

# Advanced topics in Databases

Hellenic Mediterranean University

Prof. Demos Akoumianakis ([da@hmu.gr](mailto:da@hmu.gr))

# Plan for today

- Introductory material
  - Course overview and workload
  - Methods and tools
  - Learning objectives and expected outcomes

# Course organization

- Lectures and lab every Wednesday (5PK)
- Themes and focus
- Assessment

# Themes and focus

- Part A – Introduction
  - Data management
  - Reflections on the relational model of data
  - Challenges and the process of normalization
- Part B – Functional dependency theory
  - Functional dependencies
  - Properties of functional dependencies
  - Axioms of Armstrong
  - Algorithmic concepts
  - Relational decomposition

# Themes and focus (cont.)

- Part C – Complex data & Object-relational extensions
  - New data types, semi-structured / unstructured data, open data sets, big data
  - Implications on relational model
  - Querying complex data
- Part D – NoSQL
  - Key-value stores, Document-based systems, Native graphs
  - Relations and graphs

# Reference materials

- Open datasets
  - Mimic III data set (in progress)
    - CSV files: <https://physionet.org/content/mimiciii-demo/1.4/>
    - Relational Schema: <https://mit-lcp.github.io/mimic-schema-spy/relationships.html>
    - Tables: <https://mimic.mit.edu/docs/iii/tables/admissions/>
  - Kaggle data sets
    - Uber Requests
    - Airbnb
    - Netflix case (in progress)
- Advanced Postgres project
  - for students committed to relational technology

# Assessment

- No final exam!
- Students are assessed in class based on a small number of assignments, typically three or four
- Student assessment entails
  - Writing skills assessment
    - Typically, each assignment is coupled with a two-page extended abstract where each student summarizes the problem addressed, methods used and results
  - Presentation skills assessment
    - In class presentation of the work
  - Development/coding/implementation assessment

# Audience

- Join the course in eclass
  - <https://eclass.hmu.gr/courses/ECE158/>
- ERASMUS students to contact me directly
- Students in the Informatics Engineering program interested to undertake a lab mini project should join a dedicated group in the course eclass

# Data management - I

Basics from the introductory course on Databases

# Definition

- Data management is the practice of collecting, organizing, protecting, and storing an organization's data so it can be analyzed for business decisions

# Properties of data (examined so far)

- In the past we examined data with certain properties
  - Specific types (VARCHAR, TEXT, INT, REAL, DATA, etc.)
  - Strict constraints (well defined domains, size, min/max τιμές, etc.)
  - Limited representational capacity
  - Data belonging to an ‘owner’, either a person or an institution
  - Processed using a specific computational methods / tools e.g. relational algebra operators such as  $\cap$ ,  $\cup$ ,  $\times$

# Data models

- Τι εξετάσαμε μέχρι πρόσφατα
  - ✓ *Δεδομένα που είναι*
    - ✓ *σχετικά απλά (π.χ. VARCHAR, TEXT, INT, REAL, DATA, etc.)*
    - ✓ *υπακούουν σε αυστηρούς περιορισμούς ακεραιότητας*
      - ✓ *Καλά ορισμένο πεδίο ορισμού, μέγεθος, όρια, min/max τιμές, κλπ.*
    - ✓ *διαθέτουν περιορισμένη εκφραστική δύναμη*
    - ✓ *έχουν 'ιδιοκτήτη' φυσικό πρόσωπο / φορέα*
      - ✓ *δηλαδή ανήκουν σε κάποιον*
    - ✓ *'υποτάσσονται' σε συγκεκριμένο μοντέλο υπολογισμού*
      - ✓ *Με ακεραίους κάνω προσθέσεις, αφαιρέσεις, κοκ.*
      - ✓ *Με πίνακες κάνω  $\cap$ ,  $\cup$ ,  $\times$ ,  $-$ , κλπ.*
- Μοντέλα δεδομένων
  - με εκφραστική δύναμη που καλύπτει το εύρος των προβλημάτων που προκύπτουν από δεδομένα με τις παραπάνω ιδιότητες

# Changing nature of data

- Τα τελευταία χρόνια όλο και περισσότερο αλλάζει η ‘φύση’ των δεδομένων
  - Δεδομένα που
    - είναι σύνθετων τύπων, μεγάλου όγκου και αυξημένης πολυπλοκότητας
      - Τι τύπος δεδομένου είναι ένα *email*;
      - Τι τύπος δεδομένου είναι ένα η *τροχιά* ενός κινούμενου οχήματος;
    - εγείρουν ζητήματα ακεραιότητας
      - Είναι έγκυρο το *video* που αποτυπώνει την εμπορική αμαξοστοιχία στα Τέμπη;
    - απαιτούν εξειδίκευση για να είναι χρήσιμα
      - Πως αναλύω δεδομένα τύπου *email*, *τροχιές* και *video*;
    - είναι διασυνδεδεμένα παρότι μπορεί να διαφέρουν ως προς τον ‘ιδιοκτήτη’
      - Πως ενσωματώνω αλληλεξαρτώμενα δεδομένα διαφορετικού τύπου και προέλευσης;
    - ‘μεταβάλλουν’ το μοντέλο υπολογισμού
      - Γιατί τα παραδοσιακά μοντέλα δεν αρκούν και ποια νέα εμφανίζονται;
  - Νέα μοντέλα δεδομένων που
    - υιοθετούν αρχές & πρακτικές ‘ανοικτών’ δεδομένων (open data)
    - αξιοποιούν την ‘δομημένη’ φύση δεδομένων (‘structured’ data)
    - διαθέτουν αυξημένη εκφραστική δύναμη

# Emerging data models

- Τα τελευταία χρόνια όλο και περισσότερο αλλάζει η 'φύση' των δεδομένων
  - Δεδομένα που
    - είναι σύνθετων τύπων, μεγάλου όγκου και αυξημένης πολυπλοκότητας
      - Τι τύπος δεδομένου είναι ένα email;
      - Τι τύπος δεδομένου είναι ένα η τροχιά ενός κινούμενου οχήματος;
    - εγείρουν ζητήματα ακεραιότητας
      - Είναι έγκυρο το video που αποτυπώνει την εμπορική αμαξοστοιχία στα Τέμπη;
    - απαιτούν εξειδίκευση για να είναι χρήσιμα
      - Πως αναλύω δεδομένα τύπου email, τροχιάς και video;
    - είναι διασυνδεδεμένα παρότι μπορεί να διαφέρουν ως προς τον 'ιδιοκτήτη'
      - Πως ενσωματώνω αλληλεξαρτώμενα δεδομένα διαφορετικού τύπου και προέλευσης;
    - 'μεταβάλλουν' το μοντέλο υπολογισμού
  - Μοντέλα δεδομένων που
    - αξιοποιούν 'structured' data
    - υιοθετούν αρχές & πρακτικές 'ανοικτών' δεδομένων
    - διαθέτουν αυξημένη εκφραστική δύναμη

# Ποικιλία δεδομένων και πηγών

- Δεδομένα που συλλέγονται και αποθηκεύονται
  - Ηλεκτρονικές συναλλαγές (e-commerce)
  - Οικονομικές πράξεις (financial transactions, bank/credit transactions)
  - Ηλεκτρονικό εμπόριο, αγορές και συνήθειες καταναλωτών
  - Δεδομένα κοινωνικής δικτύωσης (Social Networking)



- Τι ζητήματα προκύπτουν;

# Ένα ενδεικτικό παράδειγμα

Πανδημία COVID 19

# Συζήτηση στην τάξη

- Αναλογιστείτε την περίοδο της πανδημίας
  - Ασθενείς που νοσούν ή νόσησαν από COVID-19
  - Ιχνηλάτιση επαφών τους

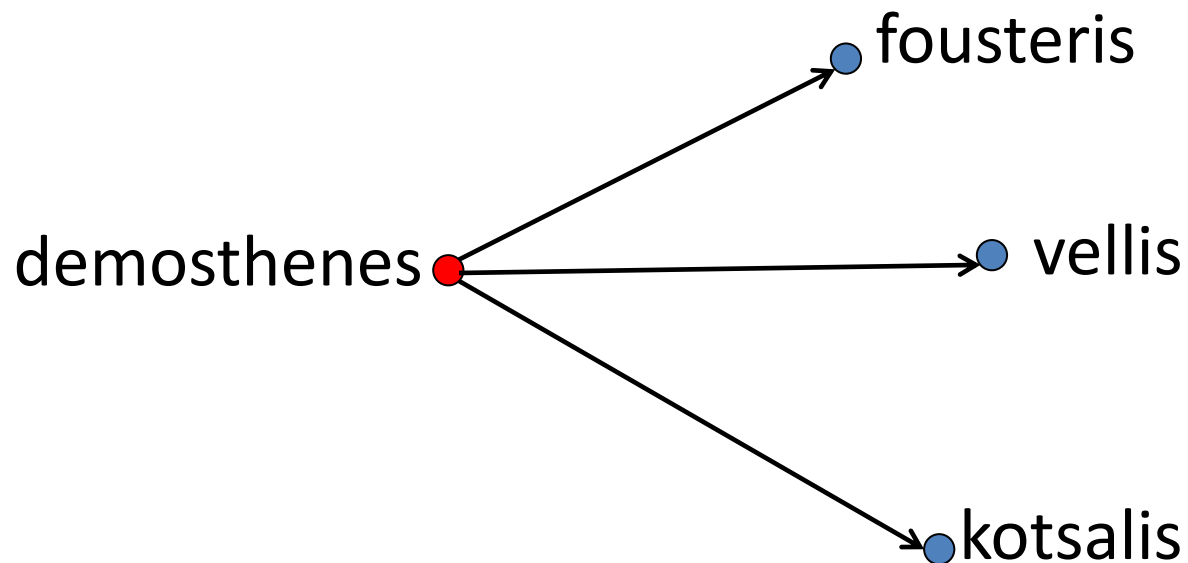
# Παράδειγμα

- Έστω ότι *demosthenes* είναι το πρώτο επιβεβαιωμένο κρούσμα

demosthenes●

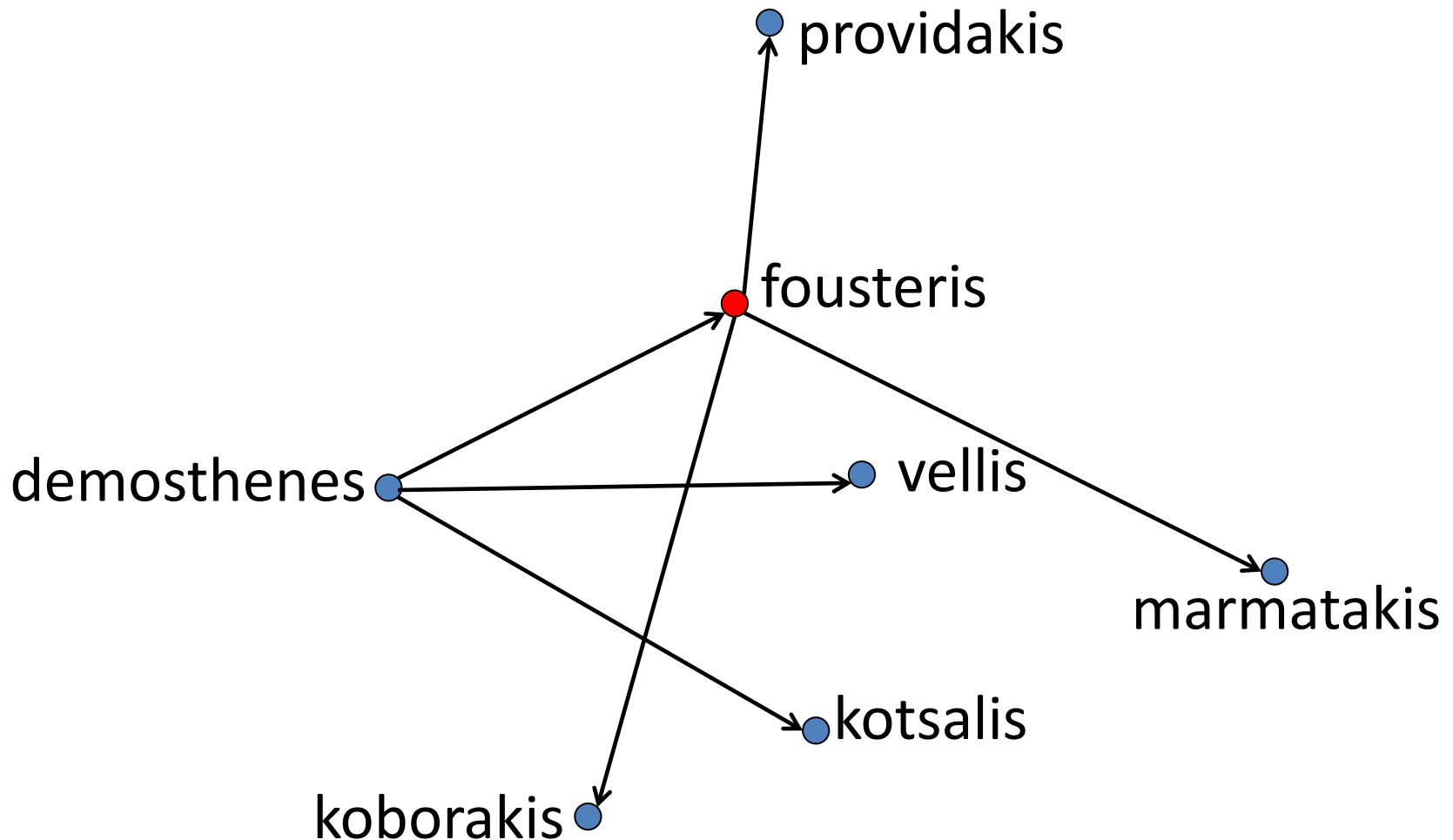
# Παράδειγμα (συν.)

- Οι επαφές του *demosthenes*



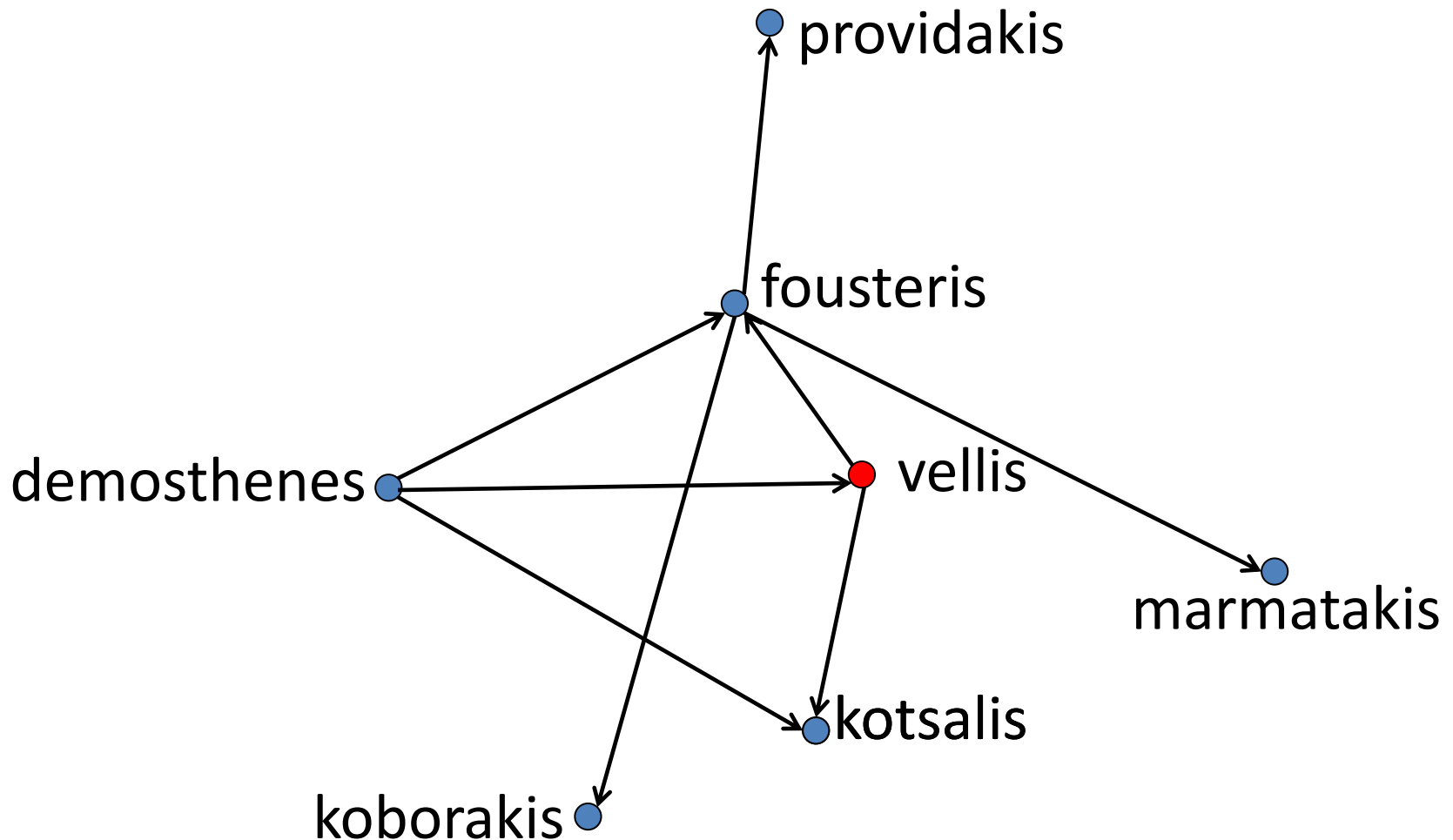
# Παράδειγμα (συν.)

- Οι επαφές του *fousteris*



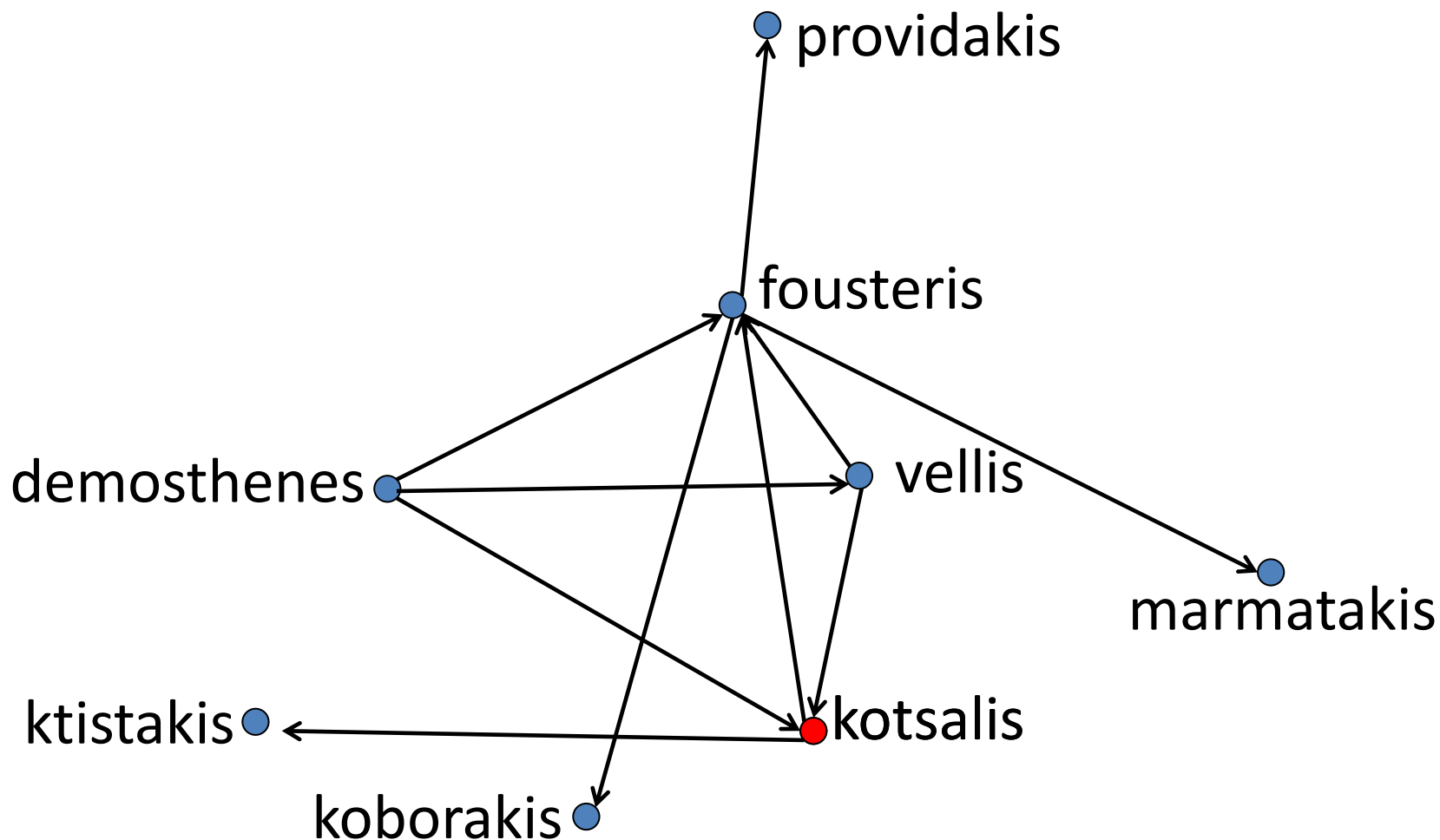
# Παράδειγμα (συν.)

- Οι επαφές του *vellis*



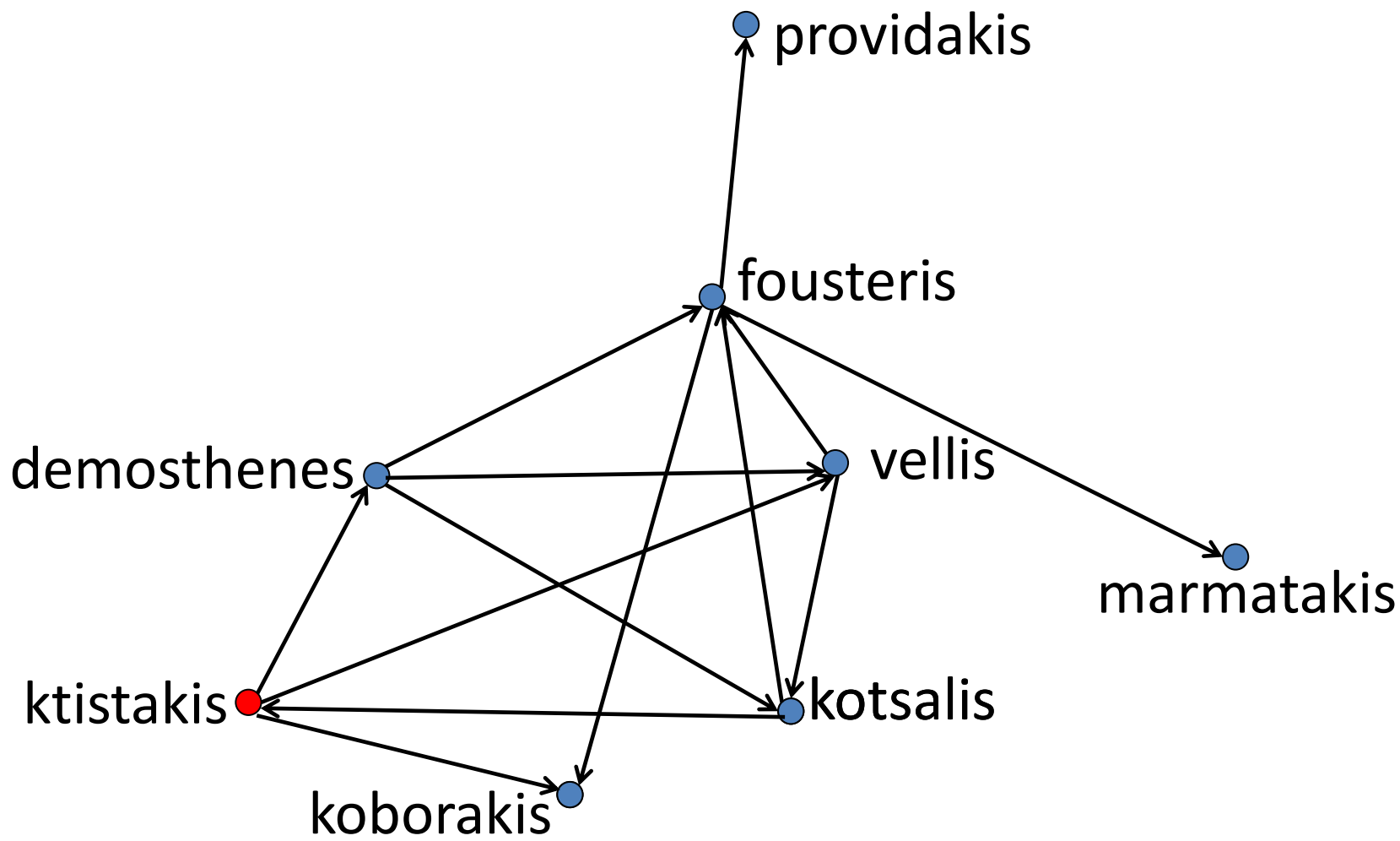
# Παράδειγμα (συν.)

- Οι επαφές του *kotsalis*



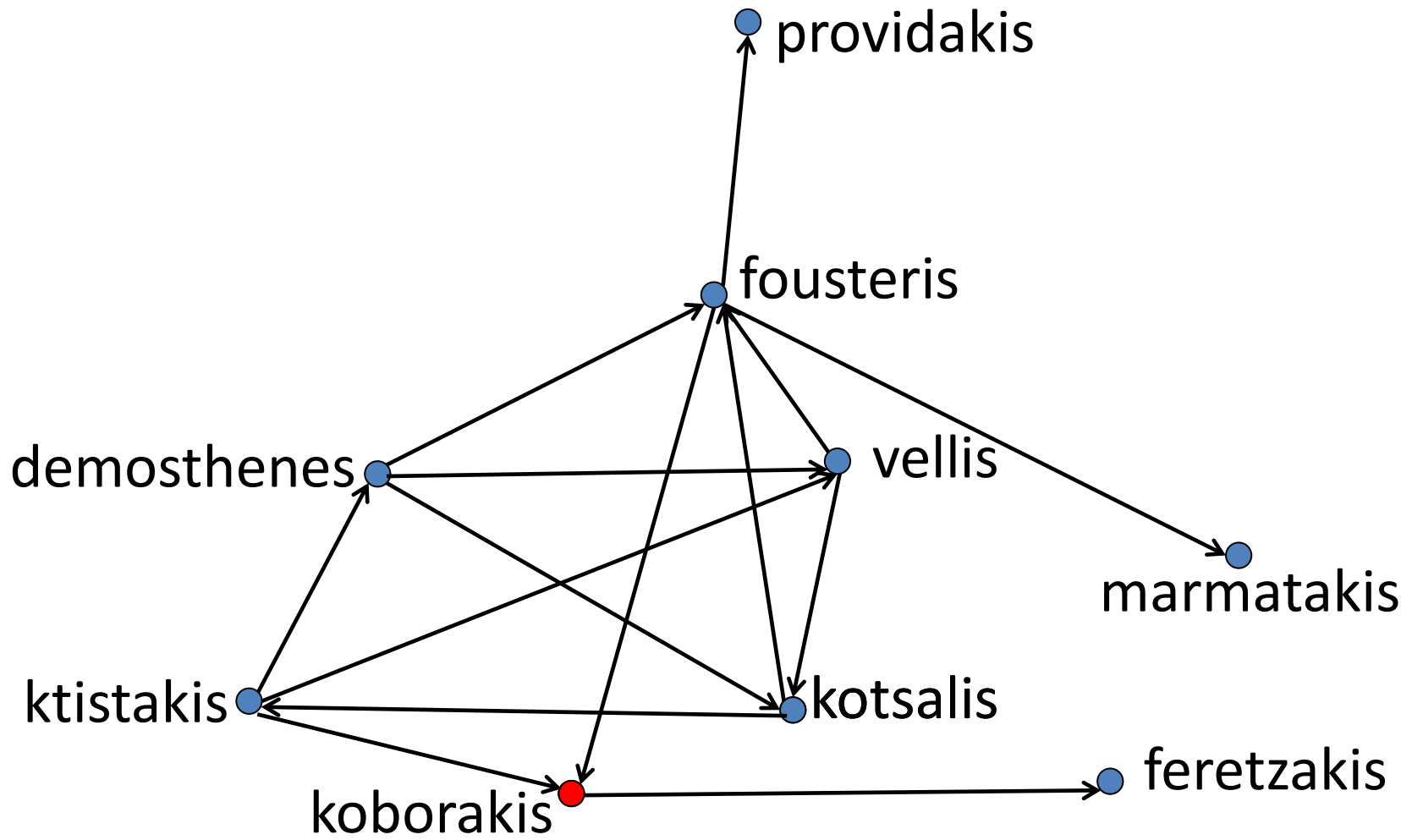
# Παράδειγμα (συν.)

- Οι επαφές του *ktistakis*



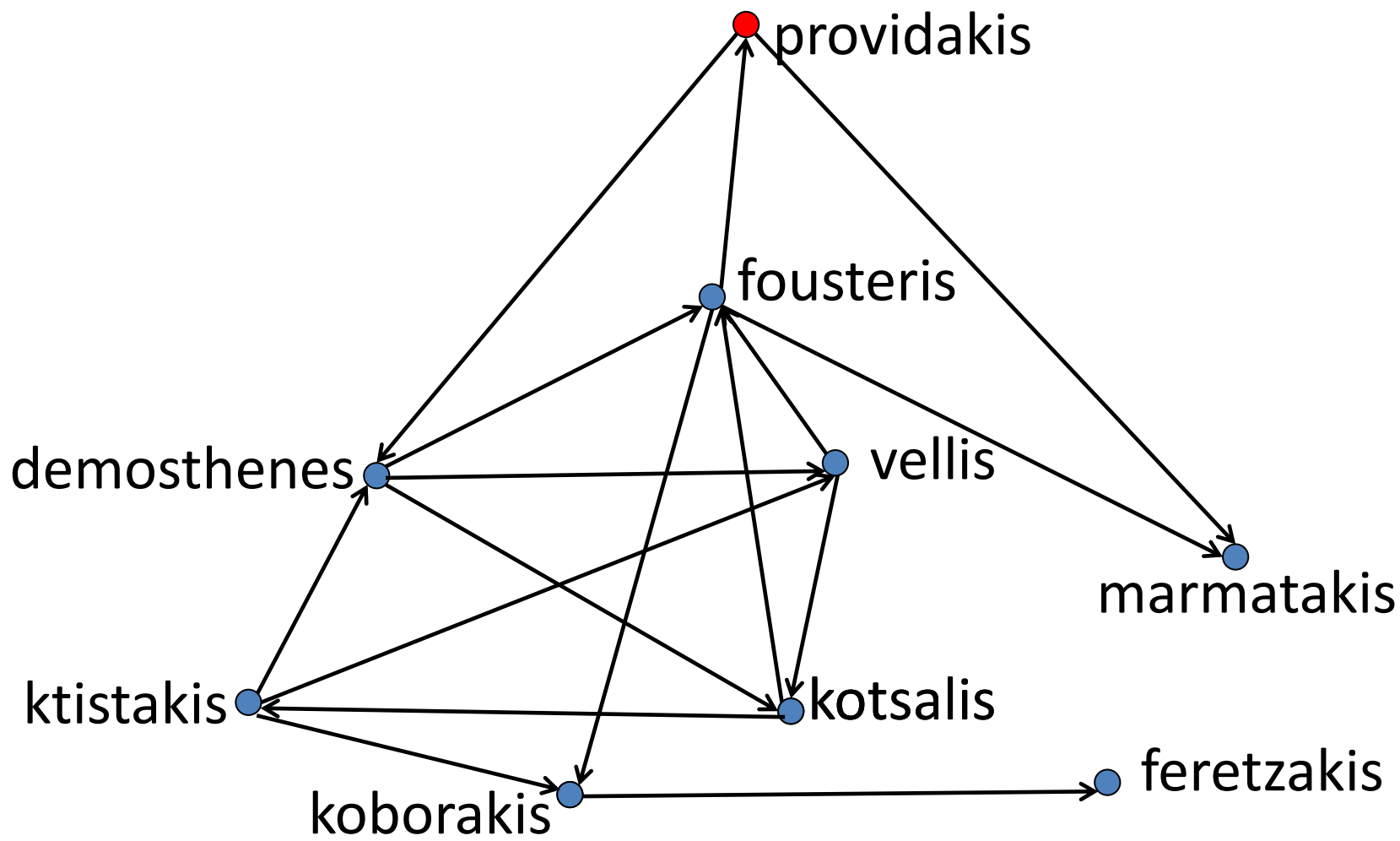
# Παράδειγμα (συν.)

- Οι επαφές του *koborakis*



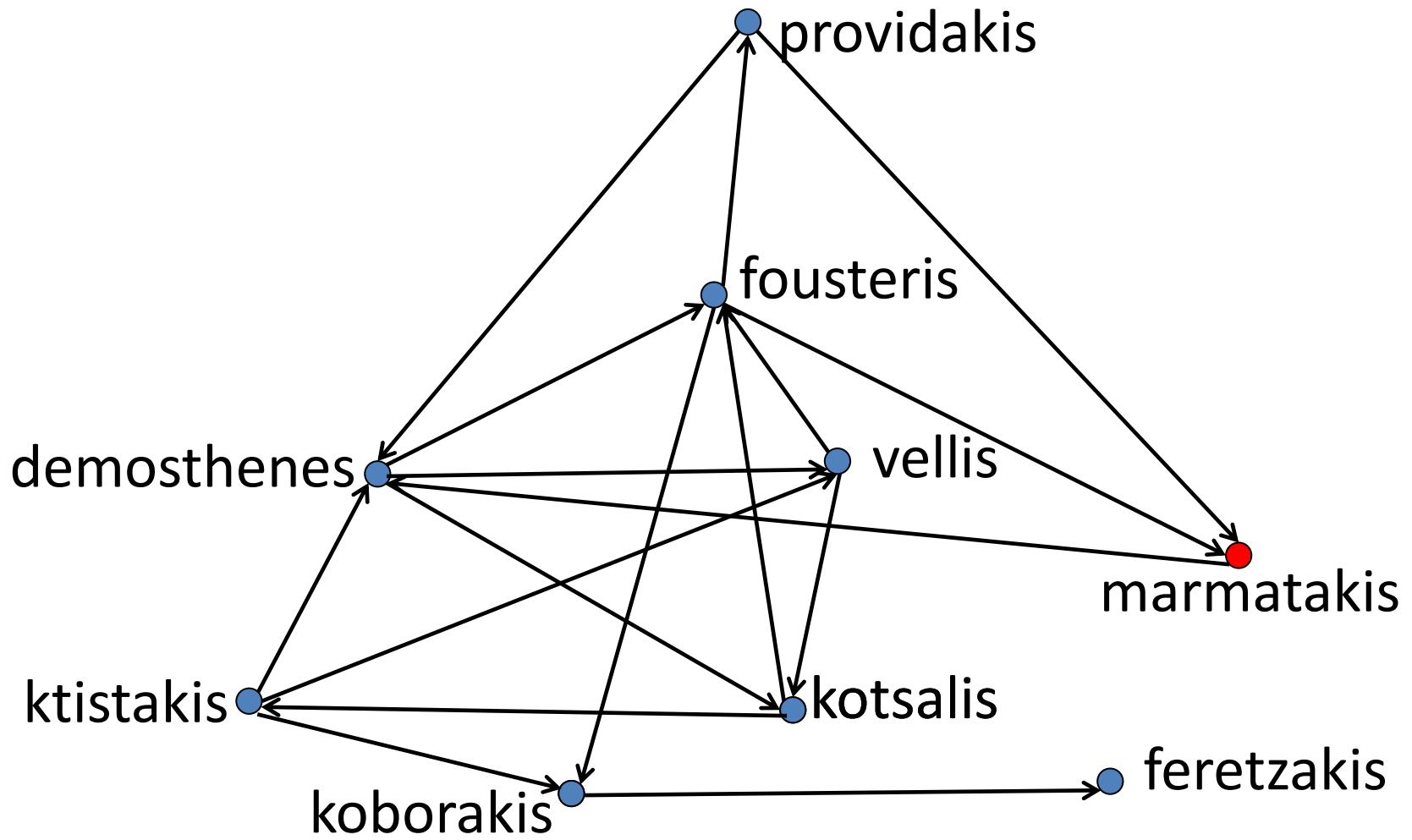
# Παράδειγμα (συν.)

- Οι επαφές του *providakis*



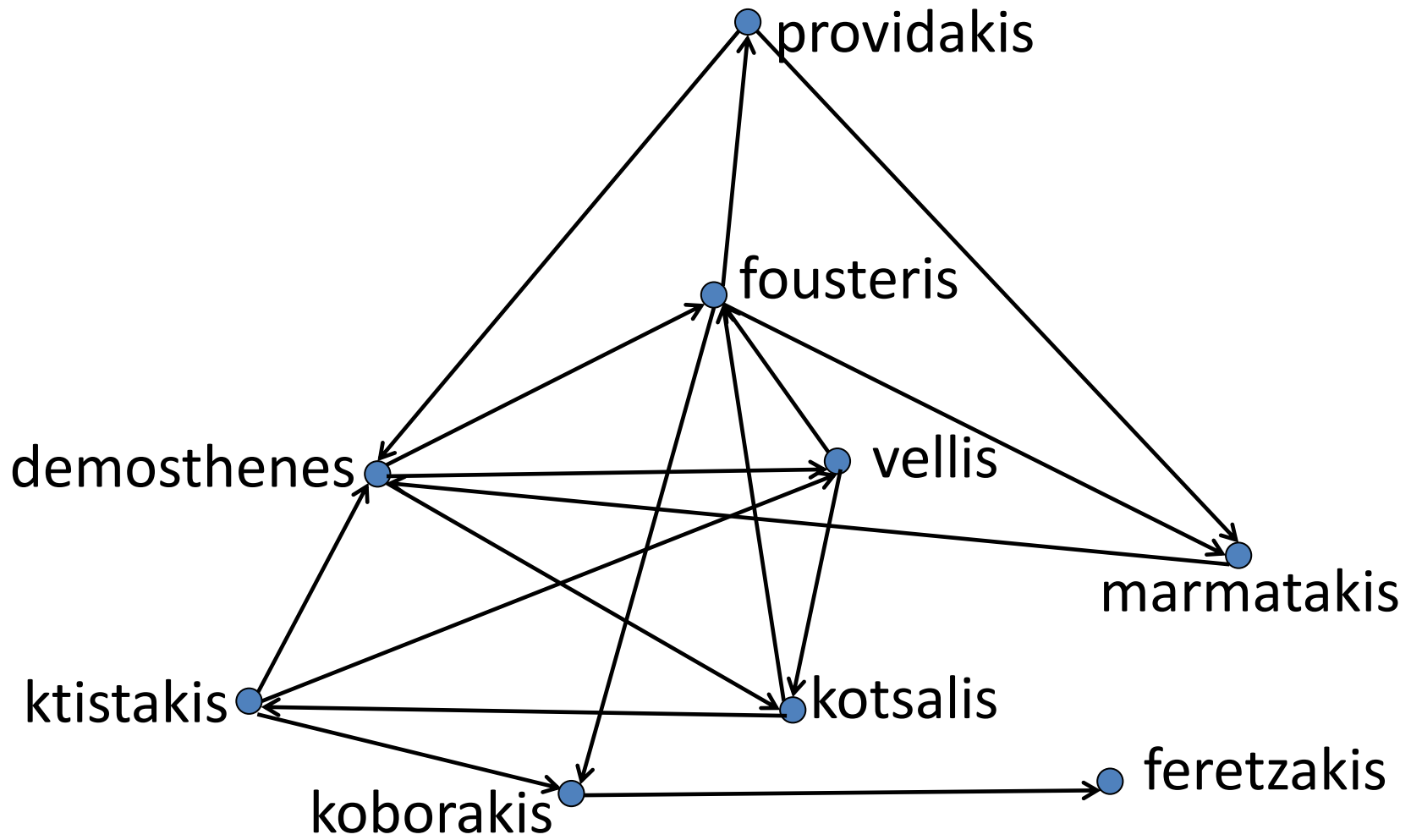
# Παράδειγμα (συν.)

- Οι επαφές του *marmatakis*



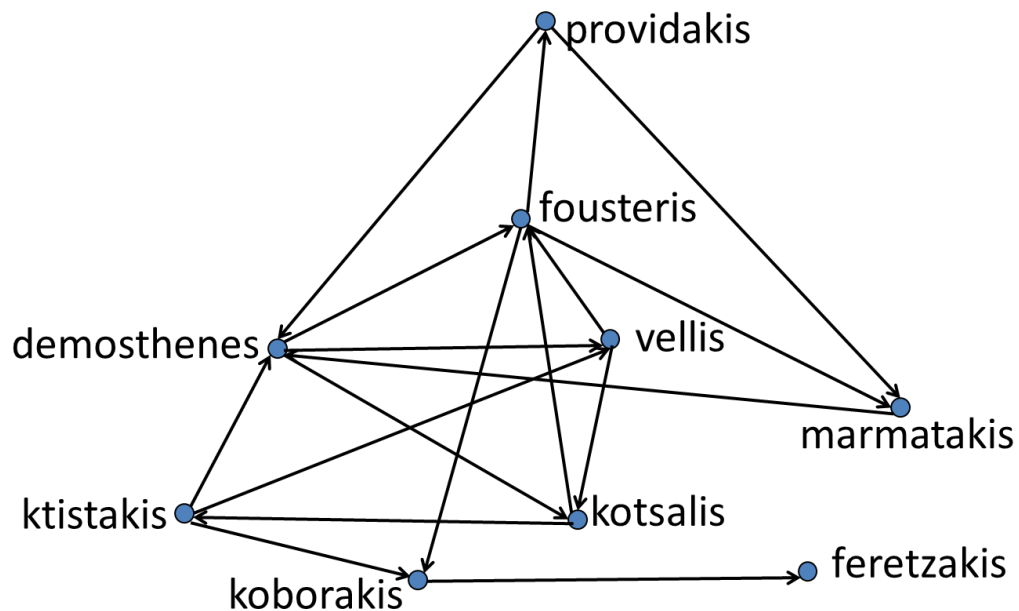
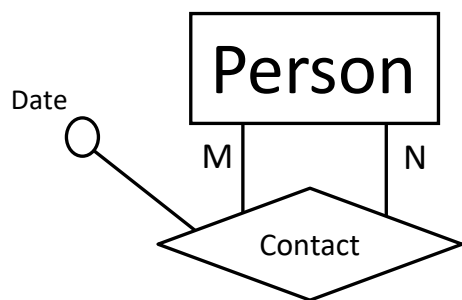
# Ζήτημα προς διερεύνηση

- Πως αναπαριστώ μια τέτοια δομή;



# Αναπαράσταση (συν.)

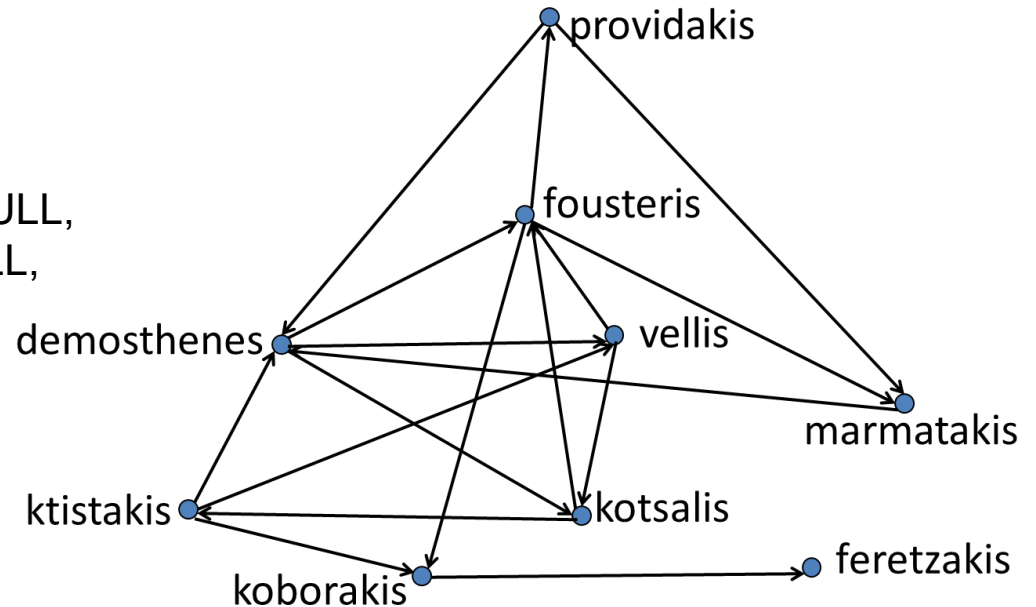
- Σε ΔΟΣ



# Αναπαράσταση (συν.)

- Σχεσιακό σχήμα

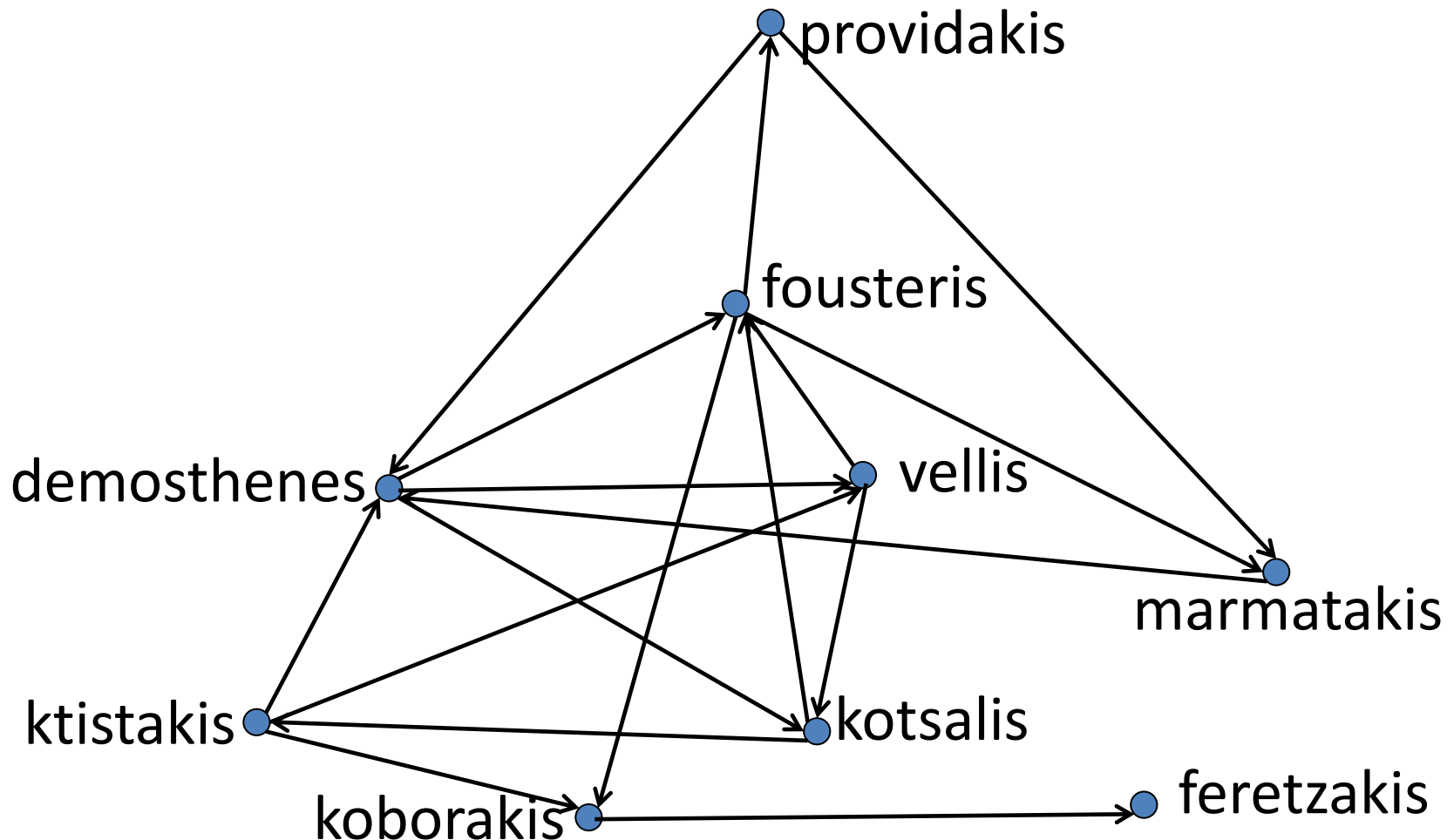
```
CREATE TABLE PERSON (  
  KODIKOS VARCHAR(30) NOT NULL,  
  ONOMA VARCHAR(60) NOT NULL,  
  EPONYMO VARCHAR(120),  
  PRIMARY KEY(KODIKOS));
```



```
CREATE TABLE CONTACT(  
  KODIKOS1 VARCHAR(30) NOT NULL,  
  KODIKOS2 VARCHAR(30) NOT NULL, CONTACTDATE DATE NOT NULL,  
  FOREIGN KEY (KODIKOS1) REFERENCES PERSON (KODIKOS),  
  FOREIGN KEY (KODIKOS2) REFERENCES PERSON (KODIKOS),  
  PRIMARY KEY (KODIKOS1, KODIKOS2));
```

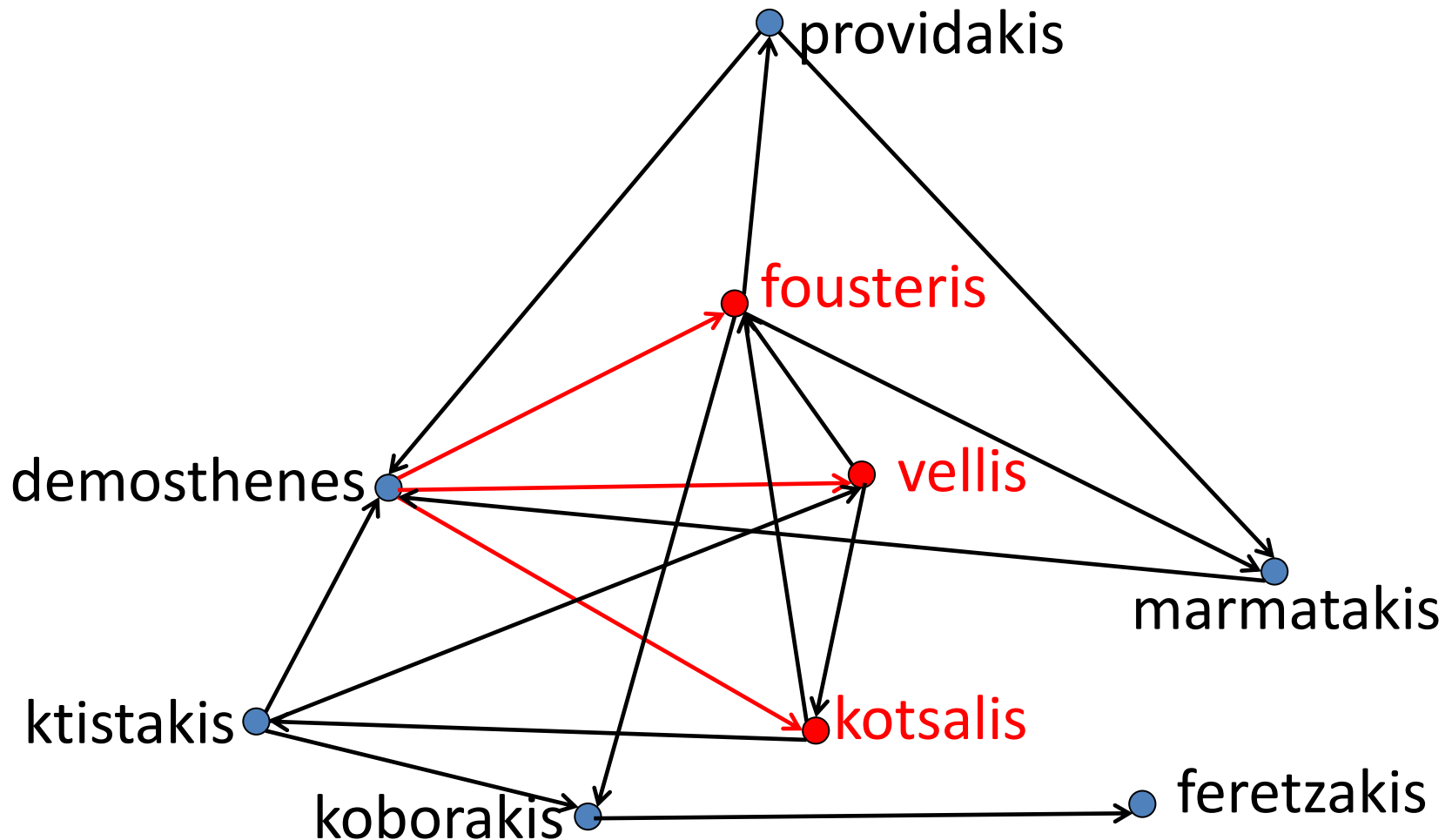
# Πως θα απαντούσατε το ερώτημα;

- Με ποιους ήρθε σε επαφή ο demosthenes;



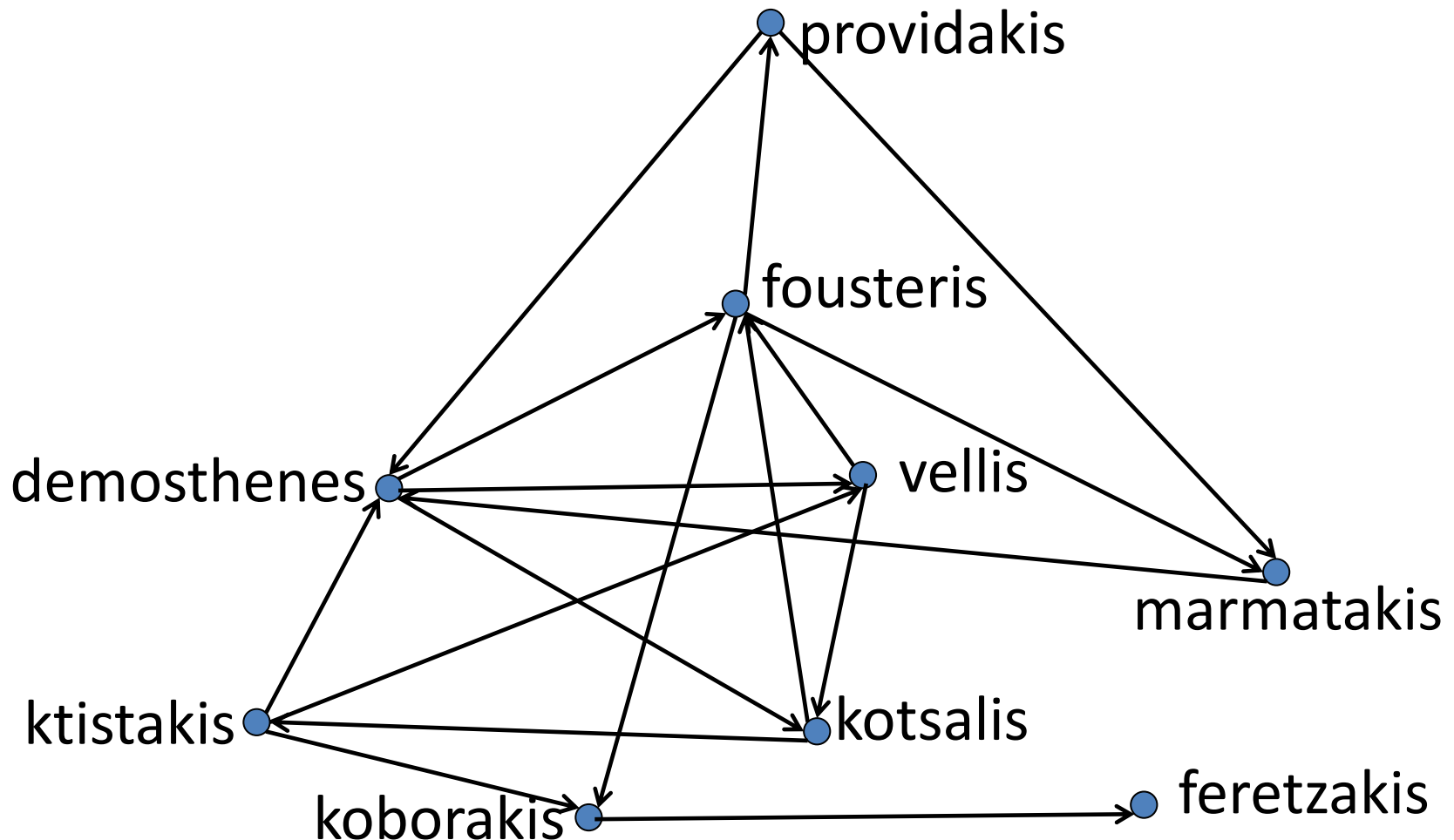
# Απάντηση

- Με ποιους ήρθε σε επαφή ο demosthenes;



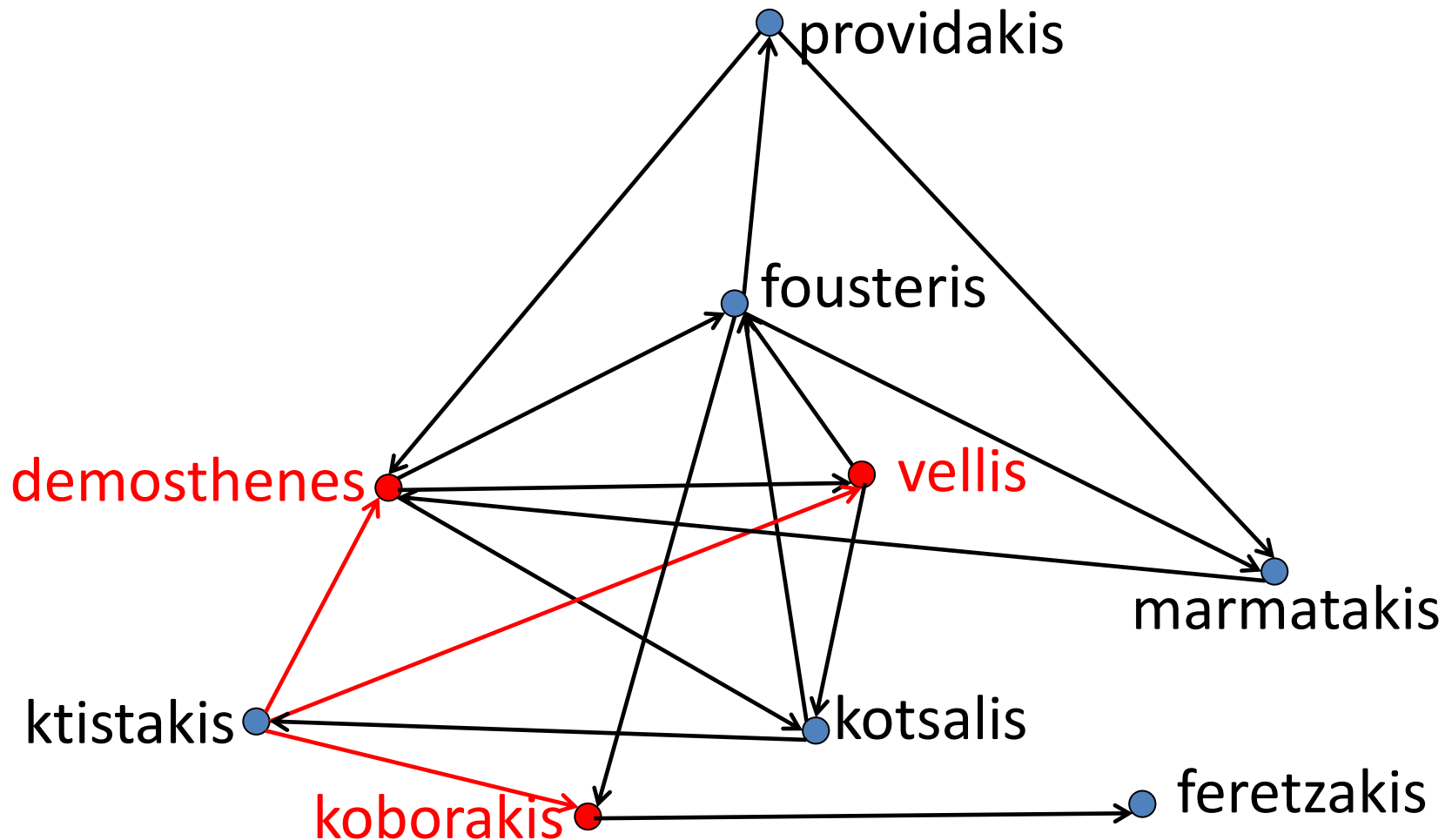
# Πως θα απαντούσατε το ερώτημα;

- Με ποιους ήρθε σε επαφή ο ktistakis;



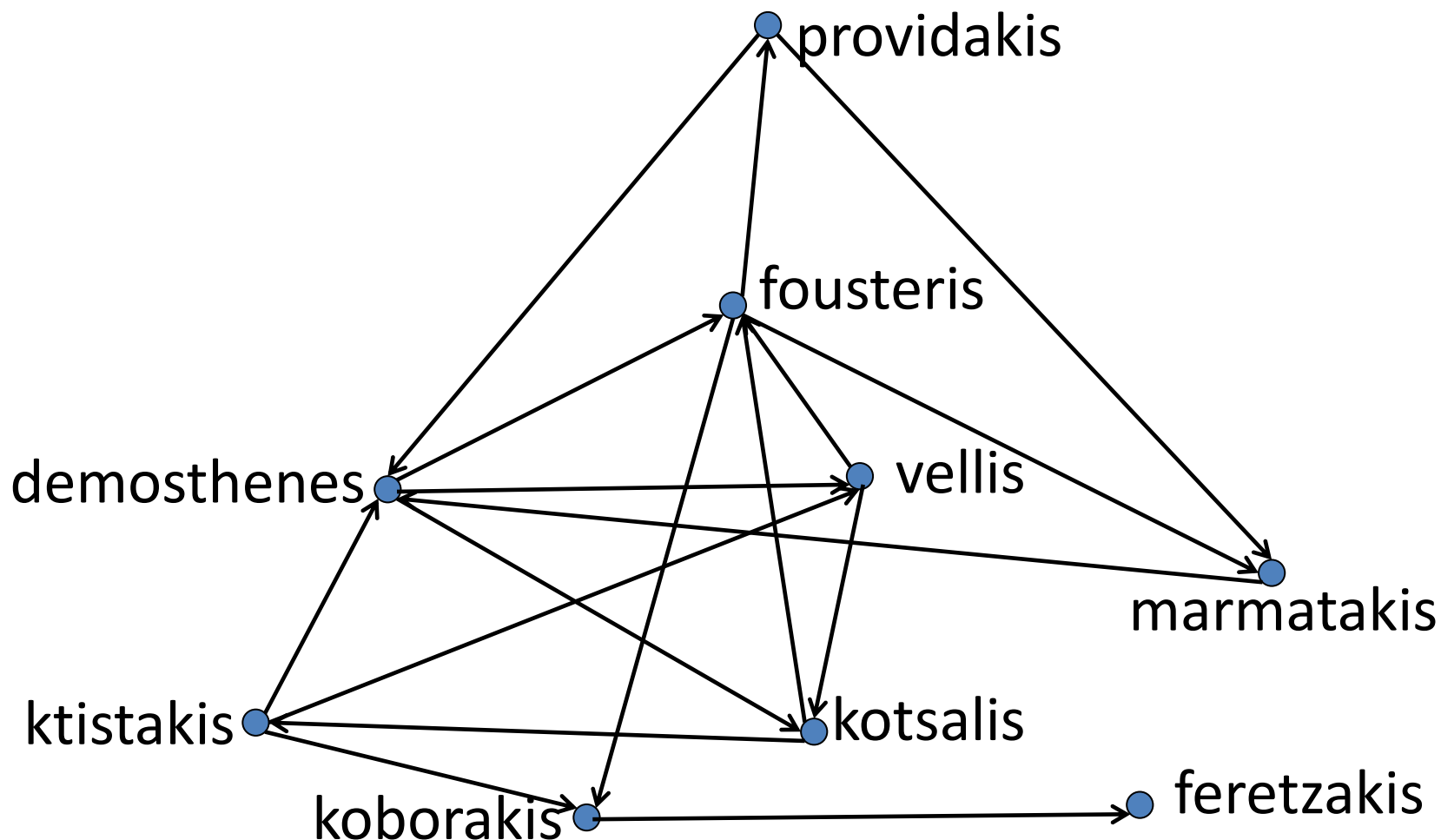
# Απάντηση

- Με ποιους ήρθε σε επαφή ο ktistakis;



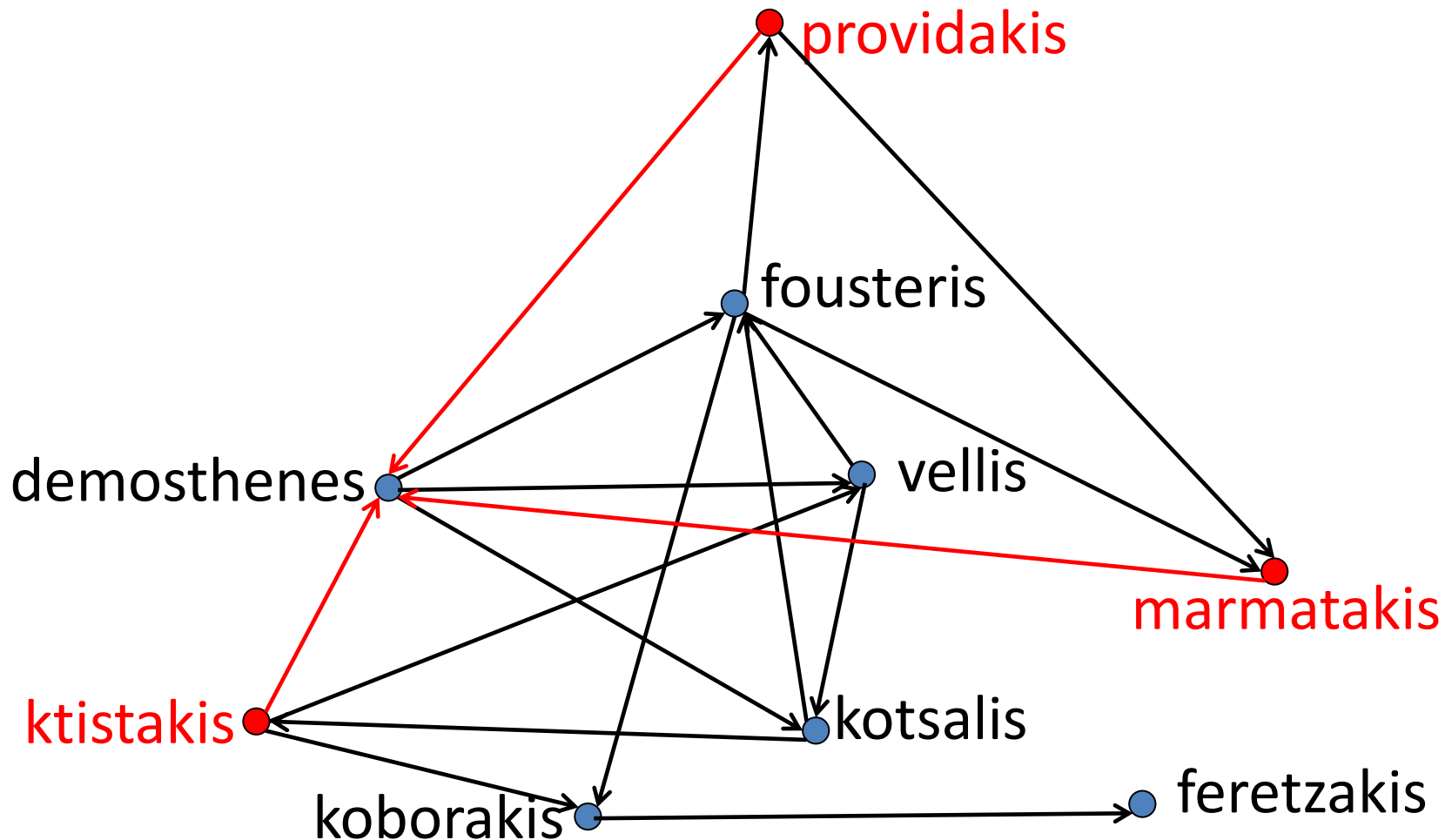
# Πως θα απαντούσατε το ερώτημα;

- Ποιοι ήρθαν σε επαφή με τον demosthenes;



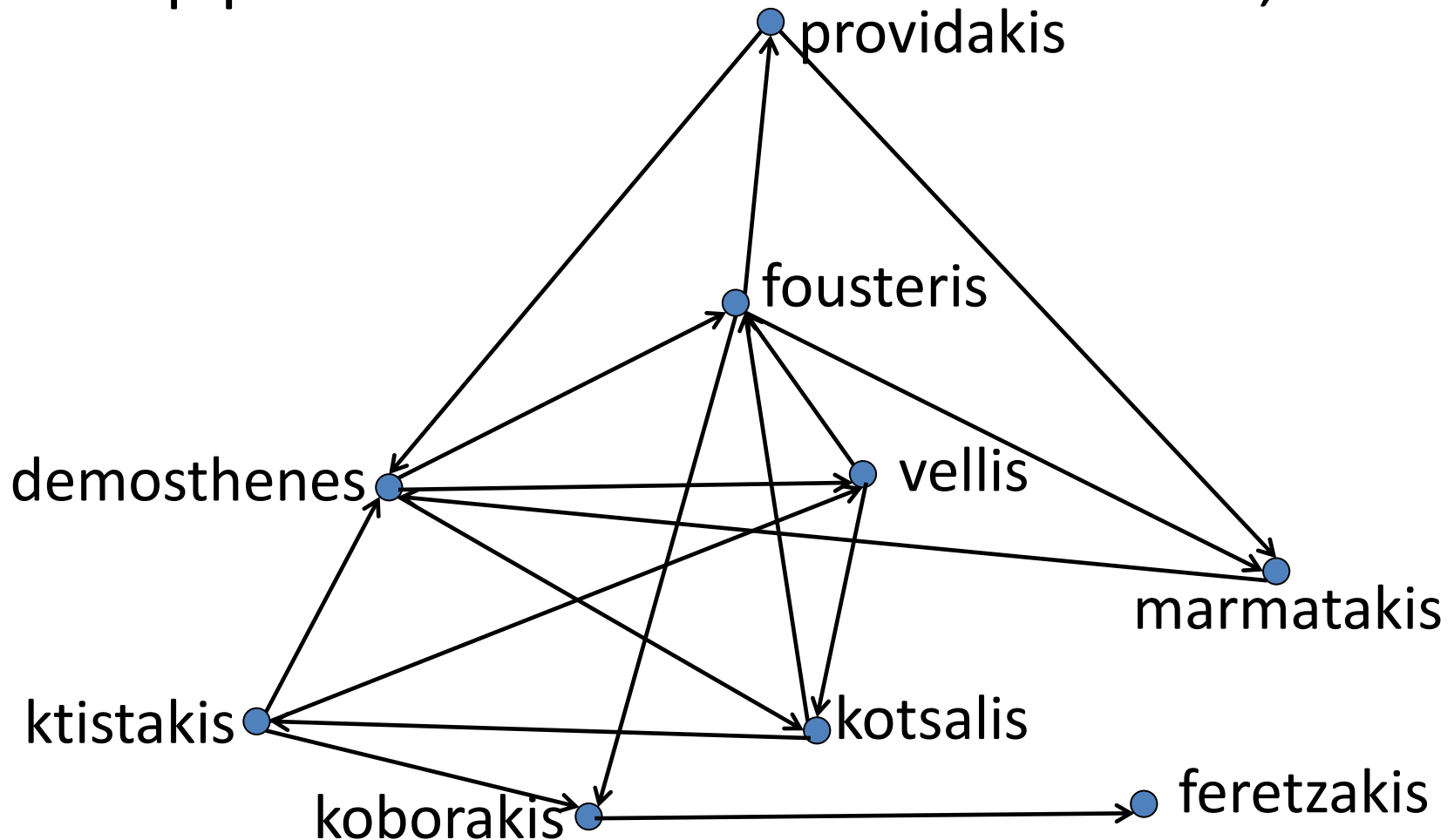
# Απάντηση

- Ποιοι ήρθαν σε επαφή με τον demosthenes;



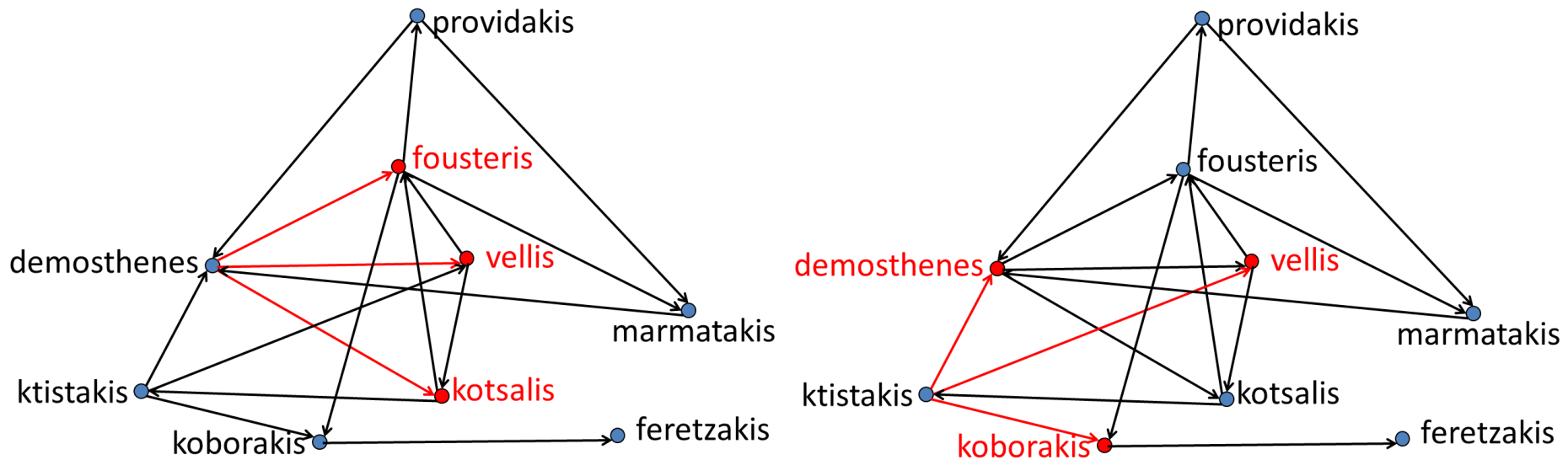
# Πως θα απαντούσατε το ερώτημα;

- Ποιοι είναι αυτοί με τους οποίους ήρθαν σε επαφή και ο demosthenes και ο ktistakis;



# Απάντηση

- Τομή ( $\cap$ ) όσων ήρθε σε επαφή ο demosthenes με αυτούς που ήρθε σε επαφή ο ktistakis;

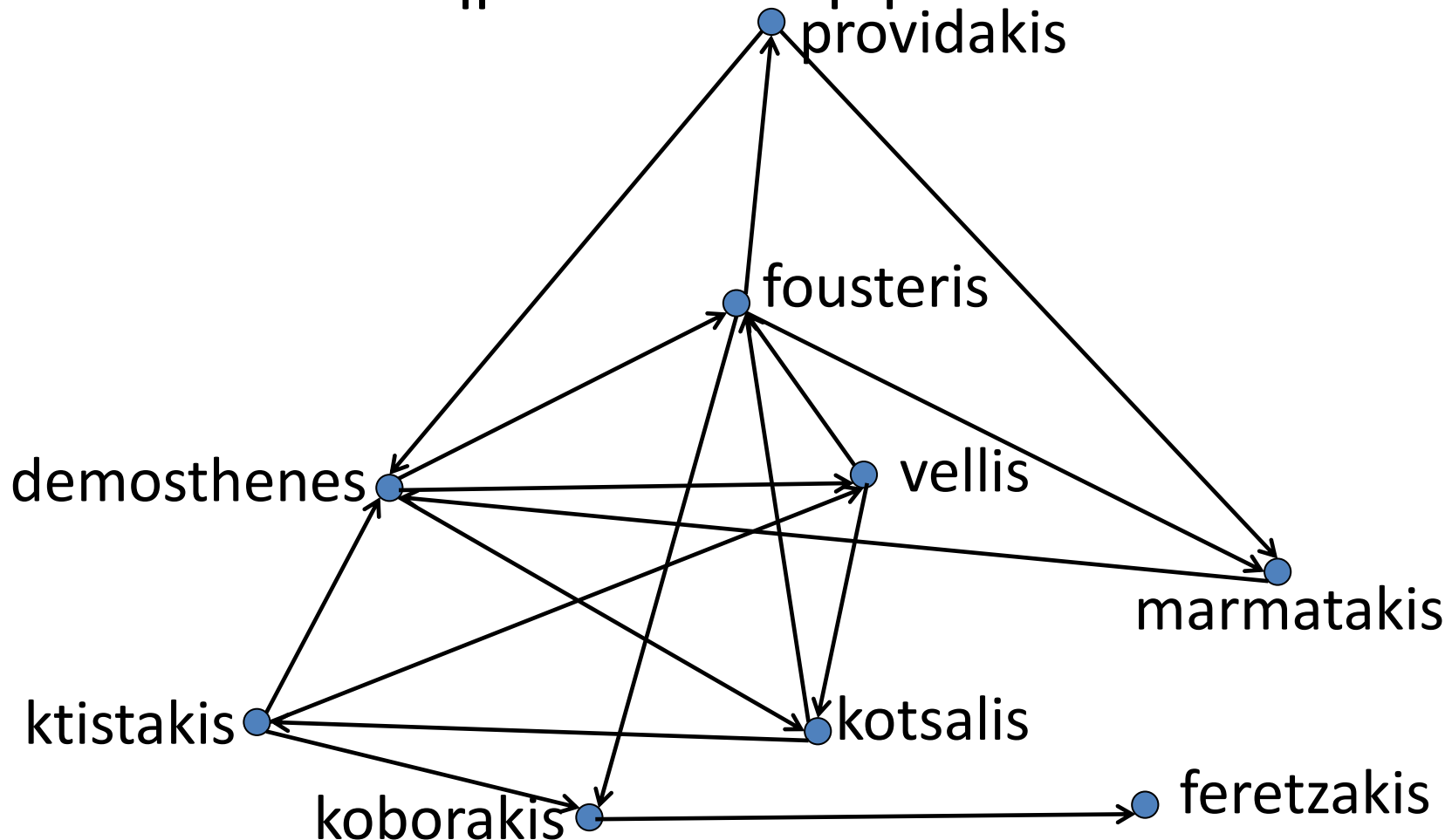


- Η απάντηση είναι

$$\{\text{fousteris, vellis, kotsalis}\} \cap \{\text{demosthenes, vellis, koborakis}\} = \{\text{vellis}\}$$

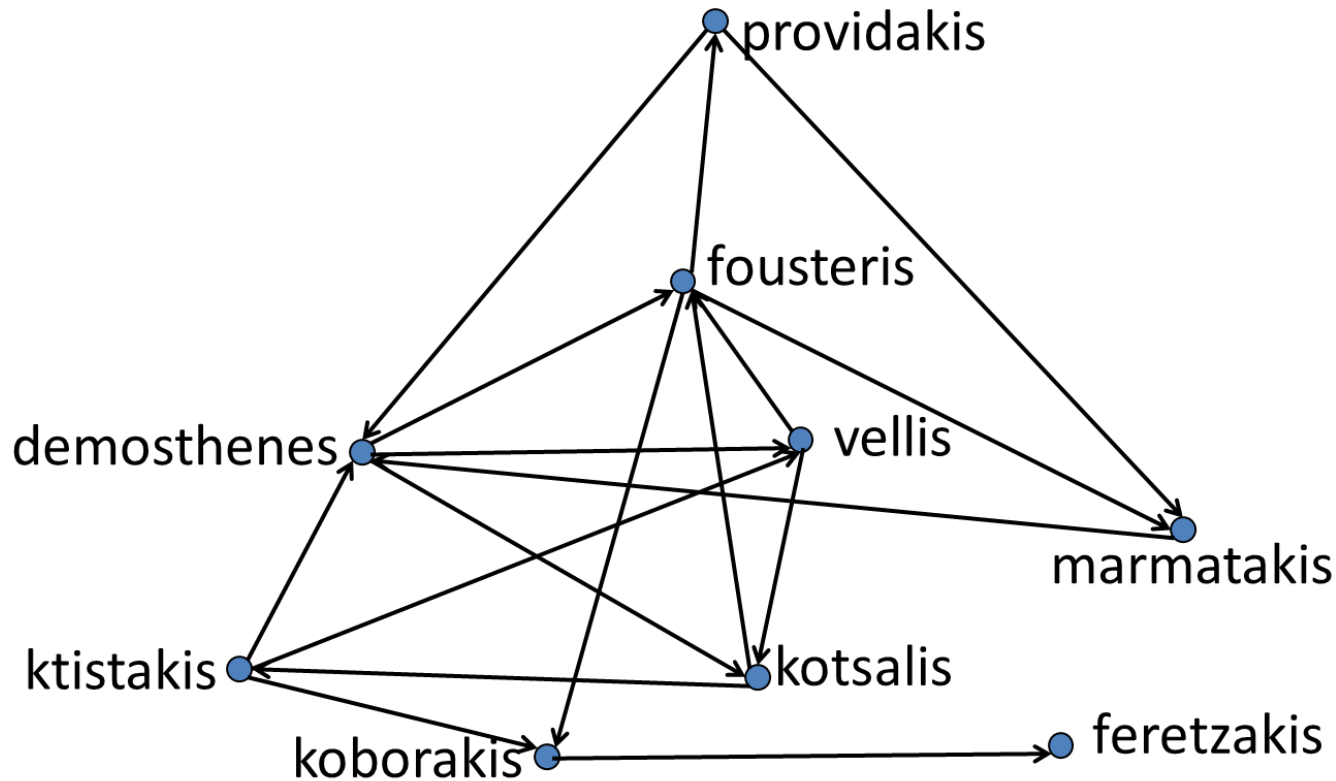
# Πως θα απαντούσατε το ερώτημα;

- Υπολογίστε αυτούς που ήρθαν σε επαφή με κάποιον που ήρθε σε επαφή και ο demosthenes;



# Απάντηση

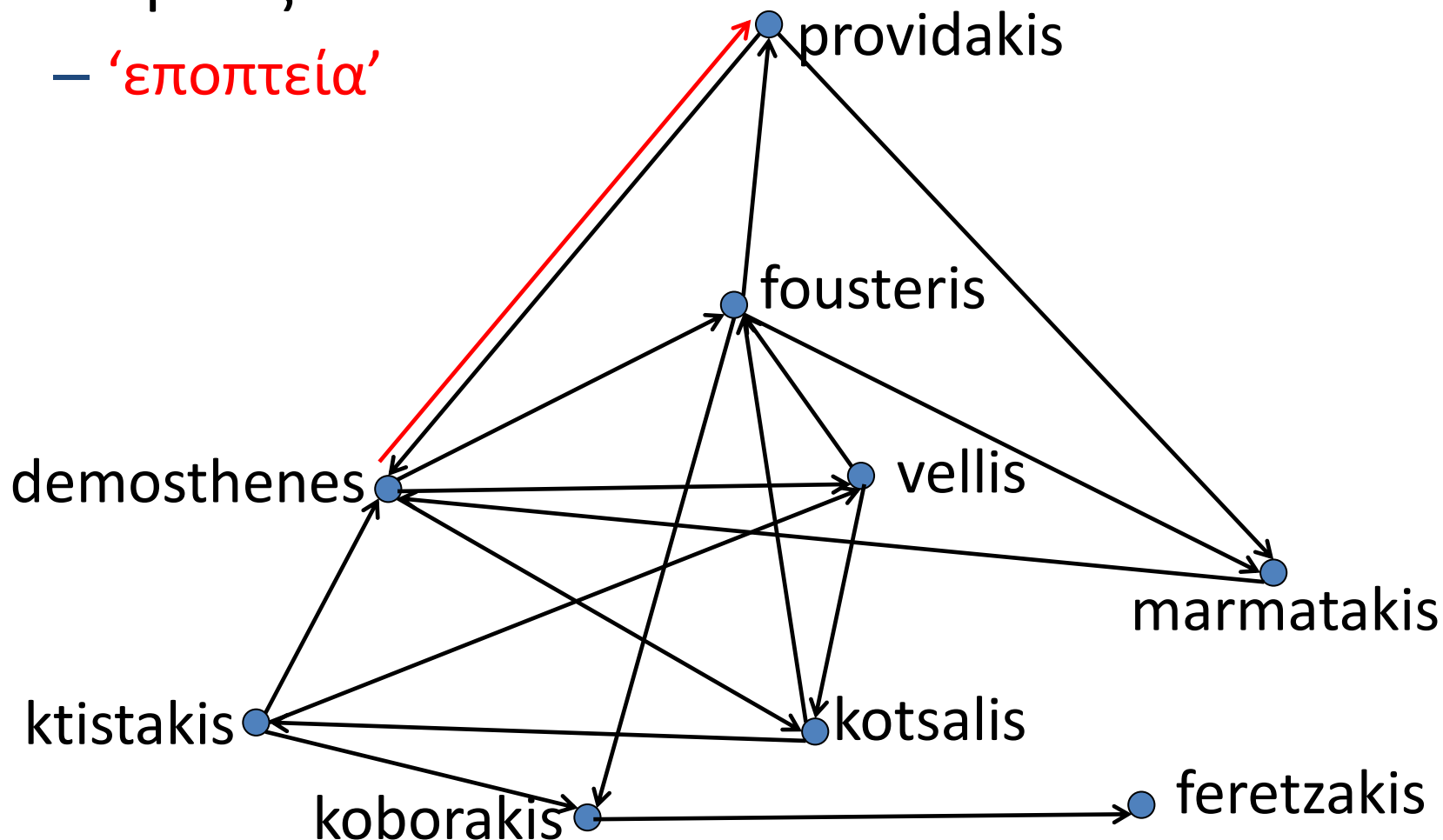
- Ανάγκη αναδρομικής επεξεργασίας
  - Για καθένα που ήρθε σε επαφή ο demosthenes, βρες αυτούς με τους οποίους ήρθε σε επαφή



# Πολλαπλοί δεσμοί

- Χρήστες μπορεί να συνδέονται με διαφορετικούς δεσμούς

– ‘εποπτεία’

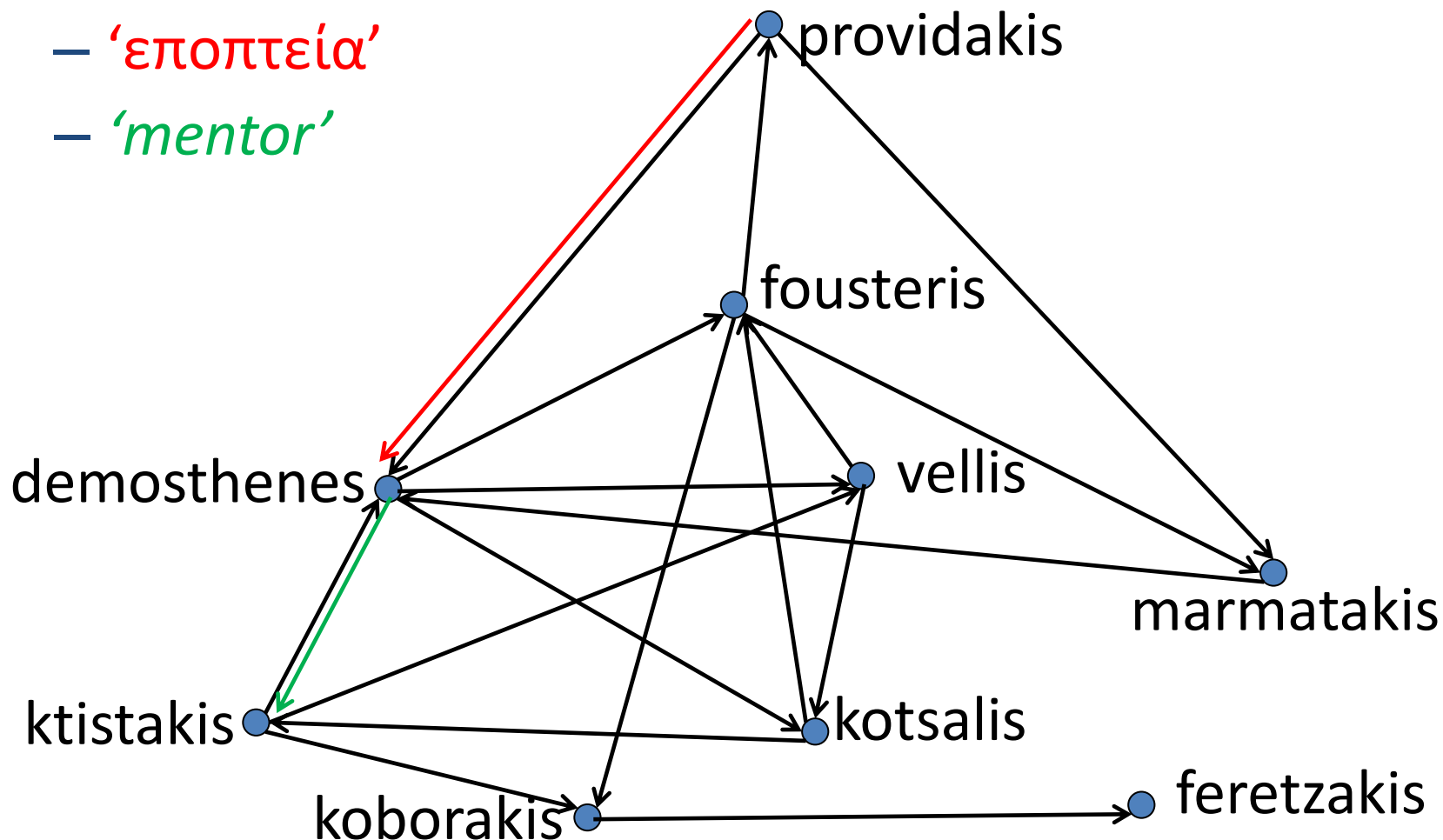


# Πολλαπλοί δεσμοί

- Χρήστες μπορεί να συνδέονται με διαφορετικούς δεσμούς

- *‘εποπτεία’*

- *‘mentor’*

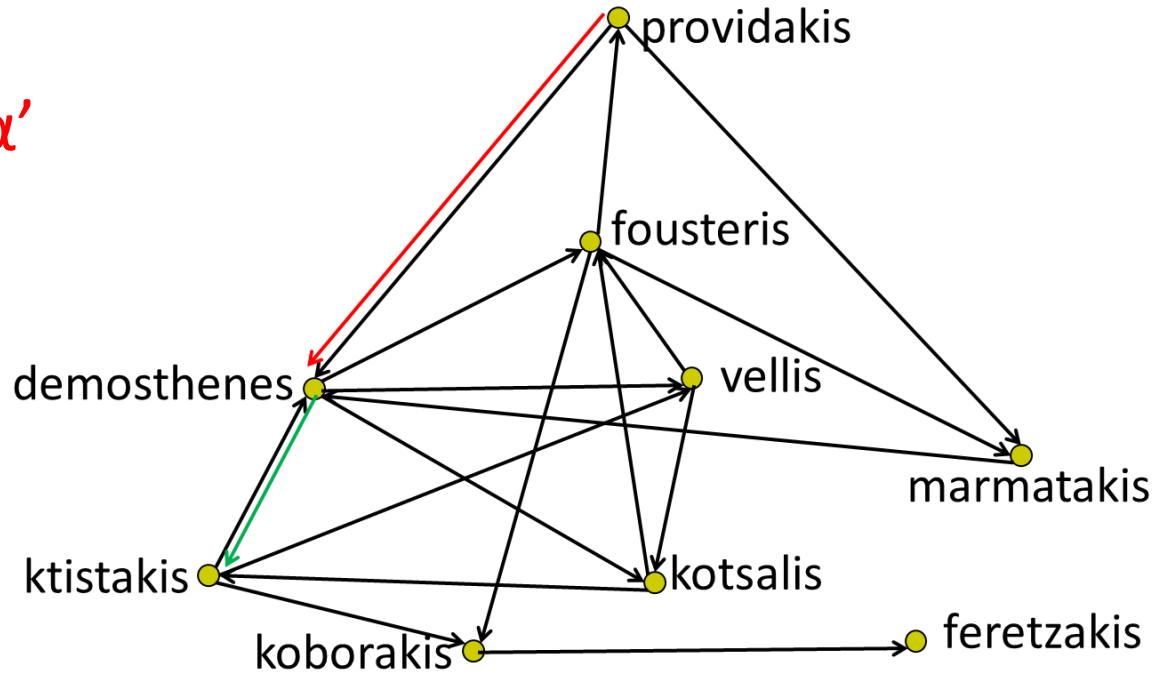


# Πολλαπλοί δεσμοί

- Χρήστες μπορεί να συνδέονται με διαφορετικούς δεσμούς

- ‘εποπτεία’

- ‘mentor’



- Ποιο μονοπάτι / διαδρομή διαλέγω όταν βρεθώ σε ένα κόμβο από τον οποίο διέρχονται πολλοί σύνδεσμοι;

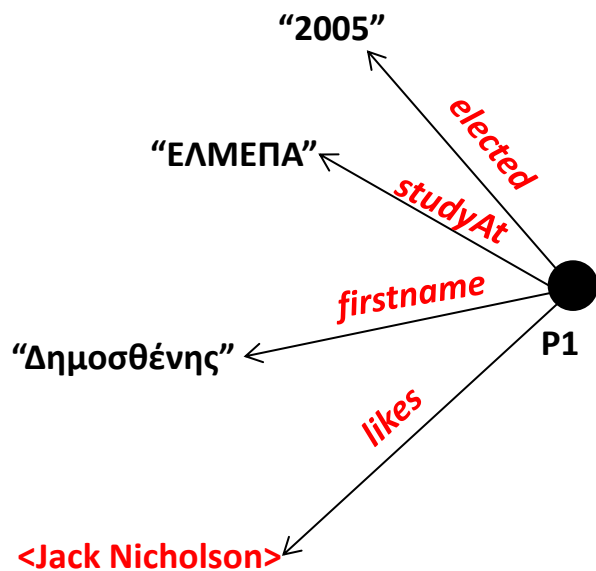
Τι γίνεται όταν δεν υπάρχει ενιαίο  
σχήμα για τα δεδομένα;

# Πολυπλοκότητα

- Φανταστείτε ότι για κάθε χρήστη γνωρίζουμε διαφορετικά πράγματα

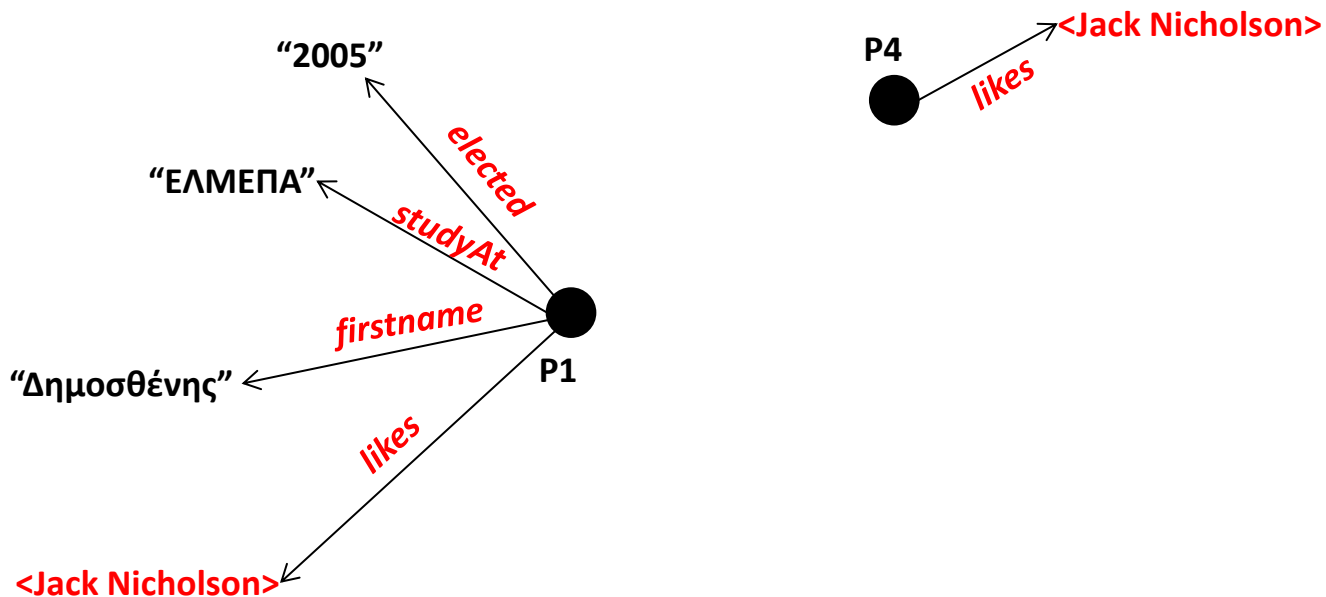
# Παράδειγμα (συν.)

- ... για τον P1



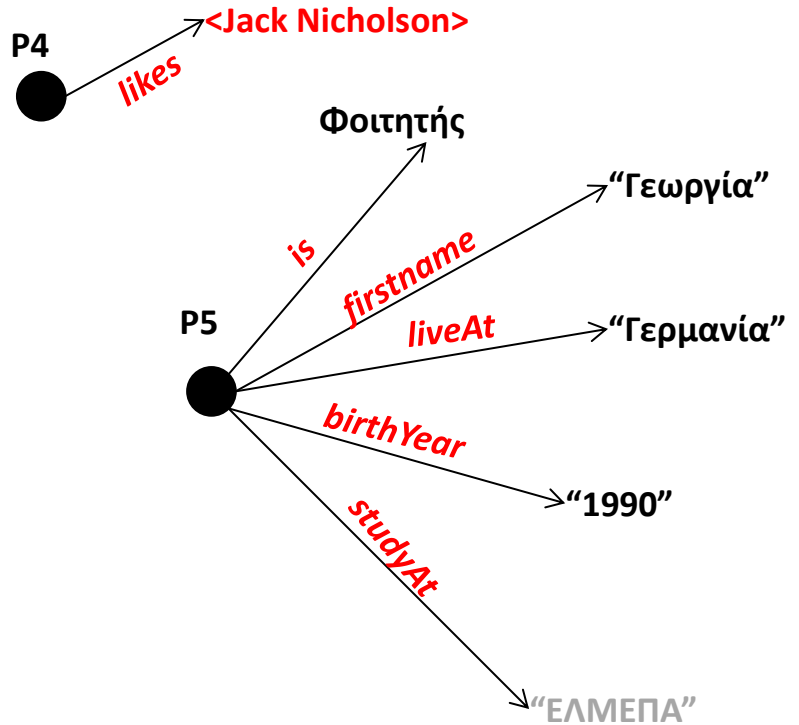
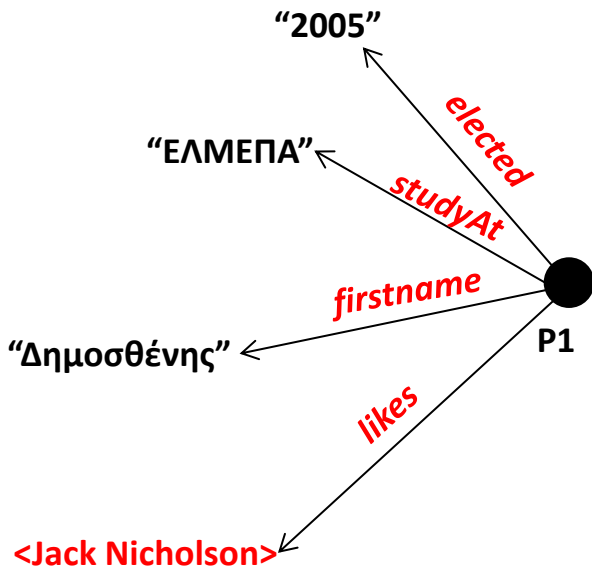
# Παράδειγμα (συν.)

- ... για τον P4



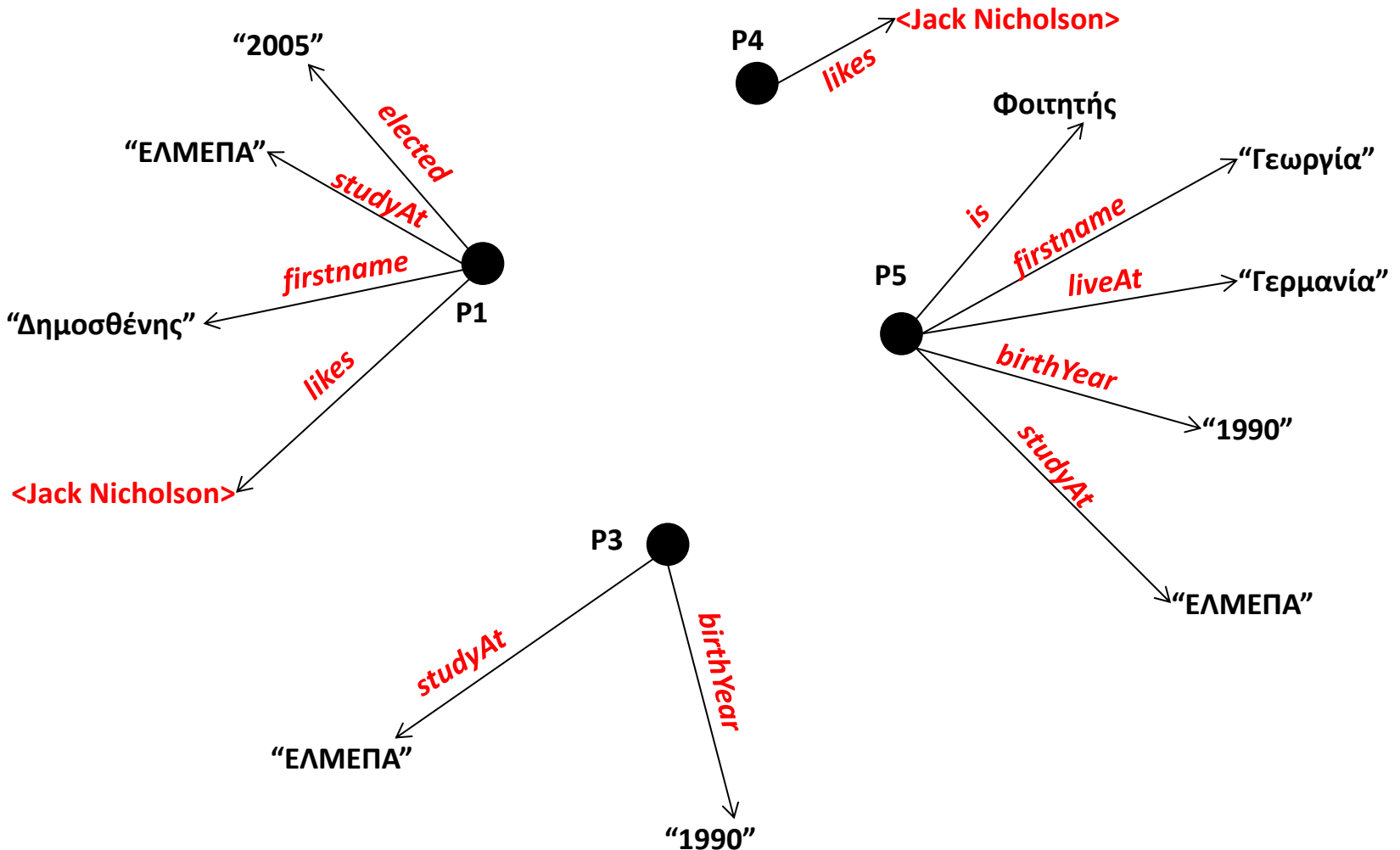
# Παράδειγμα (συν.)

- ... για τον P5



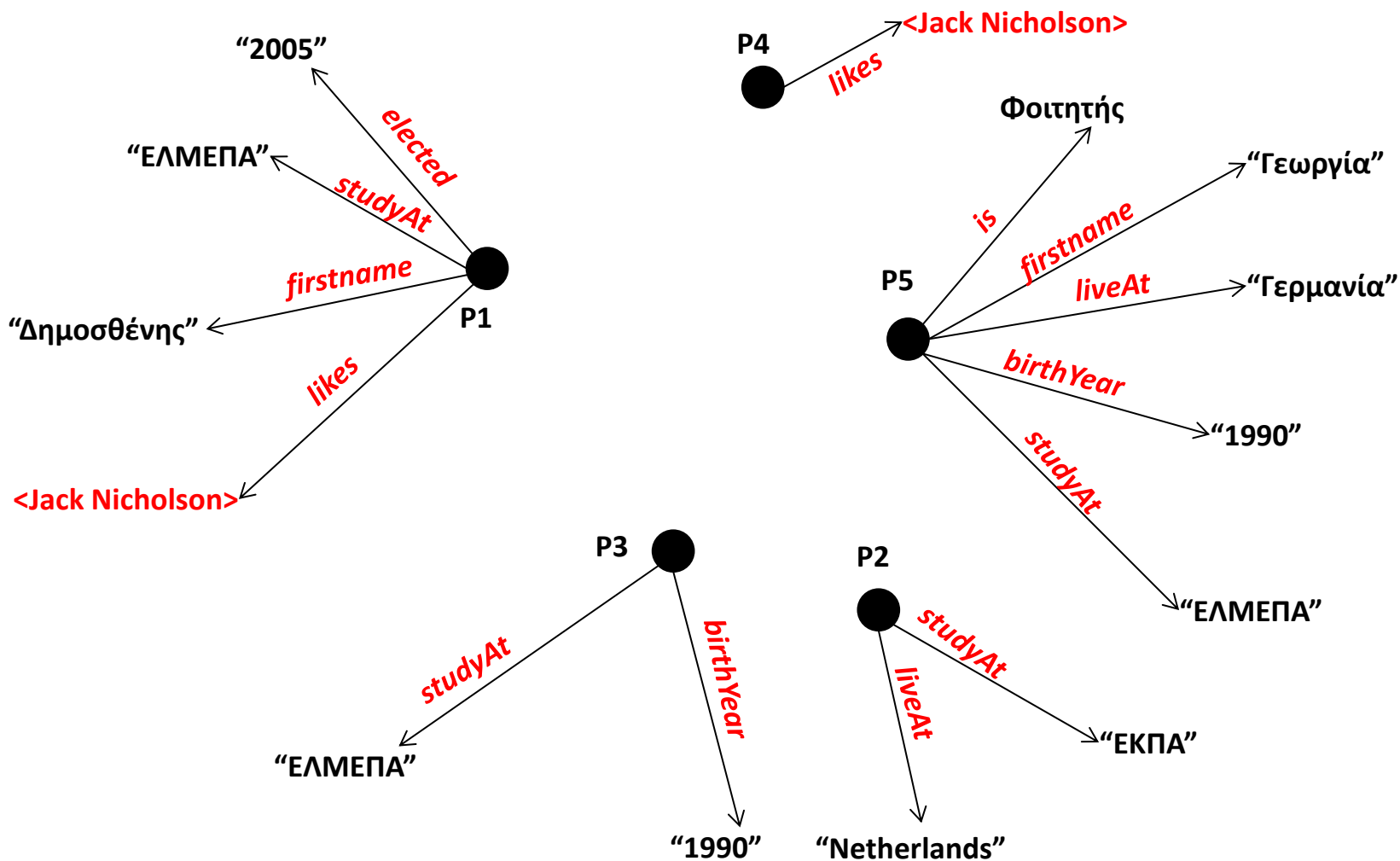
# Παράδειγμα (συν.)

- ... για τον P3



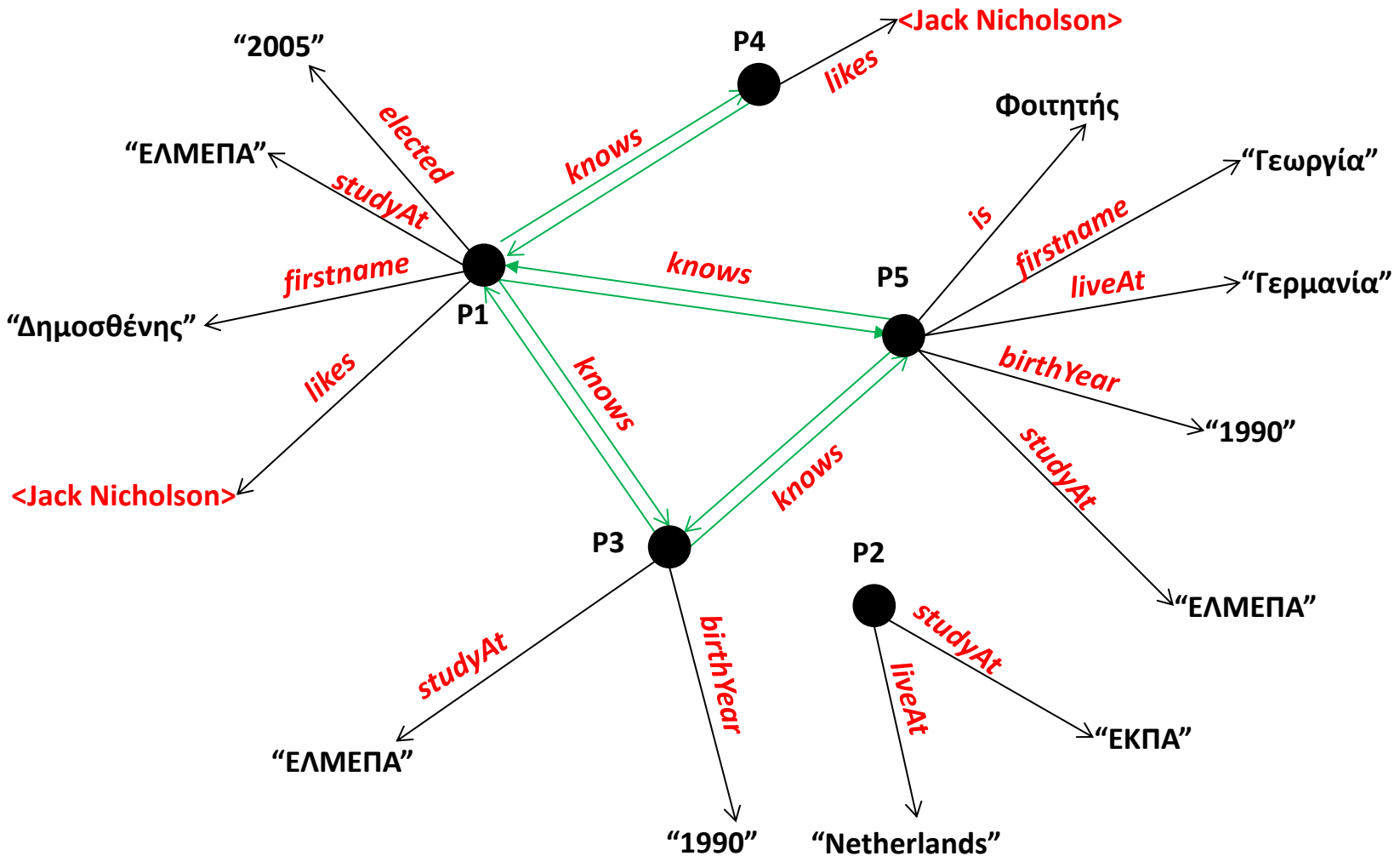
# Παράδειγμα (συν.)

- ... για τον P2



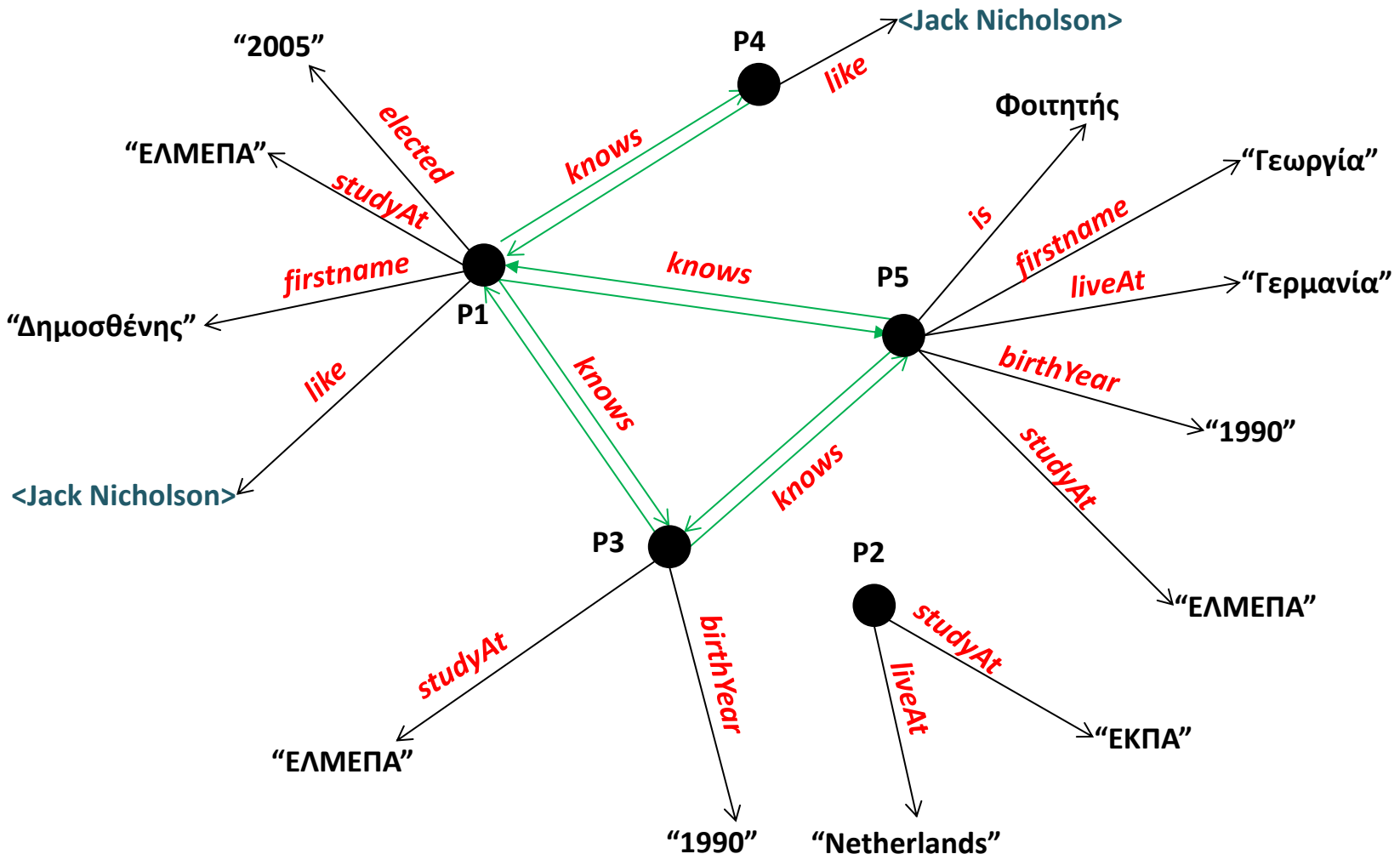
# Παράδειγμα (συν.)

- ... και ο μεταξύ τους δεσμός



# Πως θα απαντούσατε το ερώτημα

- ... σε ποιους αρέσει ο Jack Nicholson



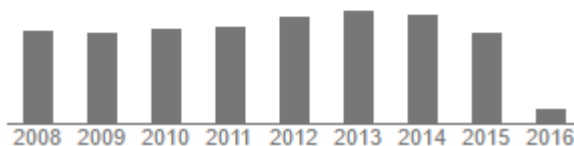
Τι γίνεται όταν τα δεδομένα είναι  
διασυνδεδεμένα;

# Αναδυόμενα θέματα

- Έμφαση στη σημασία & γνώση
  - Π.χ. Τι είναι γνωστό γενικά για ένα 'πρόσωπο', 'φαινόμενο', 'διαδικασία', 'έννοια', κλπ

Google Scholar

Citation indices	All	Since 2011
Citations	66883	21731
h-index	103	65
i10-index	414	228



Title	1-20	Cited by	Year
Designing the user interface	B Shneiderman Pearson Education India	13210	2003
Readings in information visualization: using vision to think	Morgan Kaufmann	4375	1999
The eyes have it: A task by data type taxonomy for information visualizations	B Shneiderman Visual Languages, 1996. Proceedings., IEEE Symposium on, 336-343	3524	1996
1.1 direct manipulation: a step beyond programming languages	B Shneiderman Sparks of innovation in human-computer interaction 17, 1993	2158	1993
Tree-maps: A space-filling approach to the visualization of hierarchical information structures	B Johnson, B Shneiderman Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on, 284-291	1539	1991
Tree visualization with tree-maps: 2-d space-filling approach	B Shneiderman ACM Transactions on graphics (TOG) 11 (1), 92-99	1516	1992
Visual information seeking: Tight coupling of dynamic query filters with starfield displays	C Ahlberg, B Shneiderman Proceedings of the SIGCHI conference on Human factors in computing systems ...	1194	1994
Software psychology	B Shneiderman Winthrop	1136	1980
The information visualizer, an information workspace	SK Card, GG Robertson, JD Mackinlay Proceedings of the SIGCHI Conference on Human factors in computing systems ...	761	1991
Dynamic queries for visual information seeking	B Shneiderman Software, IEEE 11 (6), 70-77	739	1994

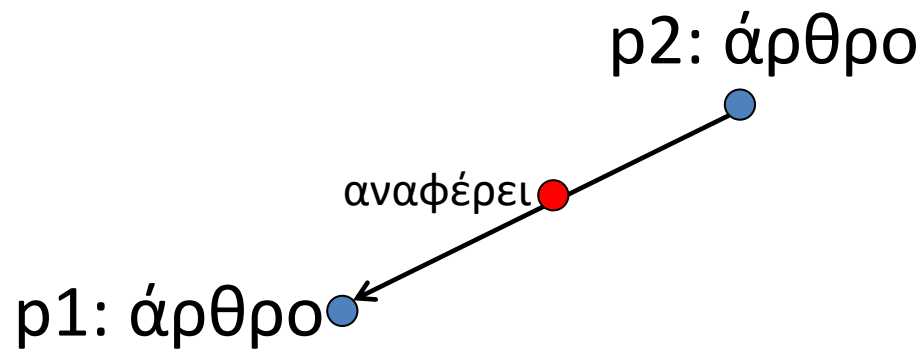
# Παράδειγμα

- Πως θα μελετούσατε ετεροαναφορές ή άρθρα που αναφέρονται το ένα στο άλλο
  - Έστω το άρθρο p1

p1: άρθρο●

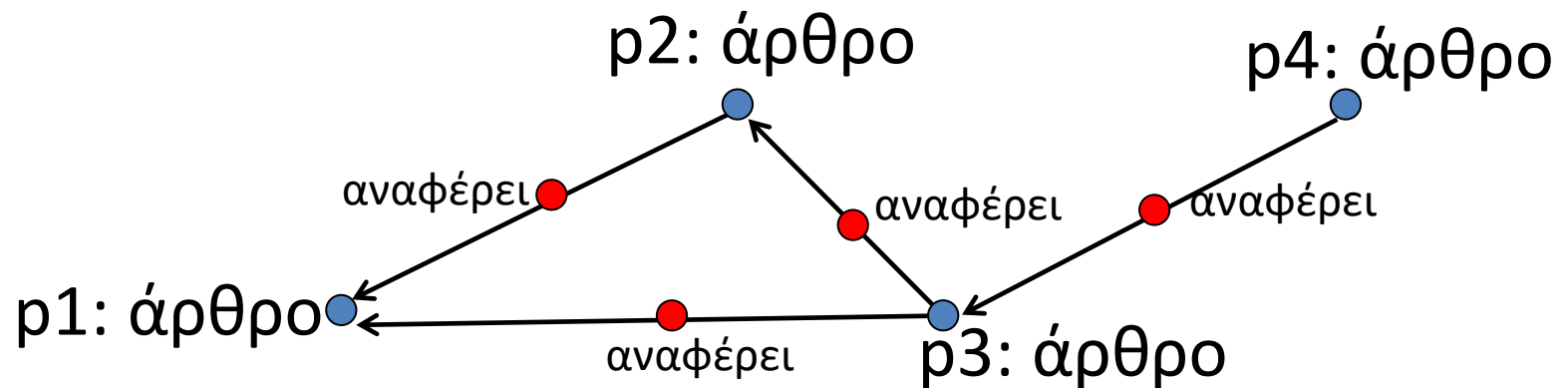
# Παράδειγμα

- Πως θα μελετούσατε ετεροαναφορές ή άρθρα που αναφέρονται το ένα στο άλλο
  - Έστω το άρθρο p1
  - Το άρθρο p2 αναφέρει το p1



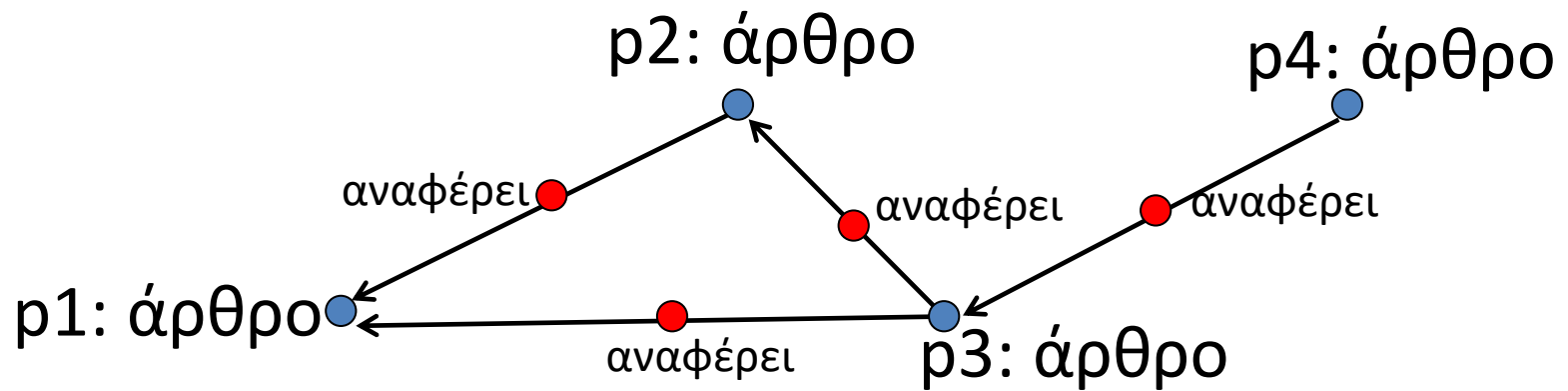
# Παράδειγμα

- Πως θα μελετούσατε ετεροαναφορές ή άρθρα που αναφέρονται το ένα στο άλλο
  - Έστω το άρθρο p1
  - Το άρθρο p2 αναφέρει το p1
  - Ακολουθούν τα άρθρα p3 και p4 ως εξής



# Παράδειγμα

- Πως θα μελετούσατε ετεροαναφορές ή άρθρα που αναφέρονται το ένα στο άλλο
  - Έστω το άρθρο p1
  - Το άρθρο p2 αναφέρει το p1
  - Ακολουθούν τα άρθρα p3 και p4 ως εξής



- Πως θα υπολογίζατε
  - το/τα αρχαιότερο/α άρθρο/α;
  - τα έτη που γίνονται οι αναφορές;

Τι γίνεται όταν τα δεδομένα  
διανέμονται υπό μορφή αρχείων  
(κάποιου τύπου);

# Datasets

- CSV αρχείο

ΠΙΘΑΝΗ ΑΠΩΛΕΙΑ ΔΕΔΟΜΕΝΩΝ Ορισμένες δυνατότητες ενδέχεται να χαθούν εάν αποθηκεύσετε αυτό το βιβλίο εργασίας σε μορφή α


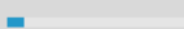
T11

	A	B	C	D	E	F	G	H	I	J	K	L
1	raceId	year	round	circuitId	name	date	time	url				
2	1,2009	1,1	Australian Grand Prix	29/03/2009	00:00	06:00:00	http://en.wikipedia.org/wiki/2009_Australian_Grand_Prix					
3	2,2009	2,2	Malaysian Grand Prix	05/04/2009	00:00	09:00:00	http://en.wikipedia.org/wiki/2009_Malaysian_Grand_Prix					
4	3,2009	3,17	Chinese Grand Prix	19/04/2009	00:00	07:00:00	http://en.wikipedia.org/wiki/2009_Chinese_Grand_Prix					
5	4,2009	4,3	Bahrain Grand Prix	26/04/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Bahrain_Grand_Prix					
6	5,2009	5,4	Spanish Grand Prix	10/05/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Spanish_Grand_Prix					
7	6,2009	6,6	Monaco Grand Prix	24/05/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Monaco_Grand_Prix					
8	7,2009	7,5	Turkish Grand Prix	07/06/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Turkish_Grand_Prix					
9	8,2009	8,9	British Grand Prix	21/06/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_British_Grand_Prix					
10	9,2009	9,20	German Grand Prix	12/07/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_German_Grand_Prix					
11	10,2009	10,11	Hungarian Grand Prix	26/07/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Hungarian_Grand_Prix					
12	11,2009	11,12	European Grand Prix	23/08/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_European_Grand_Prix					
13	12,2009	12,13	Belgian Grand Prix	30/08/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Belgian_Grand_Prix					
14	13,2009	13,14	Italian Grand Prix	13/09/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Italian_Grand_Prix					
15	14,2009	14,15	Singapore Grand Prix	27/09/2009	00:00	12:00:00	http://en.wikipedia.org/wiki/2009_Singapore_Grand_Prix					
16	15,2009	15,22	Japanese Grand Prix	04/10/2009	00:00	05:00:00	http://en.wikipedia.org/wiki/2009_Japanese_Grand_Prix					
17	16,2009	16,18	Brazilian Grand Prix	18/10/2009	00:00	16:00:00	http://en.wikipedia.org/wiki/2009_Brazilian_Grand_Prix					
18	17,2009	17,24	Abu Dhabi Grand Prix	01/11/2009	00:00	11:00:00	http://en.wikipedia.org/wiki/2009_Abu_Dhabi_Grand_Prix					
19	18,2008	1,1	Australian Grand Prix	16/03/2008	00:00	04:30:00	http://en.wikipedia.org/wiki/2008_Australian_Grand_Prix					
20	19,2008	2,2	Malaysian Grand Prix	23/03/2008	00:00	07:00:00	http://en.wikipedia.org/wiki/2008_Malaysian_Grand_Prix					
21	20,2008	3,3	Bahrain Grand Prix	06/04/2008	00:00	11:30:00	http://en.wikipedia.org/wiki/2008_Bahrain_Grand_Prix					
22	21,2008	4,4	Spanish Grand Prix	27/04/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Spanish_Grand_Prix					
23	22,2008	5,5	Turkish Grand Prix	11/05/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Turkish_Grand_Prix					
24	23,2008	6,6	Monaco Grand Prix	25/05/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Monaco_Grand_Prix					
25	24,2008	7,7	Canadian Grand Prix	08/06/2008	00:00	17:00:00	http://en.wikipedia.org/wiki/2008_Canadian_Grand_Prix					
26	25,2008	8,8	French Grand Prix	22/06/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_French_Grand_Prix					
27	26,2008	9,9	British Grand Prix	06/07/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_British_Grand_Prix					
28	27,2008	10,10	German Grand Prix	20/07/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_German_Grand_Prix					
29	28,2008	11,11	Hungarian Grand Prix	03/08/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Hungarian_Grand_Prix					
30	29,2008	12,12	European Grand Prix	24/08/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_European_Grand_Prix					
31	30,2008	13,13	Belgian Grand Prix	07/09/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Belgian_Grand_Prix					
32	31,2008	14,14	Italian Grand Prix	14/09/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Italian_Grand_Prix					
33	32,2008	15,15	Singapore Grand Prix	28/09/2008	00:00	12:00:00	http://en.wikipedia.org/wiki/2008_Singapore_Grand_Prix					
34	33,2008	16,16	Japanese Grand Prix	12/10/2008	00:00	04:30:00	http://en.wikipedia.org/wiki/2008_Japanese_Grand_Prix					
35	34,2008	17,17	Chinese Grand Prix	19/10/2008	00:00	07:00:00	http://en.wikipedia.org/wiki/2008_Chinese_Grand_Prix					
36	35,2008	18,18	Brazilian Grand Prix	02/11/2008	00:00	17:00:00	http://en.wikipedia.org/wiki/2008_Brazilian_Grand_Prix					
37	36,2007	1,1	Australian Grand Prix	18/03/2007	00:00	03:00:00	http://en.wikipedia.org/wiki/2007_Australian_Grand_Prix					
38	37,2007	2,2	Malaysian Grand Prix	08/04/2007	00:00	07:00:00	http://en.wikipedia.org/wiki/2007_Malaysian_Grand_Prix					
39	38,2007	3,3	Bahrain Grand Prix	15/04/2007	00:00	11:30:00	http://en.wikipedia.org/wiki/2007_Bahrain_Grand_Prix					
40	39,2007	4,4	Spanish Grand Prix	13/05/2007	00:00	12:00:00	http://en.wikipedia.org/wiki/2007_Spanish_Grand_Prix					
41	40,2007	5,6	Monaco Grand Prix	27/05/2007	00:00	12:00:00	http://en.wikipedia.org/wiki/2007_Monaco_Grand_Prix					

racas +

# Datasets (cont.)


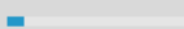
- Bus movement in the Greater area of the city of London

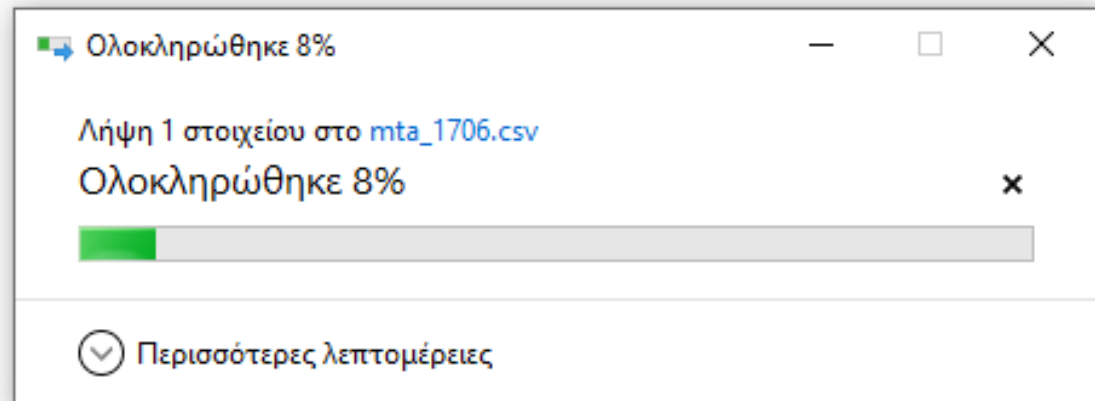
Όνομα	Κατάσταση	Ημερομηνία τροποποι...	Τύπος	Μέγεθος
 mta_1706		15/2/2023 12:49 μμ	Αρχείο τιμών δια...	1.380.524



# Datasets (cont.)


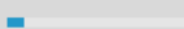
- Bus movement in the Greater area of the city of London

Όνομα	Κατάσταση	Ημερομηνία τροποποι...	Τύπος	Μέγεθος
 mta_1706		15/2/2023 12:49 μμ	Αρχείο τιμών δια...	1.380.524



# Datasets (cont.)

- Bus movement in the Greater area of the city of London

Όνομα	Κατάσταση	Ημερομηνία τροποποι...	Τύπος	Μέγεθος
 mta_1706		15/2/2023 12:49 μμ	Αρχείο τιμών δια...	1.380.524

 Microsoft Excel



Αυτό το σύνολο δεδομένων είναι πολύ μεγάλο για το πλέγμα του Excel. Εάν αποθηκεύσετε αυτό το βιβλίο εργασίας, θα χάσετε δεδομένα που δεν έχουν φορτωθεί.

OK

# Datasets (cont.)

- Once available for review

The screenshot shows an Excel spreadsheet titled "mta\_1706 - Excel" with the following columns: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, AB, AC. The rows contain detailed bus route information, including route IDs (e.g., 2017-06-01 00:03:34), line names (e.g., B8,4 AV/95 ST), origin and destination names (e.g., BROWNSVILLE ROCKAWAY AV, FOSTER AV), and arrival times (e.g., 2017-06-01 00:03:59, 24:06:14).

Route ID	Direction	Line Name	Origin Name	Origin Lat	Origin Long	Destination Name	Destination Lat	Destination Long	Vehicle Ref	Vehicle Location	Latitude	Longitude	Next Stop Point Name	Arrival Proximity Text	Distance From Stop	Expected Arrival Time	Scheduled Arrival Time		
2017-06-01 00:03:34	B8,4 AV/95 ST	40.616104	-74.031143	BROWNSVILLE ROCKAWAY AV	40.656048	-73.907379	NYCT_430	40.63517	-73.960803	FOSTER AV/E 18 ST	approaching	76	2017-06-01 00:03:59	24:06:14					
2017-06-01 00:03:43	1,S61,ST GEORGE FERRY/S61 & S91	40.643169	-74.073494	S I MALL YUKON AV	40.575935	-74.167686	NYCT_8263	40.590802	-74.15834	MERRYMOUNT ST/TRAVIS AV	approaching	62	2017-06-01 00:03:56	23:58:02					
2017-06-01 00:03:49	0,Bx10,E 206 ST/BAINBRIDGE AV	40.875008	-73.880142	RIVERDALE 263 ST	40.912376	-73.902534	NYCT_4223	40.88601	-73.912647	HENRY HUDSON PKY E/W 235 ST	at stop	5	2017-06-01 00:03:56	24:00:53					
2017-06-01 00:03:31	0,Q5,TEARDROP/LAYOVER	40.701748	-73.802399	ROSEDALE LIRR STA	via MERRICK	40.666012	-73.735939	NYCT_8422	40.668002	-73.729348	HOOK CREEK BL/SUNRISE HY	< 1 stop away	267	2017-06-01 00:04:03	24:03:00				
2017-06-01 00:03:22	1,Bx1,RIVERDALE AV/W 231 ST	40.881187	-73.90934	MOTT HAVEN 136 ST	via CONCOURSE	40.809654	-73.92836	NYCT_4710	40.868134	-73.893032	GRAND CONCOURSE/E 196 ST	at stop	11	2017-06-01 00:03:56	23:59:38				
2017-06-01 00:03:40	0,M1,4 AV/E 10 ST	40.731342	-73.990288	HARLEM 147 ST	via MADISON	40.82111	-73.935898	NYCT_3831	40.792897	-73.950023	MADISON AV/E 106 ST	approaching	73	2017-06-01 00:03:56	24:02:35				
2017-06-01 00:03:24	0,B31,GERRITSEN AV/GERRITSEN BEACH	40.587101	-73.918503	MIDWOOD KINGS HWY STA	40.608433	-73.9571	NYCT_4611	40.587024	-73.918623	GERRITSEN AV/GERRITSEN BEACH	at stop	0	NA	24:08:00					
2017-06-01 00:03:29	0,B83,GATEWAY CTR TERM/GATEWAY DR	40.652649	-73.877029	BWAY JCT VN SNDRN AV	40.678139	-73.903572	NYCT_4841	40.648801	-73.882682	PENNSYLVANIA AV/DELMAR LOOP N	< 1 stop away	196	2017-06-01 00:04:13	23:58:47					
2017-06-01 00:03:27	0,B82,STILLWELL TERMINAL BUS LOOP	40.57708	-73.981293	SPRING CRK TWRS SEVIEW AV	via KINGS HWY	40.64299	-73.878326	NYCT_6592	40.632258	-73.918318	FLATLANDS AV/RALPH AV	approaching	35	2017-06-01 00:03:56	24:00:00				
2017-06-01 00:03:51	1,S59,RICHMOND TER/PARK AV #3	40.640167	-74.130966	HYLAN BL	40.53426	-74.154213	NYCT_8279	40.590689	-74.165811	RICHMOND AV/NOME AV	approaching	31	2017-06-01 00:03:56	24:01:14					
2017-06-01 00:03:48	0,Bx28,VALENTINE AV/E 192 ST	40.864079	-73.894615	CO-OP CITY EARHART LA	via GUNHILL	40.860828	-73.82267	NYCT_8334	40.876032	-73.829543	DREISER LOOP/DEBS PL	< 1 stop away	207	2017-06-01 00:04:29	23:48:35				
2017-06-01 00:03:52	1,B1,87 ST/4 Av	40.621925	-74.028496	MANHATTAN BEACH KINGSBORO CC	40.578094	-73.94017	NYCT_7141	40.602164	-73.994162	86 ST/BAY PY	at stop	22	2017-06-01 00:03:57	24:05:55					
2017-06-01 00:03:35	1,B26,WYCKOFF ST/PALMETTO ST	40.69915	-73.910477	DNTWN BKLYN TILLARY ST	via HALSEY	40.693939	-73.990311	NYCT_6503	40.693922	-73.990411	CADMAN PZ W/TILLARY ST	approaching	47	NA					
2017-06-01 00:03:34	0,Bx39,SOUNDVIEW AV/STEPHENS AV	40.807869	-73.852715	WAKEFIELD 241 ST	via WHITE PLS RD	40.903309	-73.849922	NYCT_4723	40.869763	-73.867221	WHITE PLAINS RD/BURKE AV	< 1 stop away	170	2017-06-01 00:04:18	24:04:12				
2017-06-01 00:03:24	1,M66,E 67 ST/YORK AV	40.763908	-73.956299	LINCOLN CTR WEST END AV	XTOWN	40.775906	-73.986809	NYCT_5131	40.773762	-73.981962	W 66 ST/BROADWAY	approaching	76	2017-06-01 00:04:15	24:00:37				
2017-06-01 00:03:36	1,Bx31,VAN CORTLANDT PK E/E 242 ST	40.90266	-73.868378	WESTCHESTER SQ	40.841076	-73.84285	NYCT_274	40.841608	-73.843465	TREMONT AV/LANE AV	approaching	78	2017-06-01 00:03:55	24:05:00					
2017-06-01 00:03:53	0,Bx36,W 179 ST/BROADWAY	40.849113	-73.937752	SOUNDVIEW PUGSLEY AV	40.820507	-73.851631	NYCT_7766	40.84745	-73.887448	E 180 ST/PROSPECT AV	approaching	113	2017-06-01 00:04:35	24:01:16					
2017-06-01 00:03:47	0,M96,W 96 ST/WEST END AV	40.794933	-73.972687	EAST SIDE 1 AV	CROSTOWN	40.784222	-73.94474	NYCT_6678	40.783676	-73.944412	E 97 ST/1 AV	approaching	86	2017-06-01 00:04:13	24:06:00				
2017-06-01 00:03:35	1,Q4,235 ST/LINDEN BL	40.69054	-73.727448	JAMAICA PRSNS - ARCHR	via LINDEN	40.701683	-73.802475	NYCT_8408	40.695912	-73.782591	MERRICK BL/110 AV	approaching	53	2017-06-01 00:03:56	24:00:37				
2017-06-01 00:03:53	1,Q54,JAMAICA AV/170 ST	40.707287	-73.789581	WILLAMSBURG BRIDGE PLZ	40.709476	-73.959534	NYCT_3988	40.709274	-73.960021	WASHINGTON PLZ/BAY 3	approaching	62	2017-06-01 00:04:10	23:46:00					
2017-06-01 00:03:41	0,B6,HARWAY AV/BAY 37 ST	40.59351	-73.993996	EAST NY NEW LOTS STA	40.66642	-73.883385	NYCT_7158	40.645676	-73.901474	GLENWOOD RD/E 98 ST	at stop	4	2017-06-01 00:03:56	24:05:00					
2017-06-01 00:03:36	0,B4,NARROWS AV/77 ST	40.631836	-74.036377	SHEEPSHEAD BAY KNAPP ST	40.586628	-73.931702	NYCT_445	40.585739	-73.937214	SHORE PKY/BROWN ST	at stop	25	2017-06-01 00:03:56	24:01:49					
2017-06-01 00:03:32	1,M101,ST NICHOLAS AV/W 192 ST	40.856289	-73.928764	EAST VILLAGE 6 ST	via LEX AV	40.729115	-73.990257	NYCT_5642	40.752234	-73.975473	LXINGTON AV/E 42 ST	< 1 stop away	178	2017-06-01 00:04:33	23:55:00				
2017-06-01 00:03:53	0,Bx5,SOUTHERN BL/WESTCHESTER AV	40.823673	-73.891945	WHITE PLAINS RD	40.823059	-73.855736	NYCT_1280	40.823108	-73.85593	TURNBULL AV/ PUGSLEY AV	at stop	14	2017-06-01 00:03:56	24:02:00					
2017-06-01 00:03:45	0,Q12,ROOSEVELT AV/LAYOVER	40.758999	-73.831757	LITTLE NECK GLENWOOD ST	via NORTHERN	40.717167	-73.733307	NYCT_7379	40.759866	-73.828936	ROOSEVELT AV/MAIN ST	approaching	106	2017-06-01 00:06:12	24:05:00				
2017-06-01 00:03:32	1,B43,BOX ST/MANHATTAN AV	40.737194	-73.955673	LEFRTS GDNS PROSPECT PK STA	40.660858	-73.96138	NYCT_342	40.721454	-73.947169	GRAHAM AV/ENGERT AV	approaching	50	2017-06-01 00:03:56	24:05:23					
2017-06-01 00:03:45	1,M100,BROADWAY/W 219 ST	40.871967	-73.91304	E HARLM 1 AV-125 ST	via BDVY	via AMSTRDM	40.802273	-73.931145	NYCT_4394	40.804815	-73.938466	E 125 ST/LXINGTON AV	approaching	136	2017-06-01 00:04:52	24:00:36			
2017-06-01 00:03:35	1,M11,BROADWAY/W 133 ST	40.818493	-73.95594	WEST VILL AMBUDON SQ	via 9 AV	40.737072	-74.005867	NYCT_6372	40.77347	-73.981669	COLUMBUS AV/W 66 ST	at stop	12	2017-06-01 00:03:56	23:57:00				
2017-06-01 00:03:29	1,B11,FLATBUSH AV/E 31 ST	40.631992	-73.946701	SUNSET PARK 1 AV - 58 ST	40.645618	-74.023521	NYCT_251	40.644872	-74.017483	55 ST/3 AV	approaching	53	2017-06-01 00:03:56	24:05:50					
2017-06-01 00:03:33	1,M2,168 ST/AUDUBON AV	40.840645	-73.938545	EAST VILLAGE 8 ST	via 5 AV	40.730282	-73.990532	NYCT_3870	40.74089	-73.99006	5 AV/W 20 ST	< 1 stop away	193	2017-06-01 00:04:02	23:52:51				
2017-06-01 00:03:25	1,B41,CADMAN PLAZA WEST/JOHNSON ST	40.695534	-73.990829	BERGEN BCH VETERANS AV	via FLATBSH	40.620052	-73.908203	NYCT_5139	40.690245	-73.987316	LIVINGSTON ST/HOYT ST	< 1 stop away	174	2017-06-01 00:04:15	23:58:38				
2017-06-01 00:03:31	0,M34A-S85,9 AV/W 42 ST	40.758617	-73.992638	WATERSIDE	via 34 ST	via 40	40.737553	-73.973846	NYCT_5851	40.756669	-73.993913	W 34 ST/9 AV	< 1 stop away	464	2017-06-01 00:05:07	24:06:02			
2017-06-01 00:03:30	0,B8,4 AV/95 ST	40.616104	-74.031143	BROWNSVILLE ROCKAWAY AV	40.656048	-73.907379	NYCT_399	40.628821	-73.980484	18 AV/47 ST	approaching	120	2017-06-01 00:03:56	24:14:50					
2017-06-01 00:03:41	1,B26,WYCKOFF ST/PALMETTO ST	40.69915	-73.910477	DNTWN BKLYN TILLARY ST	via HALSEY	40.693939	-73.990311	NYCT_7260	40.686261	-73.974316	FULTON ST/S ELLIOTT PL	approaching	103	2017-06-01 00:04:11	24:03:01				
2017-06-01 00:03:38	0,B32,LINCOLN AV/E 138 ST	40.810955	-73.927139	KINGSBRIDGE VA HOSP	40.86739	-73.905464	NYCT_714	40.867352	-73.905387	VA HOSPITAL/VA HOSPITAL	at stop	5	2017-06-01 00:03:56	23:59:00					
2017-06-01 00:03:27	1,X10,E 57 ST/3 AV	40.760429	-73.967674	PT RICHMOND AV	NARROWS RD	via GANNON AV	40.633698	-74.129776	NYCT_2656	40.642433	-74.02014	gowanus py/6 av	0,6 miles away	1022	2017-06-01 00:04:29	24:02:56			
2017-06-01 00:03:46	0,B14,EASTERN PY/UTICA AV	40.668537	-73.931236	SPRING CREEK STANLEY AV	POSTAL FAC	40.664642	-73.863068	NYCT_7094	40.664865	-73.863265	POSTAL FACILITY RD/IN FRONT OF FACILITY	at stop	27	2017-06-01 00:03:56	23:49:00				

# Συμπέρασμα

- Είναι προφανές ότι αυτά τα προβλήματα δεν προσεγγίζονται εύκολα με σχεσιακούς όρους
- Δεν υπάρχει όμως ένας καθολικά αποδεκτός τρόπος αντιμετώπισης των ζητημάτων
- Λύσεις
  - Νέοι σύνθετοι τύποι δεδομένων
  - Εναλλακτικά μοντέλα (σημασιολογικά)
  - Προγραμματιστικές τεχνικές

# Παρατήρηση

- Διασύνδεση δεδομένων
- 'Ανοιχτά' δεδομένα
- Γράφοι δεδομένων αντί 'επίπεδης' ή 'σχεσιακής αναπαράστασης
- Πολυπλοκότητα δεδομένων

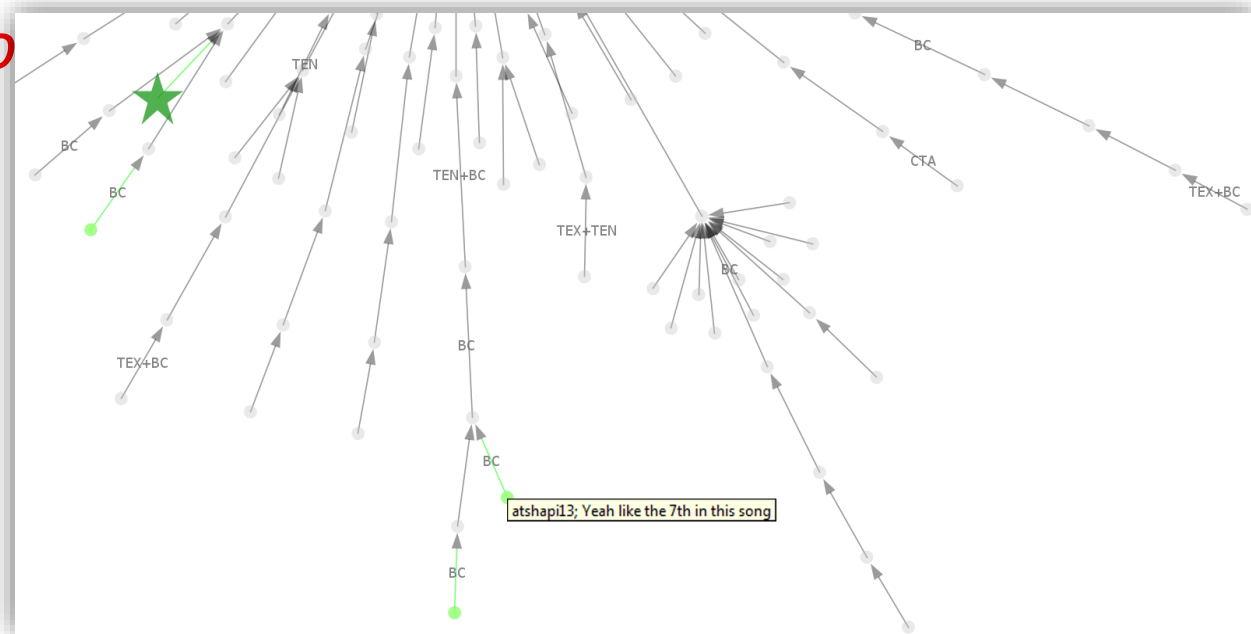
# Σύνοψη

- Το ζήτημα των μοντέλων δεδομένων συνεχίζει να αποτελεί αντικείμενο μελέτης
- Πολλές και εναλλακτικές προτάσεις, κάποιες από τις οποίες θα εξετάσουμε στη συνέχεια

Παραδείγματα προβλημάτων που  
εξετάστηκαν σε πτυχιακές

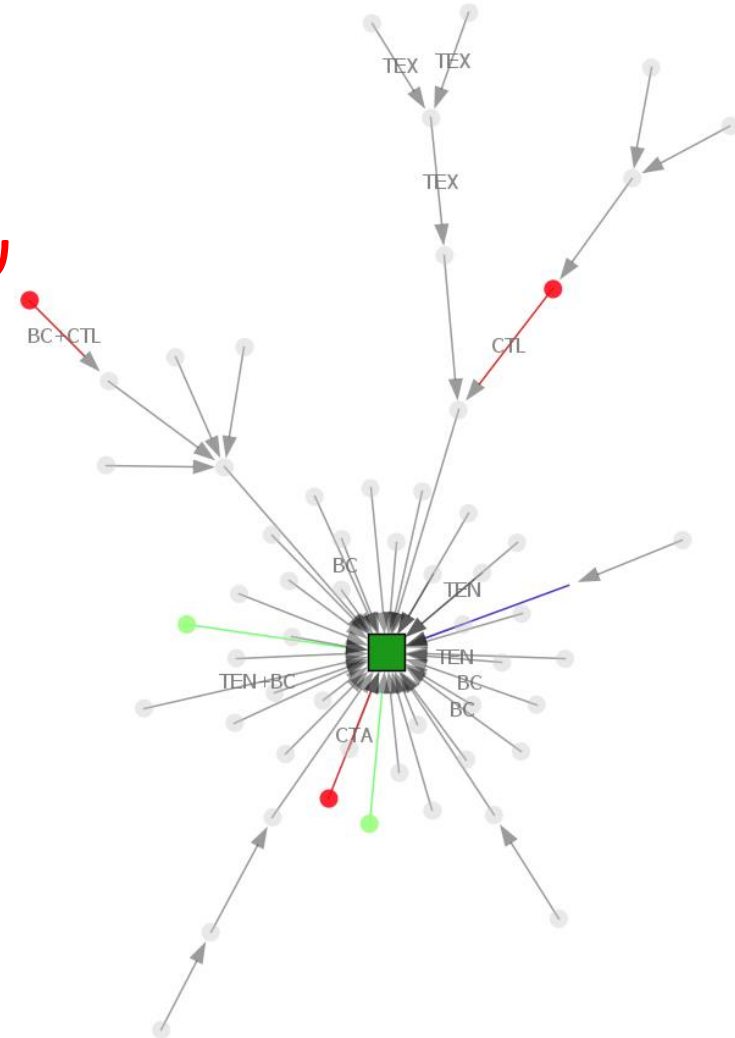
# Παρόμοια ζητήματα

- Μελέτη έντασης του δεσμού μεταξύ χρηστών ως *συνάρτηση* του *αριθμού των μηνυμάτων* που ανταλλάσσουν
  - Π.χ. ίσως ενδιαφέρει να δείτε *ποιοι χρήστες ανταλλάσσουν συνεχώς μηνύματα και ποιοι το κάνουν*



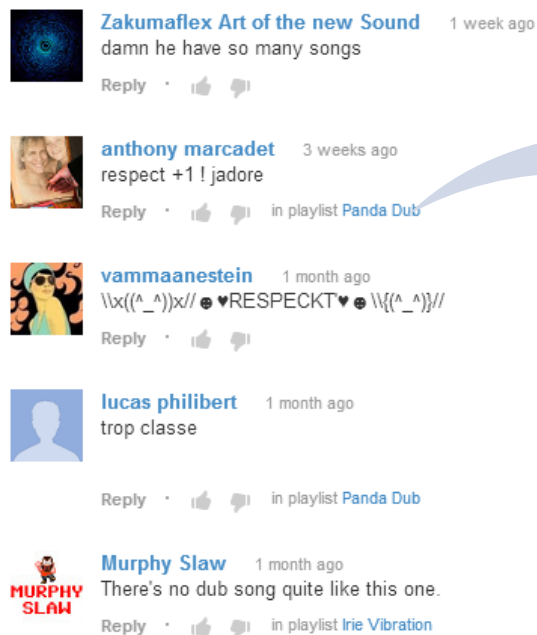
# Παρόμοια ζητήματα (συν.)

- Μελέτη της έντασης των δεσμών χρηστών ως *συνάρτηση του περιεχομένου των μηνυμάτων*
  - Ίσως ενδιαφέρει να μελετήσετε *τι λένε* οι χρήστες που ανταλλάσσουν συνεχώς μηνύματα



# Παρόμοια ζητήματα (συν.)

- Μελέτη της έντασης των δεσμών χρηστών ως **συνάρτηση** της **έκφρασης γνώμης** μεταξύ χρηστών
  - Ίσως ενδιαφέρει **ποιοι κάνουν 'like'** σε μηνύματα



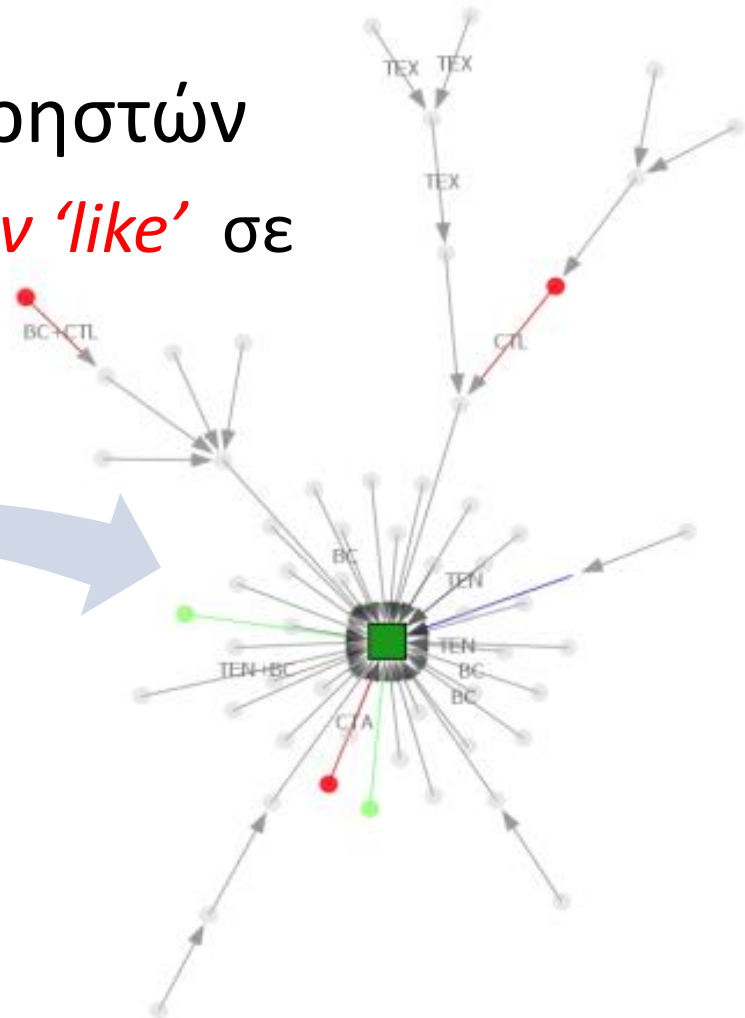
**Zakumaflex** Art of the new Sound 1 week ago  
damn he have so many songs  
Reply · 👍 👎

**anthony marcadet** 3 weeks ago  
respect +1 ! jadore  
Reply · 👍 👎 in playlist Panda Dub

**vammaanestein** 1 month ago  
 $\backslash x \{ (^ \_ \wedge ) x // \bullet \heartsuit \text{RESPECT} \heartsuit \bullet \backslash \{ (^ \_ \wedge ) \} //$   
Reply · 👍 👎

**lucas philibert** 1 month ago  
trop classe  
Reply · 👍 👎 in playlist Panda Dub

**MURPHY SLAW** 1 month ago  
There's no dub song quite like this one.  
Reply · 👍 👎 in playlist Irie Vibration



# Data management - II

Challenges faced by the community

# Types of Data We Have

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
- Social Network, Semantic Web (RDF), ...
- Streaming Data

# Working with such data

- Advanced features in established technologies
  - Recursive SQL for graph processing
- New storage and dedicated query systems
  - Cypher
- Aggregation and Statistics
  - Data warehousing and OLAP
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling