

ΜΥΕ003: Ανάκτηση Πληροφορίας

Διδάσκουσα: Ευαγγελία Πιτουρά

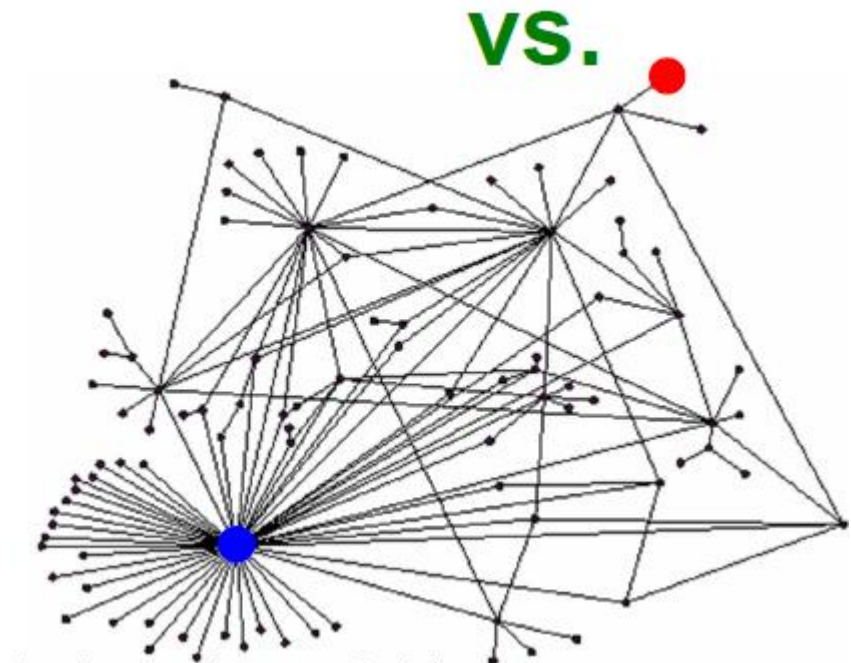
Κεφάλαιο 21: Ανάλυση Συνδέσμων.

Ακαδημαϊκό Έτος 2021-2022

Τι θα δούμε σήμερα

Πως μπορούμε να χρησιμοποιήσουμε το δίκτυο στη διάταξη των αποτελεσμάτων

Δεν είναι όλες οι σελίδες (κόμβοι) ίσες.
Ποιες σελίδες είναι «σημαντικές»?



Τι θα δούμε σήμερα

Ανάλυση συνδέσμων (link analysis)

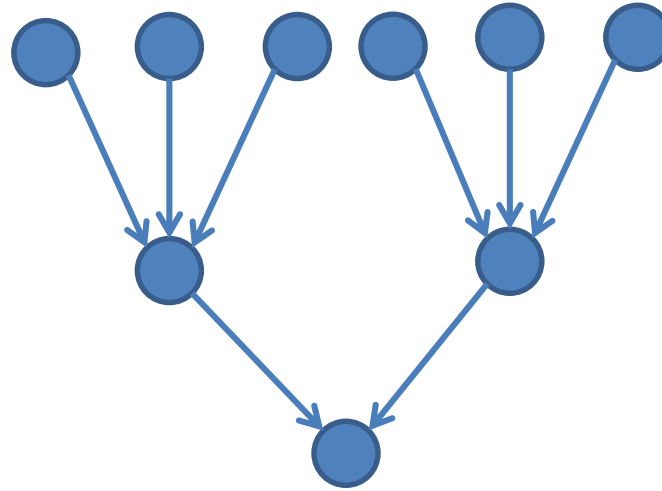
- PageRank

Υπολογισμός μιας τιμής ανά σελίδα (κόμβο) που εκφράζει το πόσο σημαντική είναι στο δίκτυο

Διάταξη με βάση τη δημοτικότητα

Διάταξη των σελίδων με βάση τον αριθμό των εισερχόμενων ακμών (**in-degree**, **degree centrality**)

Αρκεί η δημοτικότητα;



- Δεν είναι σημαντικό *πόσοι κόμβοι* δείχνουν σε μια σελίδα αλλά το *πόσο σημαντικοί* είναι αυτοί οι κόμβοι

PageRank

PageRank

- Βασική ιδέα: Μια σελίδα είναι σημαντική αν δείχνουν σε αυτήν σημαντικές σελίδες
- Η αξία (PageRank) ενός κόμβου είναι το **άθροισμα** της αξίας των φίλων του

PageRank: Βασική ιδέα

Έχουμε μια «μονάδα κύρους» που τη λέμε PageRank και την μοιράζουμε στις σελίδες.

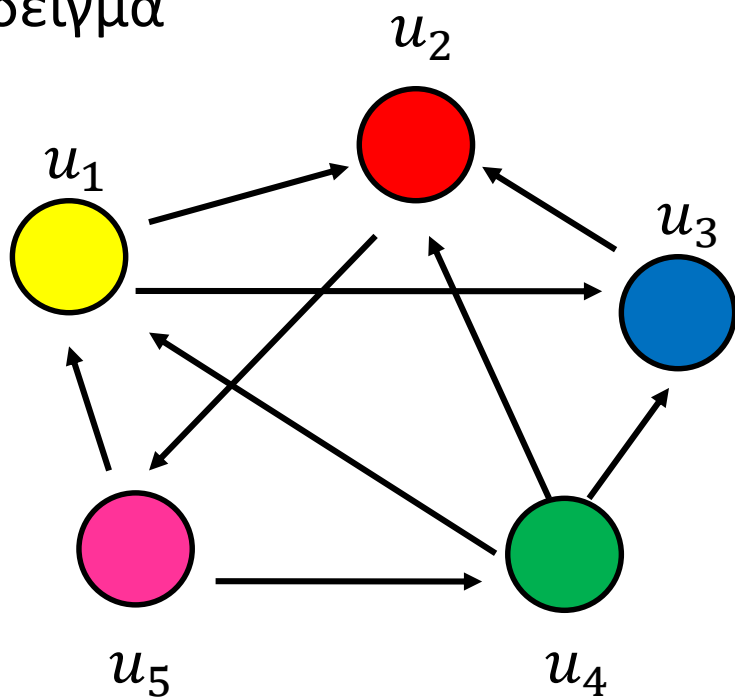
Κάθε σελίδα έχει μια τιμή PageRank

- Κάθε σελίδα *μοιράζει το PageRank της στις σελίδες που δείχνει*
- Το PageRank μιας σελίδας είναι το *άθροισμα των PageRank των σελίδων που δείχνουν σε αυτήν*

PageRank: Ορισμός

$$PageRank(v) = \sum_{u \in InNeighbors(v)} \frac{PageRank(u)}{outdegree(u)}$$

παράδειγμα



PageRank: παράδειγμα

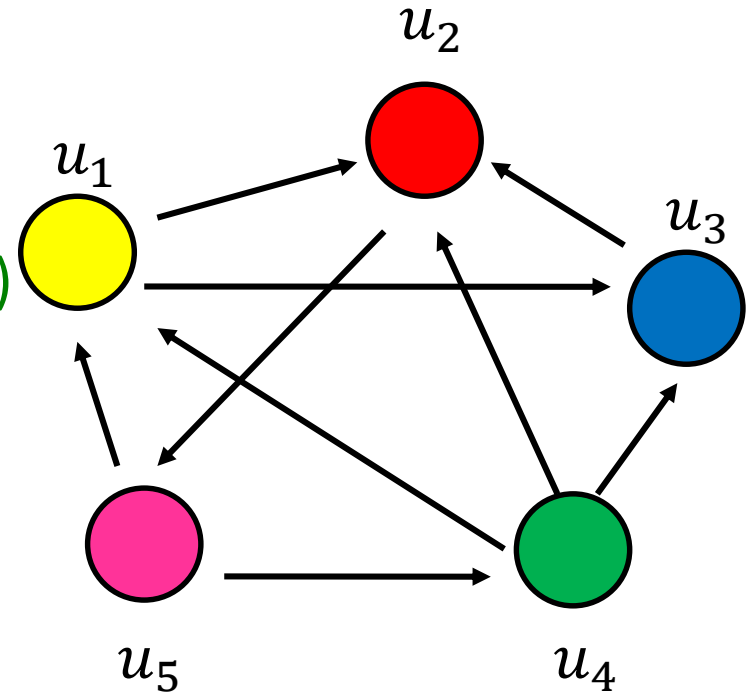
$$r(u_1) = 1/3 r(u_4) + 1/2 r(u_5)$$

$$r(u_2) = 1/2 r(u_1) + r(u_3) + 1/3 r(u_4)$$

$$r(u_3) = 1/2 r(u_1) + 1/3 r(u_4)$$

$$r(u_4) = 1/2 r(u_5)$$

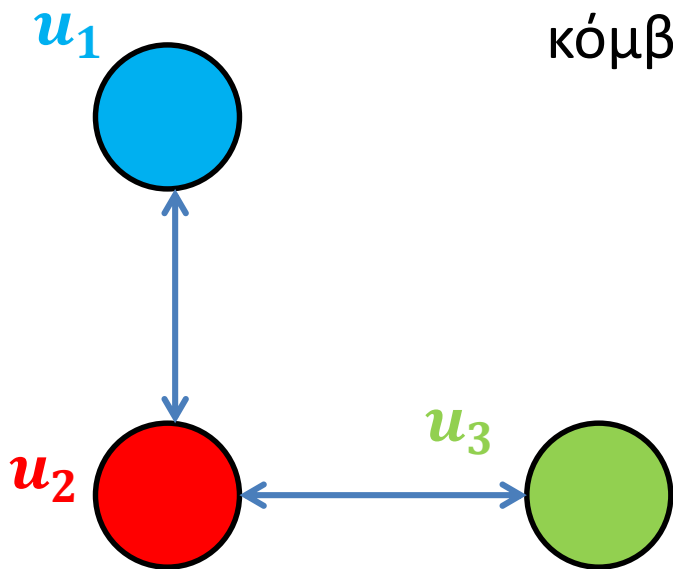
$$r(u_5) = r(u_2)$$



$$r(u_1) + r(u_2) + r(u_3) + r(u_4) + r(u_5) = 1 \quad r(u): \text{PageRank}(u)$$

Ένα απλό παράδειγμα: υπολογισμός

Το συνολικό PageRank μοιράζεται στους 3 κόμβους



$$r(u_1) + r(u_2) + r(u_3) = 1$$

$$r(u_1) = \frac{1}{2} r(u_2)$$

$$r(u_2) = r(u_1) + r(u_3)$$

$$r(u_3) = \frac{1}{2} r(u_2)$$

- Λύνοντας το σύστημα εξισώσεων παίρνουμε το PageRank των κόμβων

$$r(u_2) = \frac{1}{2} \quad r(u_1) = \frac{1}{4} \quad r(u_3) = \frac{1}{4}$$

PageRank: επαναληπτικός αλγόριθμος

Επαναληπτικός υπολογισμός (power iteration method)

Initialize $r^0(v) \leftarrow \frac{1}{n}$

$t = 1$

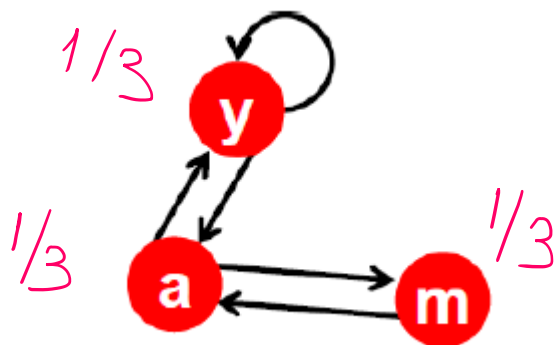
repeat

$$r^t(v) \leftarrow \sum_{u \rightarrow v} \frac{r^{t-1}(u)}{\text{outdegree}(u)}$$

$t = t + 1$

until convergence

Υπολογισμός: Παράδειγμα



y	1/3	1/3	5/12	
a	1/3	1/2	1/3	...
m	1/3	1/6	1/4	

$$r_y = r_y/2 + r_a/2$$

$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

στην πρώτη επαύριση

$$r_y^1 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}$$

$$r_a^1 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} = \frac{3}{6}$$

$$r_m^1 = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

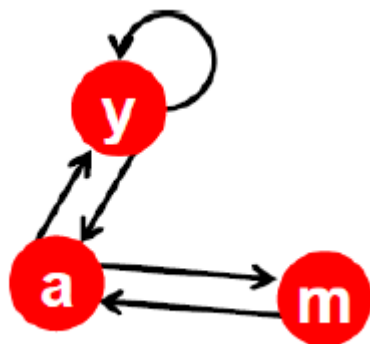
στη 2η επαύριση

$$r_y^2 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{12}$$

$$r_a^2 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{6} = \frac{1}{3}$$

$$r_m^2 = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

Υπολογισμός: Παράδειγμα



$$r_y = r_y/2 + r_a/2$$

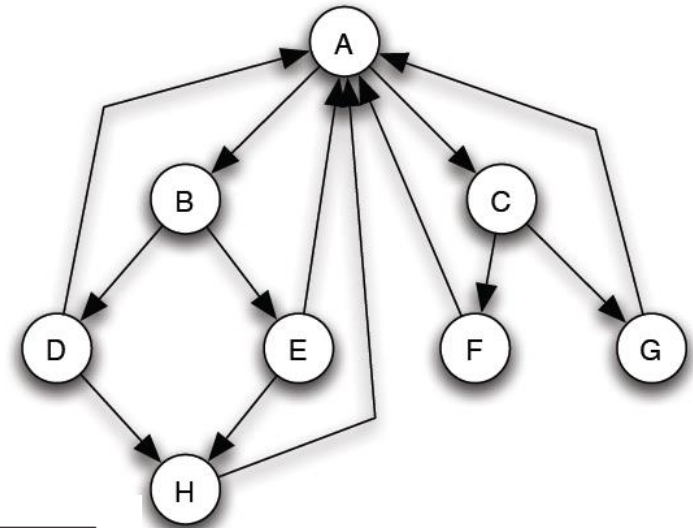
$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

$$\begin{pmatrix} r_y \\ r_a \\ r_m \end{pmatrix} = \begin{matrix} 1/3 & 1/3 & 5/12 & 9/24 & & 6/15 \\ 1/3 & 3/6 & 1/3 & 11/24 & \dots & 6/15 \\ 1/3 & 1/6 & 3/12 & 1/6 & & 3/15 \end{matrix}$$

Ένα μεγαλύτερο παράδειγμα

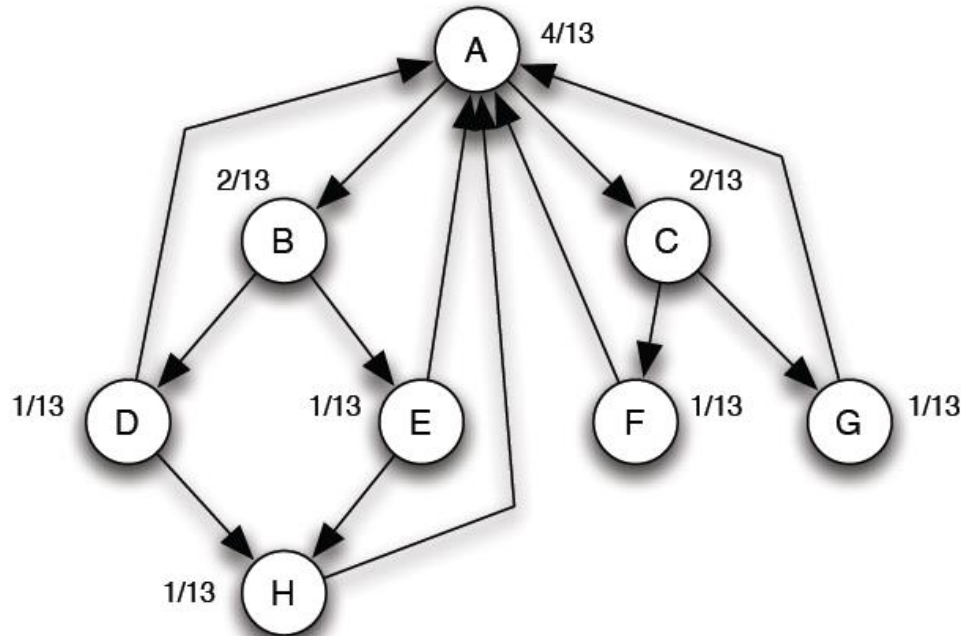
Αρχικά όλοι οι κόμβοι
PageRank $1/8$



Step	A	B	C	D	E	F	G	H
1	$1/2$	$1/16$	$1/16$	$1/16$	$1/16$	$1/16$	$1/16$	$1/8$
2	$3/16$	$1/4$	$1/4$	$1/32$	$1/32$	$1/32$	$1/32$	$1/16$

- Ένα είδος ροής (“fluid”) που κινείται στο δίκτυο
- Το συνολικό PageRank στο δίκτυο παραμένει σταθερό (δε χρειάζεται κανονικοποίηση)

Ισορροπία



- Ένας απλός τρόπος να ελέγξουμε αν το σύνολο PageRank τιμών αντιστοιχεί σε **ισορροπία**: οι τιμές αθροίζουν σε 1 και **δεν αλλάζουν** αν εφαρμόσουμε τον κανόνα ενημέρωσης

PageRank: Διανυσματική αναπαράσταση

Stochastic Adjacency Matrix – (Στοχαστικός) Πίνακας

Γειτνίασης M

Πίνακας M – πίνακας γειτνίασης του web

Αν $j \rightarrow i$, τότε $M_{ij} = 1/\text{outdegree}(j)$

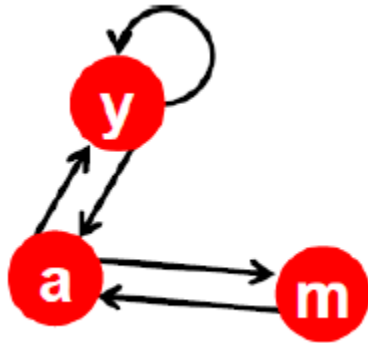
Αλλιώς, $M_{ij} = 0$

Page Rank Vector r

Ένα διάνυσμα με μία τιμή για κάθε σελίδα (το PageRank της σελίδας)

$$\sum r_i = 1$$

PageRank: Διανυσματική αναπαράσταση



	y	a	m
y	$\frac{1}{2}$	$\frac{1}{2}$	0
a	$\frac{1}{2}$	0	1
m	0	$\frac{1}{2}$	0

Column stochastic: οι τιμές στις στήλες αθροίζουν στο 1

$$\begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix}$$

$$\mathbf{r} = \mathbf{M} \cdot \mathbf{r}$$

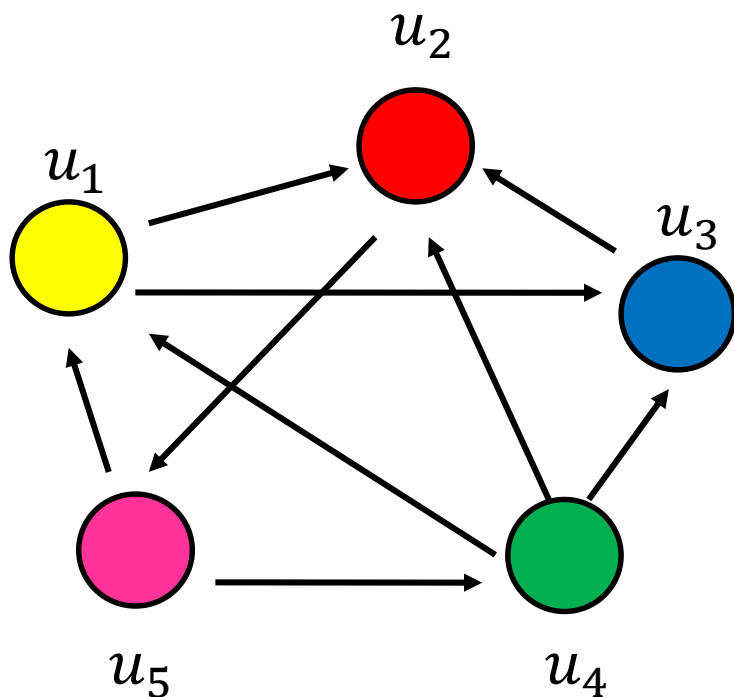
ιδιοδιάνυσμα, με ιδιοτιμή 1

Ισοδύναμα, r^T 1 x n vector (αριστερό ιδιοδιάνυσμα)

row stochastic adjacency matrix

$$r^T = r^T M^T$$

PageRank: παράδειγμα



$$A^T = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$M = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 1/2 \\ 1/2 & 0 & 1 & 1/3 & 0 \\ 1/2 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} r(u_1) \\ r(u_2) \\ r(u_3) \\ r(u_4) \\ r(u_5) \end{bmatrix}$$

$$r(u_1) = 1/3 r(u_4) + 1/2 r(u_5)$$

$$r(u_2) = 1/2 r(u_1) + r(u_3) + 1/3 r(u_4)$$

$$r(u_3) = 1/2 r(u_1) + 1/3 r(u_4)$$

$$r(u_4) = 1/2 r(u_5)$$

$$r(u_5) = r(u_2)$$

PageRank: Διανυσματική αναπαράσταση

$$r_j^{(t+1)} = \sum_{i \rightarrow j} \frac{r_i^{(t)}}{d_i} \quad r = Mr$$

- Ποια είναι η φυσική σημασία;
- Συγκλίνει ο επαναληπτικός αλγόριθμος;

Τυχαίος Περίπατος (Random Walks)

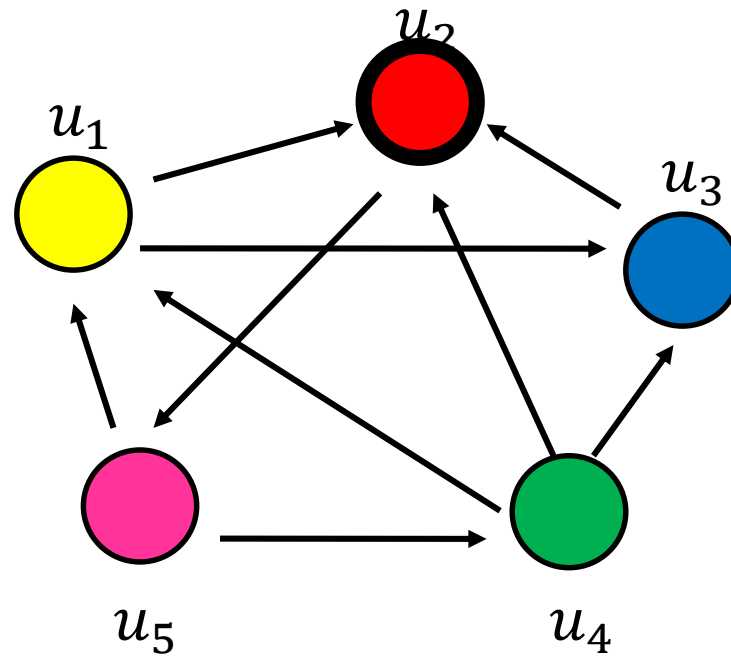
Ο αλγόριθμος προσομοιώνει ένα τυχαίο περίπατο (random walk) στο γράφο

Τυχαίος περίπατος:

- Ξεκίνα από κάποιον τυχαίο κόμβο (επιλεγμένο uniformly at random) με πιθανότητα $1/n$
- Επέλεξε τυχαία (uniformly at random) μια από τις εξερχόμενες ακμές του κόμβου
- Ακολούθησε την ακμή
- Επανάλαβε

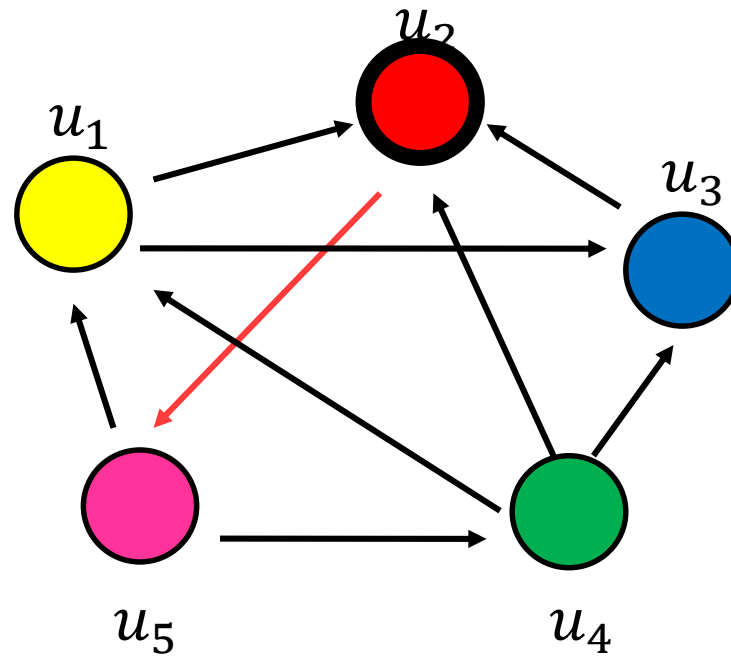
Παράδειγμα

- Step 0



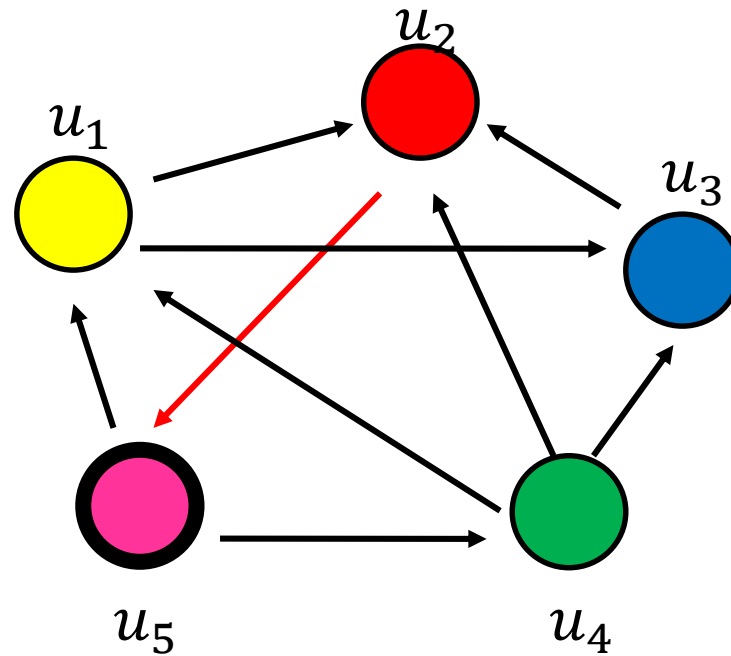
Παράδειγμα

- Step 0



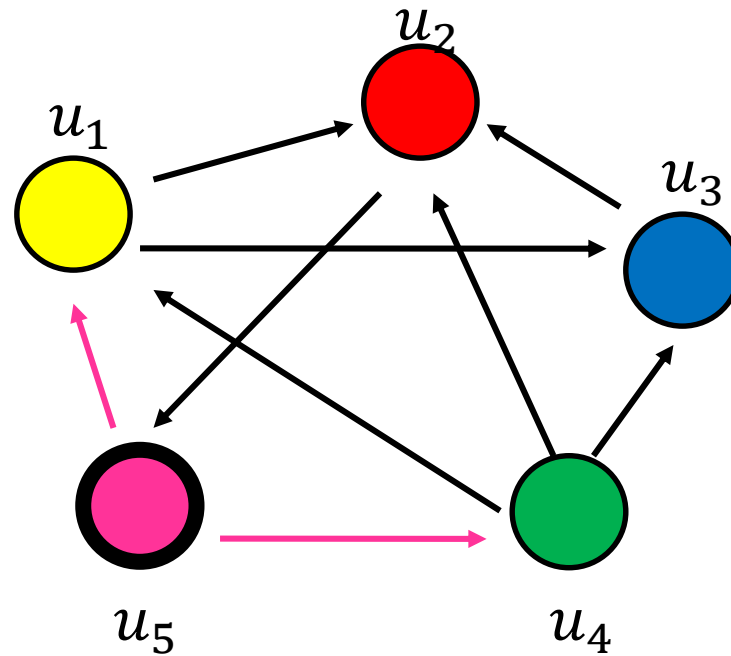
Παράδειγμα

- Step 1



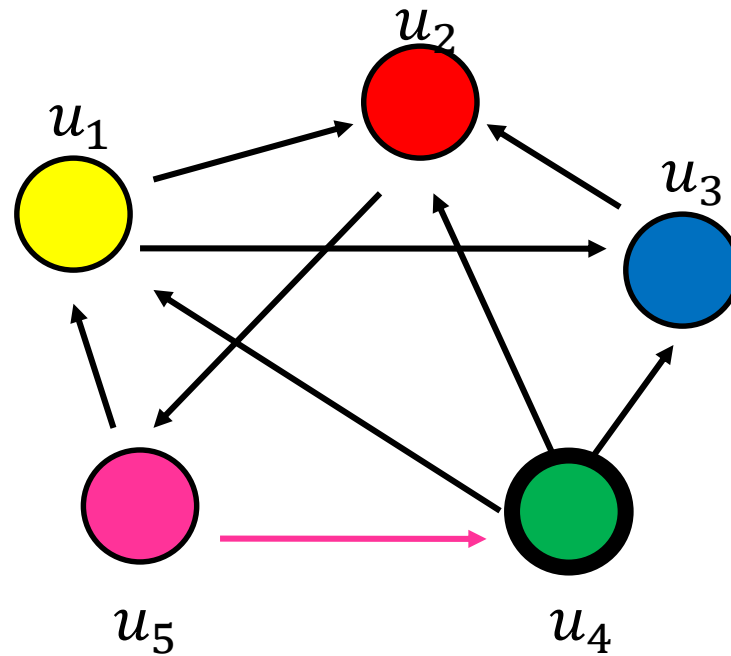
Παράδειγμα

- Step 1



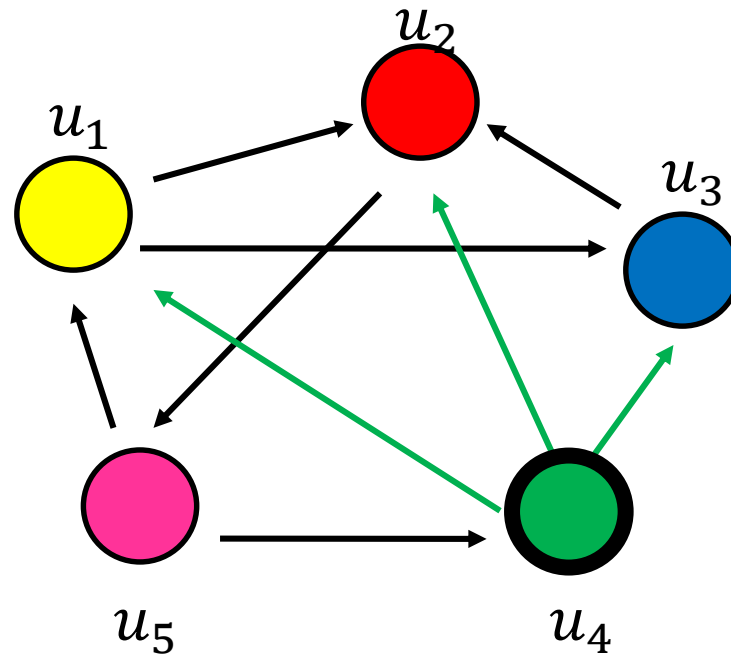
Παράδειγμα

- Step 2



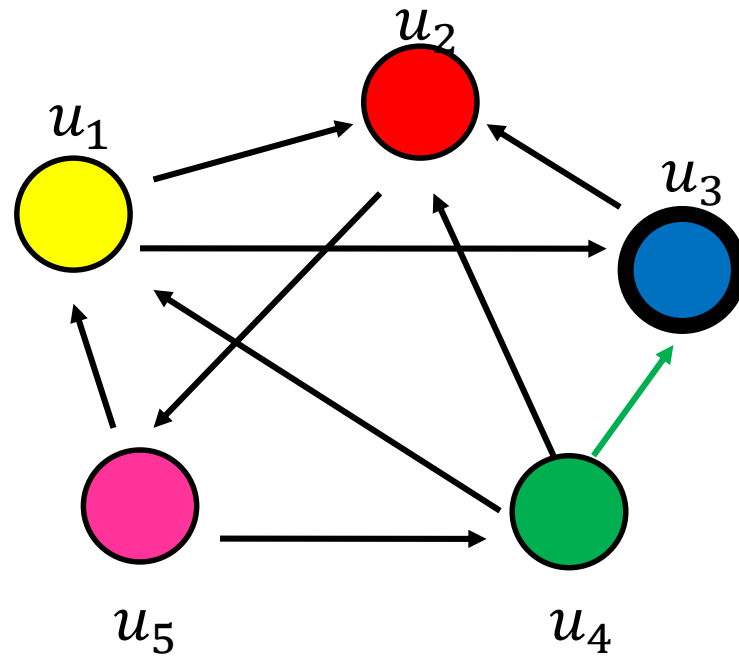
Παράδειγμα

- Step 2



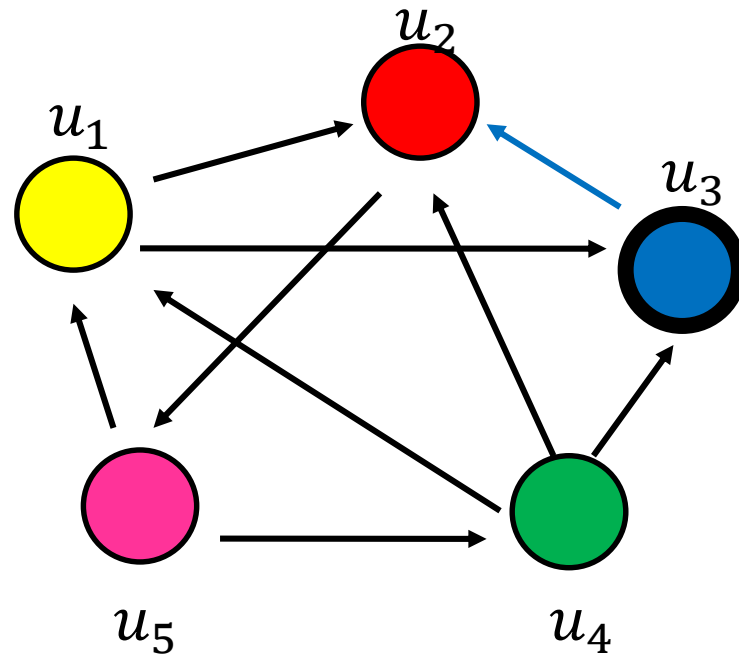
Παράδειγμα

- Step 3



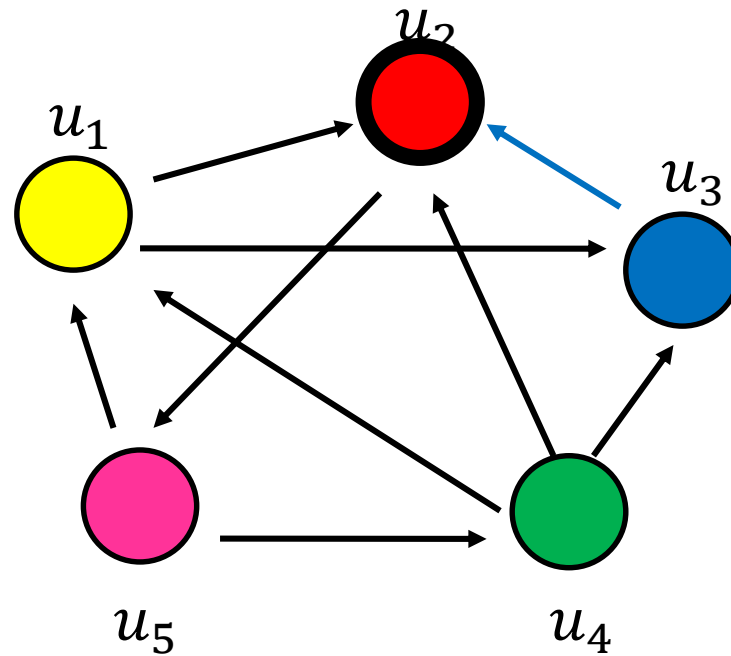
Παράδειγμα

- Step 3



Παράδειγμα

- Step 4...



Τυχαίος Περίπατος

Η πιθανότητα να είσαι στη σελίδα X μετά από k βήματα του τυχαίου περιπάτου είναι το $PageRank$ της σελίδας X μετά από k επαναλήψεις του υπολογισμού του $PageRank$

Το μοντέλο του **Random Surfer**

Του χρήστη που τριγυρνά στο web, ξεκινώντας από μια τυχαία σελίδα και συνεχίζει ακολουθώντας τυχαία κάποια από τις συνδέσεις της σελίδας

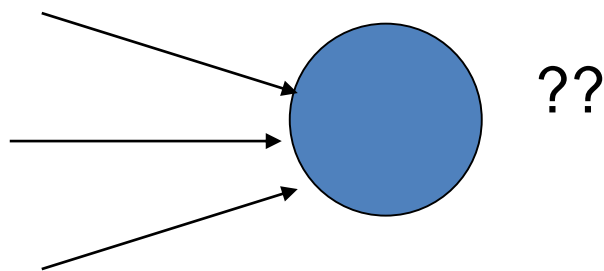
PageRank: Επεκτάσεις

Δύο προβλήματα

1. **Dead ends**: σελίδες χωρίς εξερχόμενες ακμές
Έχουν ως αποτέλεσμα να ξεφεύγει (leak out) to PageRank
2. **Spider traps**: Ομάδα σελίδων που όλες οι εξερχόμενες ακμές είναι μεταξύ τους
Τελικά απορροφούν όλο το PageRank

PageRank: Αδιέξοδα

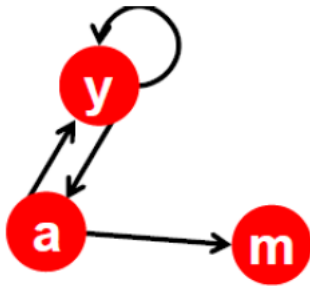
Αδιέξοδα (dead ends): σελίδες που δεν έχουν εξερχόμενες ακμές



Ο τυχαίος περίπατος μπορεί να κολλήσει σε ένα τέτοιο κόμβο

Λέγονται και **sink nodes**

PageRank: Αδιέξοδα



	y	a	m
y	1/2	1/2	0
a	1/2	0	0
m	0	1/2	0

$$\mathbf{r}_y = \mathbf{r}_y/2 + \mathbf{r}_a/2$$

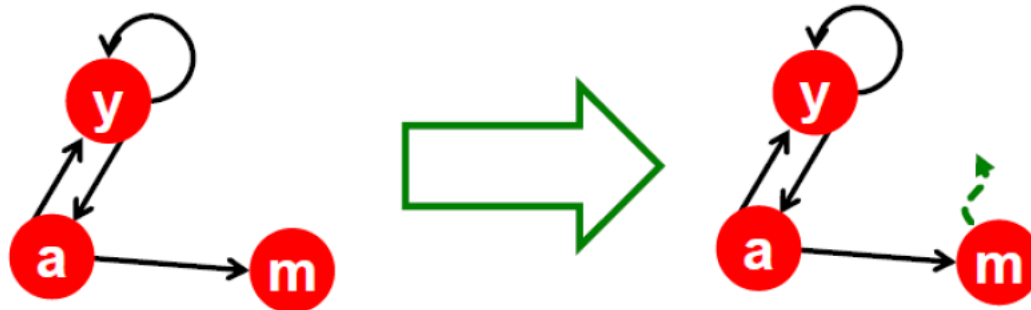
$$\mathbf{r}_a = \mathbf{r}_y/2$$

$$\mathbf{r}_m = \mathbf{r}_a/2$$

$$\begin{pmatrix} \mathbf{r}_y \\ \mathbf{r}_a \\ \mathbf{r}_m \end{pmatrix} = \begin{matrix} 1/3 & 2/6 & 3/12 & 5/24 & & 0 \\ 1/3 & 1/6 & 2/12 & 3/24 & \dots & 0 \\ 1/3 & 1/6 & 1/12 & 2/24 & & 0 \end{matrix}$$

PageRank: Αδιέξοδα

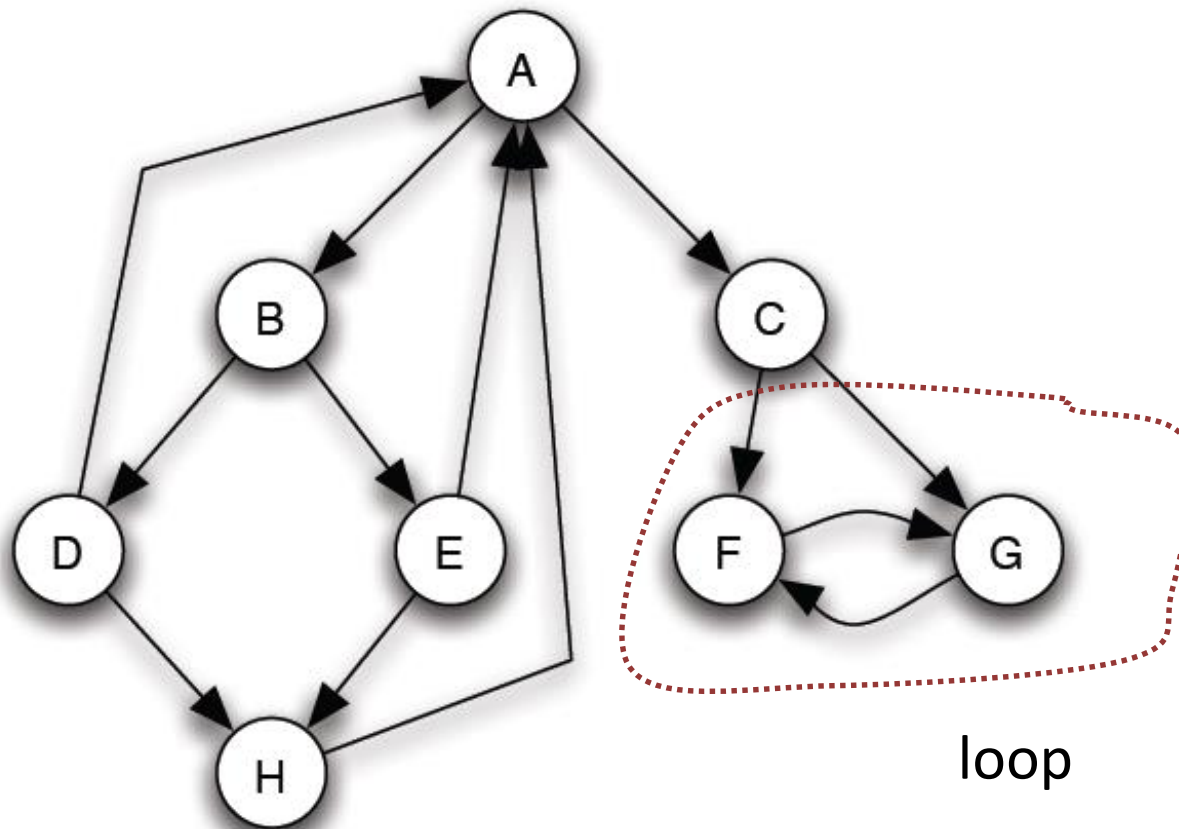
Teleports: ακολουθήσε με πιθανότητα 1 τυχαία links από τους αδιέξοδους κόμβους



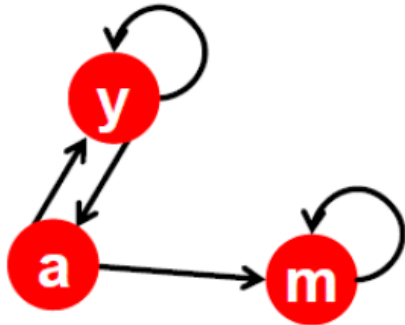
Αντίστοιχη τροποποίηση
του πίνακα

	y	a	m
y	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$
a	$\frac{1}{2}$	0	$\frac{1}{3}$
m	0	$\frac{1}{2}$	$\frac{1}{3}$

PageRank: Spider Traps



PageRank: Spider Traps



	y	a	m
y	$\frac{1}{2}$	$\frac{1}{2}$	0
a	$\frac{1}{2}$	0	0
m	0	$\frac{1}{2}$	1

$$r_y = r_y/2 + r_a/2$$

$$r_a = r_y/2$$

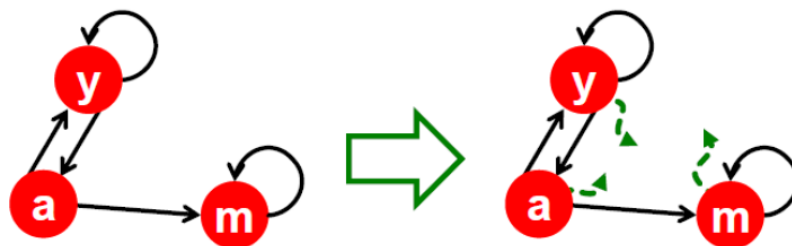
$$r_m = r_a/2 + r_m$$

$$\begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix} = \begin{matrix} 1/3 & 2/6 & 3/12 & 5/24 & & 0 \\ 1/3 & 1/6 & 2/12 & 3/24 & \dots & 0 \\ 1/3 & 3/6 & 7/12 & 16/24 & & 1 \end{matrix}$$

Random Walks with Jumps

Τυχαία περίπατοι με «άλματα»

Με πιθανότητα β , ο περιπατητής ακολουθεί μια τυχαία εξερχόμενη ακμή όπως πριν και με πιθανότητα $1-\beta$ επιλέγει (jumps) σε μια τυχαία σελίδα στο δίκτυο, επιλεγμένη με πιθανότητα $1/n$



Random Walks with Jumps

Γράφος με 3 κόμβους

Brin-Page, 1998

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{n} \quad \beta M + (1 - \beta) \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Προτεινόμενη τιμή **0,8 – 0,9**

damping factor: probability to jump $\sim 0,15$

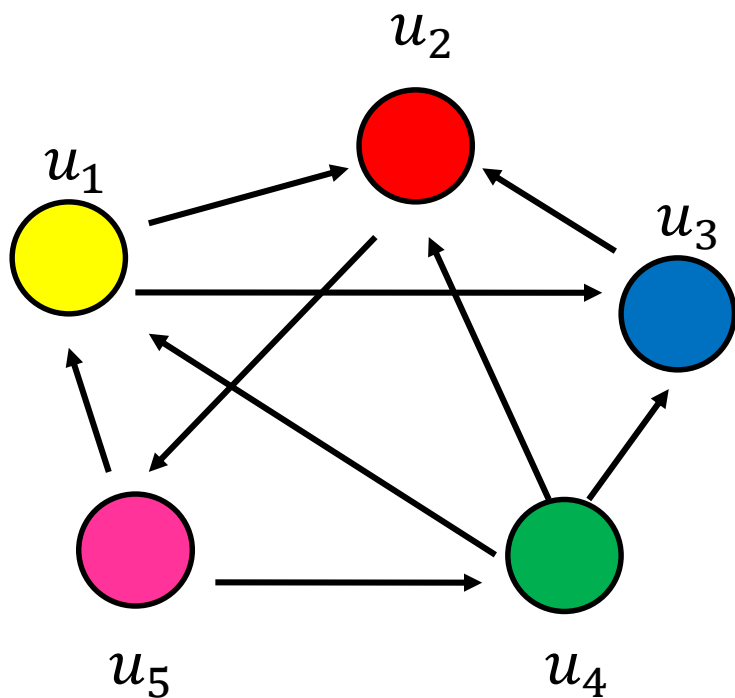
Μοντέλο του Random Surfer

Του χρήστη που τριγυρνά στο web, ξεκινώντας από μια τυχαία σελίδα συνεχίζει ακολουθώντας τυχαία συνδέσεις ή με κάποια πιθανότητα βαριέται και πάει (jumps) σε μια άλλη τυχαία σελίδα

5 links και **1 jump**

- Προσθέτουμε ένα τυχαίο άλμα σε ένα διάνυσμα v (jump vector) με πιθανότητα $1 - \beta$
 - συνήθως, το ομοιόμορφο διάνυσμα
 - β dumping factor
- Ο τυχαίος περίπατος ξαναρχίζει μετά από $1/(1 - \beta)$ βήματα in expectation

$$r = \beta M r + (1 - \beta) v^t$$



Πίνακας μετάβασης

$$M = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 1/2 \\ 1/2 & 0 & 1 & 1/3 & 0 \\ 1/2 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$r = \begin{bmatrix} r(u_1) \\ r(u_2) \\ r(u_3) \\ r(u_4) \\ r(u_5) \end{bmatrix}$$

$$r = Mr$$

$$r(u_1) = 1/3 r(u_4) + 1/2 r(u_5)$$

$$r(u_2) = 1/2 r(u_1) + r(u_3) + 1/3 r(u_4)$$

$$r(u_3) = 1/2 r(u_1) + 1/3 r(u_4)$$

$$r(u_4) = 1/2 r(u_5)$$

$$r(u_5) = r(u_2)$$

$$r = \beta M r + (1 - \beta) v^t$$

$$v = \begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix}$$

Νέος πίνακας
μετάβασης

$$P' = \beta \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{bmatrix} + (1 - \beta) \begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}$$

$$P' = \beta M + (1 - \beta) i_1 v^t \quad \text{όπου } i_1 \text{ το διάνυσμα με όλα 1}$$

PageRank, τυχαίοι περίπατοι και αλυσίδες Markov

Ποια είναι η πιθανότητα p_i^t να είσαι στον κόμβο i μετά από t βήματα;

$$p_1^0 = \frac{1}{5}$$

$$p_1^t = \frac{1}{3}p_4^{t-1} + \frac{1}{2}p_5^{t-1}$$

$$p_2^0 = \frac{1}{5}$$

$$p_2^t = \frac{1}{2}p_1^{t-1} + p_3^{t-1} + \frac{1}{3}p_4^{t-1}$$

$$p_3^0 = \frac{1}{5}$$

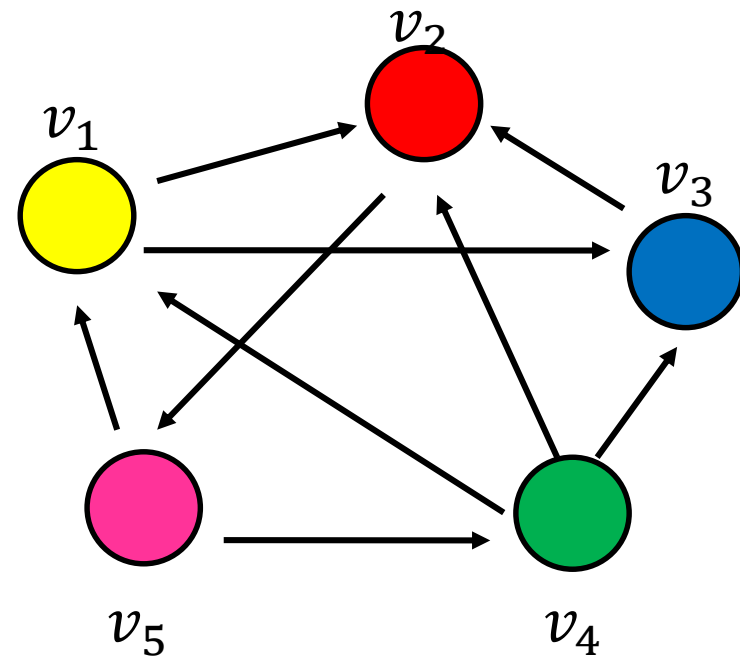
$$p_3^t = \frac{1}{2}p_1^{t-1} + \frac{1}{3}p_4^{t-1}$$

$$p_4^0 = \frac{1}{5}$$

$$p_4^t = \frac{1}{2}p_5^{t-1}$$

$$p_5^0 = \frac{1}{5}$$

$$p_5^t = p_2^{t-1}$$



Αλυσίδες Markov

- Περιγράφουν μια **στοχαστική διαδικασία διακριτού χρόνου** σε ένα σύνολο από καταστάσεις S
$$S = \{s_1, s_2, \dots, s_n\}$$
με βάση έναν πίνακα πιθανοτήτων μετάβασης (**transition probability matrix**) P
Όπου $P[i, j]$ είναι η πιθανότητα $s_i \rightarrow s_j$ να μεταβούμε στην κατάσταση s_j όταν είμαστε στην κατάσταση s_i
- Οι γραμμές αθροίζουν σε 1 (row stochastic)
- **Memoryless**: η επόμενη κατάσταση εξαρτάται μόνο από την τωρινή κατάσταση και όχι από τον παρελθόν της διαδικασίας

Αλυσίδες Markov

- State probability distribution vector

$$p^t = (p_1^t, p_2^t, \dots, p_n^t)$$

Διάνυσμα που αποθηκεύει την πιθανότητα να είμαστε στην κατάσταση p_i μετά από t βήματα

- Μπορούμε να το υπολογίσουμε ως:

$$p^t = P p^{t-1}$$

- Το state probability vector **συγκλίνει** σε μια μοναδική κατανομή αν η αλυσίδα είναι **μη περιοδική** (aperiodic) και **αμείωτη** (irreducible)

Αλυσίδες Markov

Irreducible: υπάρχει πάντα μια ακολουθία μεταβάσεων με μη μηδενική πιθανότητα από μια οποιαδήποτε κατάσταση σε μία άλλη (connectivity)

Aperiodicity: οι καταστάσεις δε μπορούν να χωριστούν σε σύνολα τέτοια ώστε όλες οι μεταβάσεις να συμβαίνουν κυκλικά από το ένα σύνολο στο άλλο

Τυχαίοι Περίπατοι

Οι τυχαίοι περίπατοι στους γράφους αντιστοιχούν σε Αλυσίδες Markov

- Το σύνολο των καταστάσεων S είναι οι κόμβοι του γράφου
- Ο πίνακας πιθανοτήτων μετάβασης είναι η πιθανότητα να ακολουθήσουμε μια ακμή από έναν κόμβο σε ένα άλλο

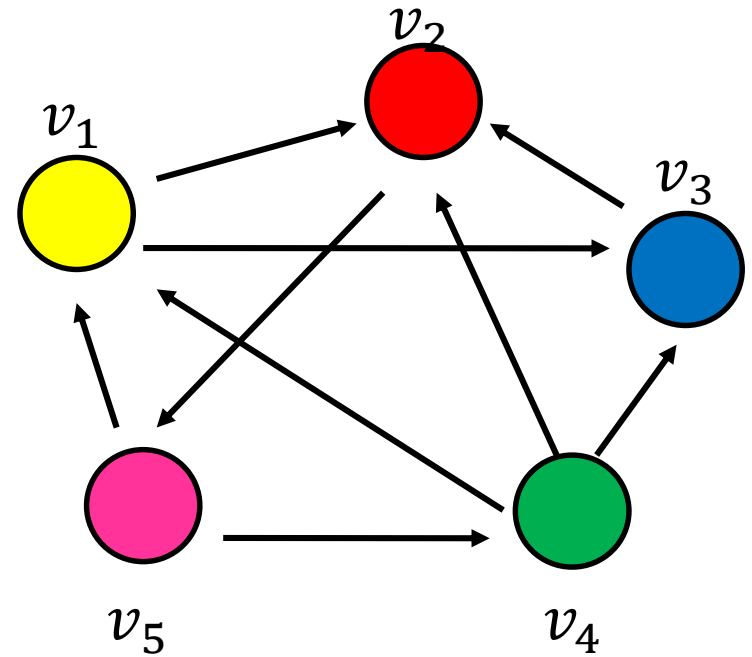
$$P[i, j] = 1/\text{outdegree}(i)$$

Στα επόμενα θα θεωρήσουμε τον ανάστροφο (transpose) του πίνακα M

Ένα παράδειγμα

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$



Το διάνυσμα πιθανοτήτων

$$p^t = (p_1^t, p_2^t, \dots, p_n^t)$$

Διάνυσμα που αποθηκεύει την πιθανότητα να είμαστε στον κόμβο u_i μετά από t βήματα

- p_i^0 πιθανότητα να αρχίσουμε από τον κόμβο i (συνήθως) ομοιόμορφη
- $p^t = P p^{t-1}$

Ένα παράδειγμα

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$

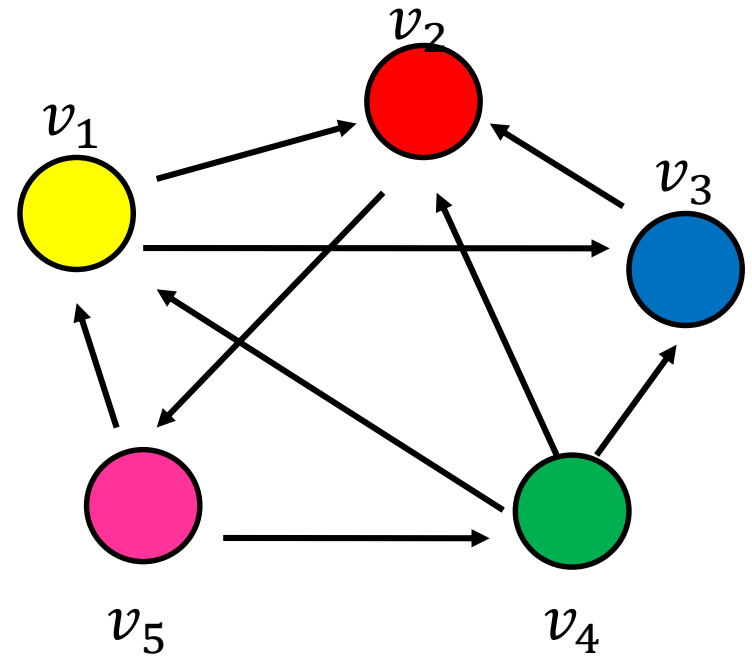
$$p_1^t = \frac{1}{3}p_4^{t-1} + \frac{1}{2}p_5^{t-1}$$

$$p_2^t = \frac{1}{2}p_1^{t-1} + p_3^{t-1} + \frac{1}{3}p_4^{t-1}$$

$$p_3^t = \frac{1}{2}p_1^{t-1} + \frac{1}{3}p_4^{t-1}$$

$$p_4^t = \frac{1}{2}p_5^{t-1}$$

$$p_5^t = p_2^{t-1}$$



Stationary distribution

- Η stationary κατανομή ενός τυχαίου περίπατου με πίνακα μετάβασης P είναι η κατανομή πιθανοτήτων π τέτοια ώστε
 - $\pi = \pi P$
- Το **ιδιοδιάνυσμα** (principal left eigenvector) του πίνακα P (οι στοχαστικοί πίνακες έχουν μέγιστη ιδιοτιμή 1)
- Το ποσοστό των φορών που επισκεπτόμαστε την κατάσταση (κόμβο) i όταν $t \rightarrow \infty$
- Θεωρία Αλυσίδων Markov: Ο τυχαίος περίπατος **συγκλίνει** σε μια μοναδική stationary distribution ανεξάρτητα από την αρχική κατάσταση αν ο γράφος είναι **ισχυρά συνεκτικός** και δεν είναι **διμερής**

Υπολογισμός

- Power Method

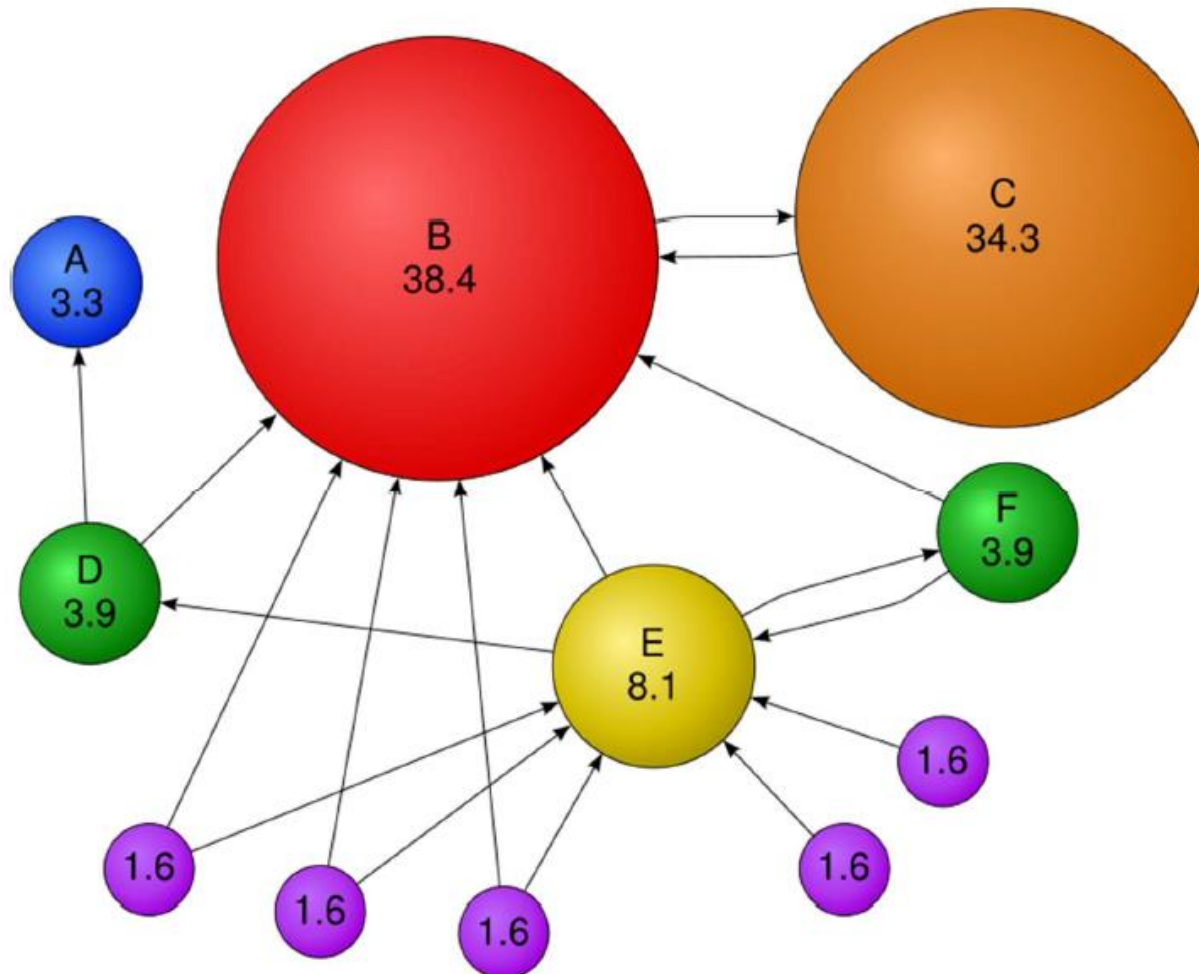
Initialize q^0 to some distribution
Repeat
 $q^t = q^{t-1}P$
Until convergence

- Μετά από πολλές επαναλήψεις $q^t \rightarrow \pi$ ανεξάρτητα από το αρχικό διάνυσμα q^0
- Power method γιατί υπολογίζει το $q^t = q^0 P^t$
- Ρυθμός σύγκλισης
 - Καθορίζεται από τη δεύτερη ιδιοτιμή λ_2^t

Stationary distribution

- Τι σημαίνει η stationary distribution π ενός τυχαίου περίπατου
- $\pi(i)$: η πιθανότητα να είμαστε στον κόμβο i μετά από ένα πολύ μεγάλο (άπειρο) αριθμό από βήματα
- $\pi = p_0 P^\infty$, όπου P ο πίνακας μετάβασης, p_0 το αρχικό διάνυσμα
 - $P(i, j)$: πιθανότητα μετάβασης από το i στο j σε ένα βήμα
 - $P^2(i, j)$: πιθανότητα μετάβασης από το i στο j σε δύο βήματα (πιθανότητα όλων των μονοπατιών μήκους 2)
 - $P^\infty(i, j) = \pi(j)$: πιθανότητα μετάβασης από το i στο j σε άπειρα βήματα – δεν έχει σημασία το αρχικό σημείο

PageRank: Παράδειγμα



PageRank: τυχαίος περίπατος

Αν το διάνυσμα v που γίνεται το jump δεν είναι uniform, τότε μια *προτίμηση* σε συγκεκριμένους κόμβους

Ονομάζεται και *τυχαίος περίπατος με επανεκκίνηση* (random walk with restart) - Personalized page rank – Rooted page rank

Κόμβοι σε «μικρή» απόσταση από τον «restart» κόμβο

Θεματικό PageRank

Η σημασία μιας σελίδας μόνο με βάση το δίκτυο – ανεξάρτητη από την ερώτηση (ή, το θέμα)

Πως μπορούμε να υπολογίσουμε «θεματικό» ή personalized PageRank;

Θεματικό PageRank

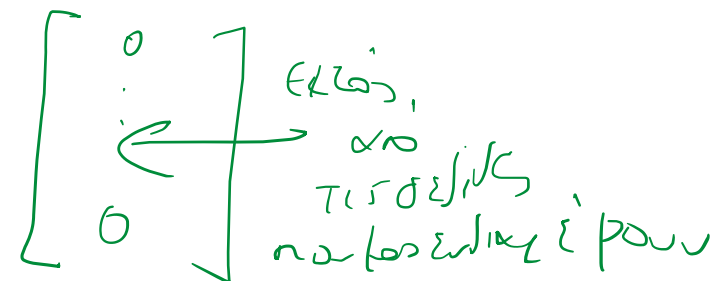
Έστω S ένα σύνολο από σελίδες «συναφείς» με το θέμα

Προσθέτουμε συνδέσεις teleports σε αυτές αντί σε τυχαίους κόμβους

$$M'_{ij} = (1 - \beta) M_{ij} + \beta / |S| \quad \text{if } i \in S$$

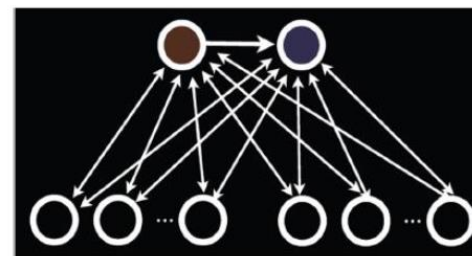
$$= (1 - \beta) M_{ij} \quad \text{otherwise}$$

Γειτνίαση με τις σελίδες στο σύνολο S



PageRank

Link Farms: δίκτυα από εκατομμύρια σελίδες που δείχνουν η μία στην άλλη με στόχο την αύξηση του PageRank κάποιων σελίδων



affordablecellphonerates.com

Search:

Results for "card phone prepaid":

- [Online prepaid phone card](#)
Over 95 prepaid calling cards sale Great quality instant ph
www.z.comm.com
- [US 1¢/min - World 2¢/min](#)
Use from Home, Office, Hotel & Cell Pre-Paid Card. Easy No
PennyTalk.com
- [Prepaid Phone Cards](#)
Your Guide To Consumer Electronics. Find Prepaid Phone C
GizmoCafe.com
- [Prepaid Phone](#)
Compare prepaid phones and plans. Boost Mobile® Officia
boostmobile.com
- [Prepaid Phone Cards](#)
Looking For Prepaid Phone Cards? See Our Prepaid Phone
kellyscornerstore.com
- [Phone Card](#)
Find Providers of Prepaid Phone Cards on the Business.co
www.business.com
- [Phone card](#)
Unlimited local and long distance Int'l rates as low as 14 p
www.village.com

Related Searches:

- [Free Prepaid Calling Card](#)
- [Retail](#)
- [International Call](#)
- [Internet Phone Card](#)
- [Calling Cards from To](#)
- [Calling Cards for India](#)
- [Cellular Phone Prepaid Phone Card](#)
- [Long Distance Card](#)
- [Cheap International Calling Cards](#)
- [Instant Calling Card Pin](#)
- [Calling Card Costa Rica](#)
- [South Africa Calling Card](#)
- [Buy a Calling Card](#)

PageRank: χρήση στην ανάκτηση

- Σελίδες με μεγάλο PageRank υψηλότερα στη διάταξη
- Τελικός βαθμός συνδυασμός πολλών χαρακτηριστικών (features)

ΤΕΛΟΣ 21^{ου} Κεφαλαίου

Ερωτήσεις?

Χρησιμοποιήθηκε κάποιο υλικό από:

- ✓ *Pandu Nayak and Prabhakar Raghavan, CS276:Information Retrieval and Web Search (Stanford)*
- ✓ *Hinrich Schütze and Christina Lioma, Stuttgart IIR class*
- ✓ *Τις αντίστοιχες διαλέξεις του μεταπτυχιακού μαθήματος «Κοινωνικά Δίκτυα και Μέσα»*