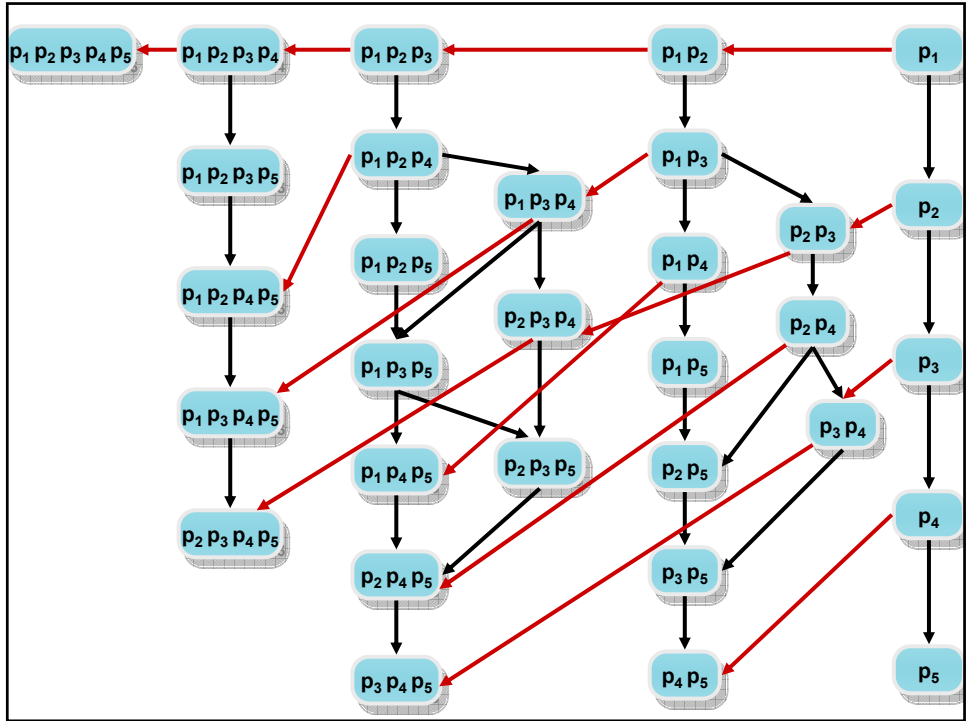
Μεταβάσεις με βάση το κόστος

- $\text{interest}(p_1 p_3 p_4) \geq \text{interest}(p_1 p_3)$
- $\text{cost}(p_1 p_3 p_4) \geq \text{cost}(p_1 p_3)$
- $\text{cost}(p_2 p_3) \leq \text{cost}(p_1 p_3)$



Μερική ταξινόμηση καταστάσεων με βάση συντακτικές αλλαγές





### 3 CQP Χώρος Καταστάσεων

#### Βασική Ιδέα Αναζήτησης



Περιοχή Α

Περιοχή Β



### 3

## CQP Χώρος Καταστάσεων

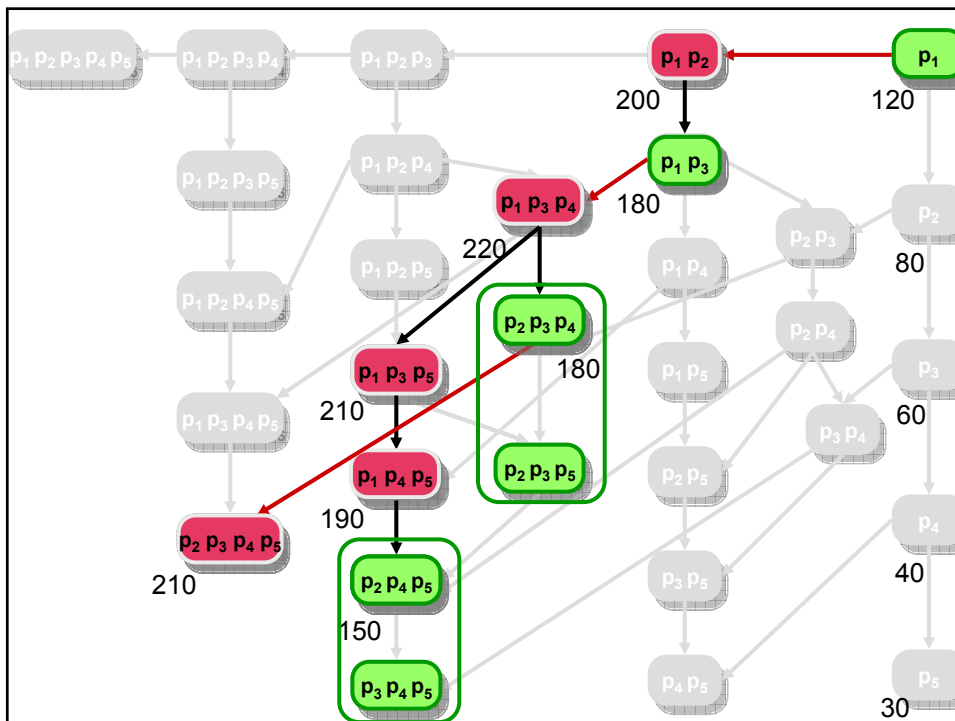
C\_Boundaries

Φάση 1: βρες boundaries

Φάση 2: βρες βέλτιστη λύση

π.χ.

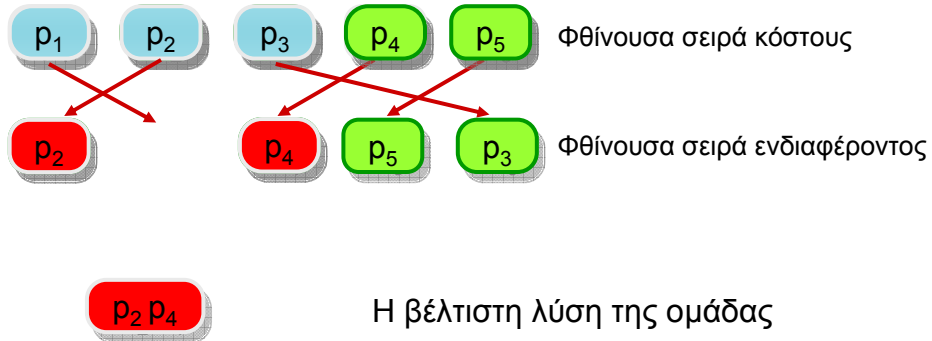
Μεγιστοποίηση του ενδιαφέροντος με κόστος < 185



### 3 CQP Χώρος Καταστάσεων

C\_Boundaries (φάση 2)

Αναζήτηση μέσα σε μία ομάδα για τη βέλτιστη λύση



### 3 CQP Χώρος Καταστάσεων

**Μεταβάσεις κόστους**

C\_Boundaries

C\_MaxBounds

**Μεταβάσεις ενδιαφέροντος**

D\_MaxDoi

D\_SingleMaxDoi

D\_HeurDoi

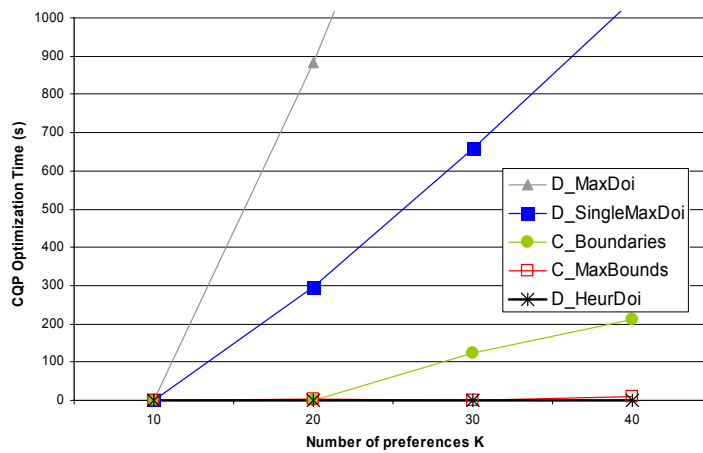
## 4 Πειράματα

### Execution Times

$K = \#$  προτιμήσεων στο  $P$

$C_{\mu\epsilon\gamma} = 400$

## 4 Πειράματα



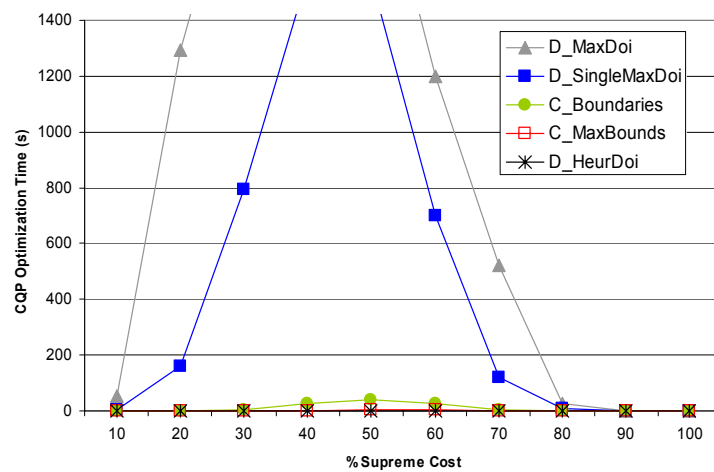
## 4 Πειράματα

### Execution Times

$C_{\mu\epsilon\gamma}$  = % του χρόνου εκτέλεσης της ερώτησης που ενσωματώνει όλες τις προτιμήσεις (Supreme Cost)

$K = 20$

## 4 Πειράματα



## 4 Πειράματα

### Quality of Solution

■  $doi_{\text{optimal}}$  by C\_Boundaries, D-MaxDoi

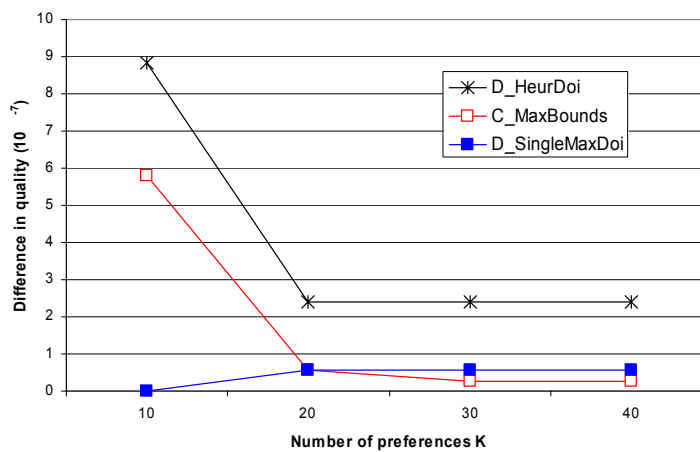
■ Quality =  $doi_{\text{optimal}} - doi_{\text{found}}$

■  $doi(P_x) = 1 - \prod_{p_i \in P_x} (1 - doi(p_i))$

K = # προτιμήσεων στο P

$C_{\text{μεγ}} = 400$

## 4 Πειράματα



## 5 Σύνοψη

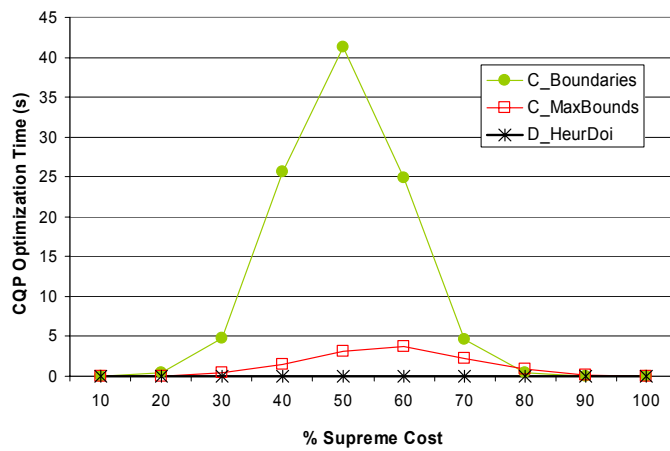
- Εισαγωγή της οικογένειας προβλημάτων εξατομίκευσης υπό περιορισμούς
- Διατύπωση του προβλήματος ως πρόβλημα αναζήτησης σε χώρο καταστάσεων
- «Συντακτικές» μεταβάσεις
- Πειράματα

## Μελλοντική Εργασία

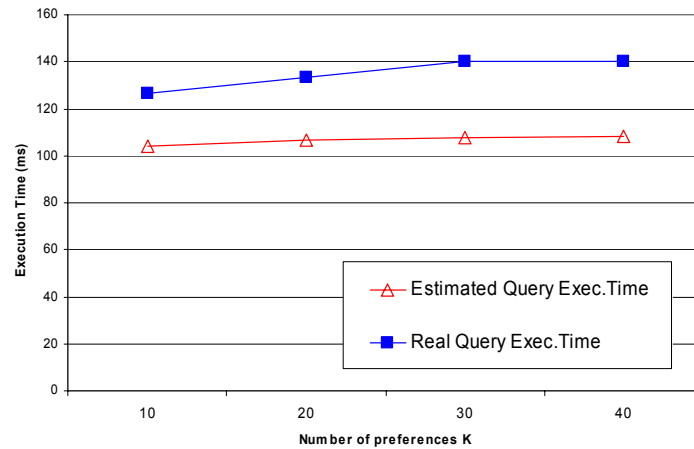
- Πολιτικές για CQP προβλήματα
- Ενσωμάτωση με location-based services
- Multi-objective constrained optimization

Ερωτήσεις ...?

## 4 Experiments



## 4 Experiments



## 2 CQP System Architecture

### Personalized Query Construction

Q: SELECT title FROM MOVIE

$p_1, p_2$  preferences

$q_1: Q \wedge p_1$

$q_2: Q \wedge p_2$

### Personalized Query

```
SELECT title FROM  $q_1$  union all  $q_2$ 
GROUP BY title HAVING count(*) > 1
```

## 2 CQP System Architecture

### Parameter Estimation

Subsets of preferences:  $P_x \subseteq P_y \subseteq P$

■  $\text{interest}(Q \wedge P_x) \leq \text{interest}(Q \wedge P_y)$

■  $\text{result\_size}(Q \wedge P_x) \geq \text{result\_size}(Q \wedge P_y)$

$Q_x := \cup q_i \Rightarrow \text{cost}(Q_x) = \text{cost}(\cup q_i) = \sum \text{cost}(q_i)$

■  $\text{cost}(Q \wedge P_x) \leq \text{cost}(Q \wedge P_y)$

