

3 - Ο αλγόριθμος *PageRank*

Δημήτριος Κοσμόπουλος

Πανεπιστήμιο Πατρών
Τμήμα Μηχανικών ΗΥ κ Πληροφορικής

7 Μαΐου 2025

Το γράφημα ως μητρώο

- ▶ Αντιμετωπίζοντας το γράφημα ως μητρώο μας επιτρέπει:
 - ▶ Να αποφασίσουμε την σημασία ενός κόμβου μέσω RW
 - ▶ Να υπολογίσουμε τις ενσωματώσεις κόμβων μέσω παραγοντοποίησης
 - ▶ Να θεωρήσουμε άλλες ενσωματώσεις ως αποτέλεσμα παραγοντοποίησης
- ▶ Τα RW, οι ενσωματώσεις και η παραγοντοποίηση είναι στενά συνδεδεμένα

Το διαδίκτυο ως γράφημα

- ▶ Οι ιστοσελίδες είναι οι κόμβοι
- ▶ Οι υπερσύνδεσμοι είναι οι ακμές
- ▶ Το γράφημα είναι κατευθυντικό
- ▶ Κάποιοι κόμβοι είναι σημαντικότεροι από άλλους

Μπορούμε να χρησιμοποιήσουμε τη δομή του γραφήματος για να ταξινομήσουμε τους κόμβους ανάλογα με τη σημασία τους.

Οι σύνδεσμοι ως ψήφοι

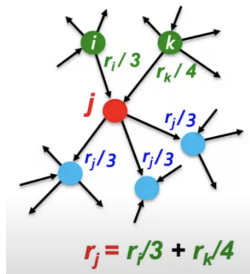
Ιδέα: οι σύνδεσμοι είναι ψήφοι.

- ▶ Μια σελίδα είναι πιο σημαντική αν έχει πολλούς συνδέσμους (εισερχόμενους; εξερχόμενους;)
- ▶ Σύνδεσμοι από σημαντικές σελίδες θα πρέπει να έχουν μεγαλύτερη βαρύτητα
- ▶ Υπάρχει αναδρομική σχέση

Το μοντέλο της ροής

Σύνδεσμοι από σημαντικές σελίδες θα πρέπει να έχουν μεγαλύτερη βαρύτητα

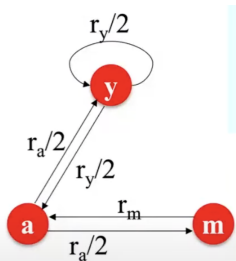
- ▶ καθε ψήφος είναι ανάλογη της σημασίας (βάρους) του κόμβου από τον οποίο προέρχεται
- ▶ για τον κόμβο i με βάρος r_i και εξερχόμενες d_i ακμές ο κάθε σύνδεσμος έχει βάρος $\frac{r_i}{d_i}$
- ▶ για τον κόμβο j το βάρος του r_j θα είναι το άθροισμα των βαρών των εισερχόμενων κόμβων



Το μοντέλο της ροής

Σύνδεσμοι από σημαντικές σελίδες θα πρέπει να έχουν μεγαλύτερη βαρύτητα

Το βάρος του κόμβου j δίνεται από: $r_j = \sum_{i \rightarrow j} \frac{r_i}{d_i}$



“Flow” equations:

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

$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

Αναπαράσταση με μητρώα

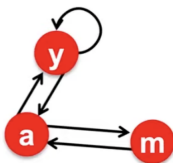
Το βάρος του κόμβου j δίνεται από: $r_j = \sum_{i \rightarrow j} \frac{r_i}{d_i}$

- ▶ και σε μορφή πίνακα:

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

- ▶ όπου $M_{i,j} = \frac{1}{d_j}$
- ▶ το άθροισμα κάθε στήλης είναι ίσο με 1

Αναπαράσταση με μητρώα



	r_y	r_a	r_m
r_y	$\frac{1}{2}$	$\frac{1}{2}$	0
r_a	$\frac{1}{2}$	0	1
r_m	0	$\frac{1}{2}$	0

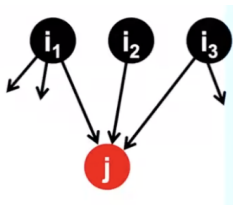
$$r_y = r_y/2 + r_a/2$$

$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

$$\begin{matrix} \boxed{\begin{matrix} r_y \\ r_a \\ r_m \end{matrix}} \\ \mathbf{r} \end{matrix} = \begin{matrix} \begin{matrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{matrix} \\ \mathbf{M} \end{matrix} \begin{matrix} \boxed{\begin{matrix} r_y \\ r_a \\ r_m \end{matrix}} \\ \mathbf{r} \end{matrix}$$

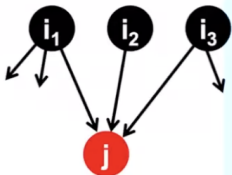
Συσχέτιση με RW



- ▶ Τη χρονική στιγμή t βρισκόμαστε στη σελίδα i
- ▶ Τη χρονική στιγμή $t + 1$ ακολουθούμε σύνδεσμο με ομοιόμορφη κατανομή
- ▶ Καταλήγουμε σε σελίδα j που συνδέεται με τη σελίδα i
- ▶ Η διαδικασία επαναλαμβάνεται

Θεωρούμε ότι $p_i(t)$ η πιθανότητα να βρισκόμαστε στη σελίδα i τη χρονική στιγμή t .

Στατική συνάρτηση πιθανότητας



Που βρισκόμαστε τη χρονική στιγμή $t + 1$;

$$\mathbf{p}(t + 1) = \mathbf{M} \cdot \mathbf{p}(t)$$

Αν φτάσουμε σε σημείο όπου $\mathbf{p}(t + 1) = \mathbf{M} \cdot \mathbf{p}(t)$ τότε η $\mathbf{p}(t)$ είναι στατική σ.π.π. του RW

Η σχέση

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

ικανοποιείται και το $\mathbf{r}$ εκφράζει την στατική σ.π.π. του RW .

Ιδιοδιανύσματα πίνακα γειτνίασης

- ▶ Το ιδιοδιάνυσμα ικανοποιεί τη σχέση
$$\mathbf{A} \cdot \mathbf{c} = \lambda \cdot \mathbf{c}$$
- ▶ ο ορισμός της κεντρικότητας βάσει ιδιοδιανύσματος για μη κατευθυντικό γράφημα
- ▶ Ο *PageRank* αφορά κατευθυντικό γράφημα
- ▶ Σύγκλιση *RW* στατικής σ.π.π., αναπαράστασης ροής, αναπαράστασης ιδιοδιανύσματος

Αναπαράσταση ιδιοδιανύσματος

- ▶ Η εξίσωση ροής μπορεί να γραφεί

$$1 \cdot \mathbf{r} = \mathbf{M} \cdot \mathbf{r}$$

- ▶ Άρα το διάνυσμα $\mathbf{r}$ είναι ιδιοδιάνυσμα του πίνακα $\mathbf{M}$ και αντιστοιχεί στην ιδιοτιμή 1
- ▶ Ξεκινώντας από οποιοδήποτε διάνυσμα $\mathbf{u}$ το όριο $\mathbf{M}(\mathbf{M}(\mathbf{M} \dots \mathbf{M}\mathbf{u}))$ δίνει εξ ορισμού τη σ.π.π. για $t \rightarrow \infty$
- ▶ Θυμηθείτε τα μονοπάτια άπειρου μήκους και το δείκτη *Katz*
- ▶ Ο *PageRank* δίνει την στατική σ.π.π. του *RW*, που αντιστοιχεί στο ιδιοδιάνυσμα ιδιοτιμής 1

Υπολογισμός *PageRank*

1. Ανάθεσε σε κάθε ένα από τους N κόμβους μια τιμή r_j^0
2. Υπολόγισε

$$r_j^{t+1} = \sum_{i \rightarrow j} \frac{r_i^t}{d_i}$$

όπου d_i είναι ο αριθμός των εξερχόμενων ακμών.

3. Επανάλαβε μέχρι να ισχύσει η συνθήκη: $\sum_i |r_i^{t+1} - r_i^t| < \epsilon$

Υπολογισμός *PageRank*

Σε διανυσματική μορφή:

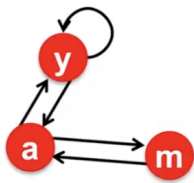
1. Ανάθεσε τιμή $\mathbf{r}^0 = [\frac{1}{N}, \dots, \frac{1}{N}]$
2. Υπολόγισε

$$\mathbf{r}^{t+1} = \mathbf{M}\mathbf{r}^t$$

3. Επανάλαβε μέχρι να ισχύσει η συνθήκη: $|\mathbf{r}^{t+1} - \mathbf{r}^t| < \epsilon$

Συγκλίνει για περίπου 50 επαναλήψεις.

Παράδειγμα υπολογισμού



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

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

$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

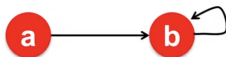
$$\begin{pmatrix} r_y \\ r_a \\ r_m \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \begin{pmatrix} 1/3 \\ 3/6 \\ 1/6 \end{pmatrix} \begin{pmatrix} 5/12 \\ 1/3 \\ 3/12 \end{pmatrix} \begin{pmatrix} 9/24 \\ 11/24 \\ 1/6 \end{pmatrix} \dots \begin{pmatrix} 6/15 \\ 6/15 \\ 3/15 \end{pmatrix}$$

Iteration 0, 1, 2, ...

Προβλήματα

Η παγίδα της αράχνης (spider trap)

$$r_j^{t+1} = \sum_{i \rightarrow j} \frac{r_i^t}{d_i}$$



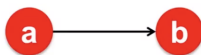
Iteration:		0,	1,	2,	3...
r_a	=	1	0	0	0
r_b		0	1	1	1

Δεν βλάπτει την μαθηματική αναπαράσταση, αλλά δεν δίνει το αποτέλεσμα που θέλουμε.

Προβλήματα

Το αδιέξοδο (dead end)

$$r_j^{t+1} = \sum_{i \rightarrow j} \frac{r_i^t}{d_i}$$



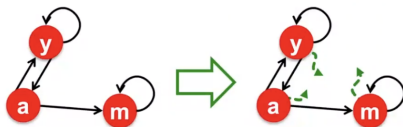
Iteration: 0,		1,	2,	3...	
r_a	=	1	0	0	0
r_b		0	1	0	0

Βλάπτει τη μαθηματική αναπαράσταση: η στήλη του πίνακα πιθανοτήτων δεν αθροίζει πια στο 1 και μηδενίζεται.

Λύση στο πρόβλημα παγίδας της αράχνης

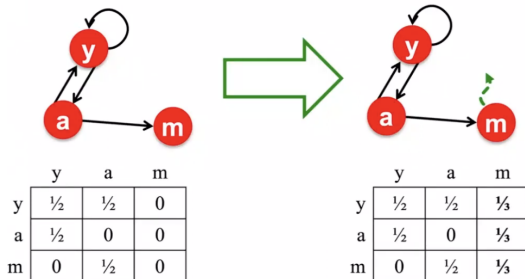
Σε κάθε βήμα έχουμε 2 επιλογές

1. Με πιθανότητα β ακολουθήσε ένα τυχαίο σύνδεσμο
 2. Με πιθανότητα $1-\beta$ διακτίνισου σε τυχαίο κόμβο
- ▶ Τυπική τιμή $\beta=0.9$
 - ▶ Η διακτίνιση από την παγίδα γίνεται σε λίγα βήματα.



Λύση στο πρόβλημα αδιεξόδου

- ▶ Διακτίνιση σε κόμβο με ομοιόμορφη κατανομή πιθανότητας αν φτάσεις σε αδιέξοδο.
- ▶ Αναπροσαρμογή πίνακα



Συνολική λύση

Σε κάθε βήμα έχουμε 2 επιλογές:

1. Με πιθανότητα β ακολουθήσε ένα τυχαίο σύνδεσμο
2. Με πιθανότητα $1-\beta$ διακτινίσου σε τυχαίο κόμβο

$$r_j^{t+1} = \beta \sum_{i \rightarrow j} \frac{r_i^t}{d_i} + (1 - \beta) \frac{1}{N}$$

Η αναπαράσταση υποθέτει ότι δεν υπάρχουν αδιέξοδα και ότι έχει ήδη προσαρμοστεί ο πίνακας M .

Συνολική λύση

- ▶ Η εξίσωση *PageRank*:

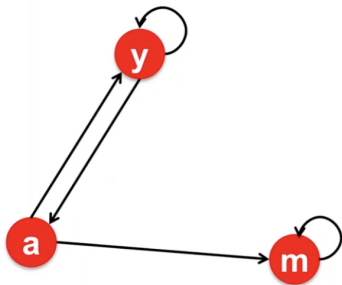
$$r_j^{t+1} = \beta \sum_{i \rightarrow j} \frac{r_i^t}{d_i} + (1 - \beta) \frac{1}{N}$$

- ▶ Ο πίνακας *Google* $\mathbf{G} = \beta \mathbf{M} + (1 - \beta) [\frac{1}{N}]_{N \times N}$
- ▶ Έχουμε αναδρομικό πρόβλημα:

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

- ▶ $\beta \approx 0.9$

Παράδειγμα



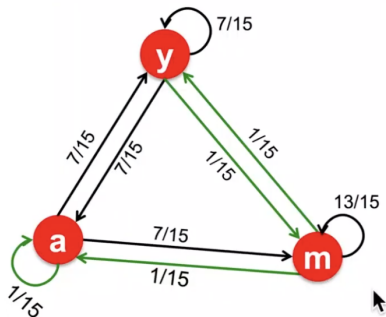
M

1/2	1/2	0
1/2	0	0
0	1/2	1

$[1/N]_{N \times N}$

$$\begin{array}{l} y \\ a \\ m \end{array} = \begin{array}{l} 1/3 \\ 1/3 \\ 1/3 \end{array}$$

Παράδειγμα



$$\begin{array}{c} \mathbf{M} \\ 0.8 \end{array} \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 1 \end{bmatrix} + 0.2 \begin{array}{c} [1/N]_{N \times N} \\ \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \mathbf{G} \\ \begin{bmatrix} y & 7/15 & 7/15 & 1/15 \\ a & 7/15 & 1/15 & 1/15 \\ m & 1/15 & 7/15 & 13/15 \end{bmatrix} \end{array}$$

y		1/3	0.33	0.24	0.26		7/33
a	=	1/3	0.20	0.20	0.18	...	5/33
m		1/3	0.46	0.52	0.56		21/33

Παράδειγμα

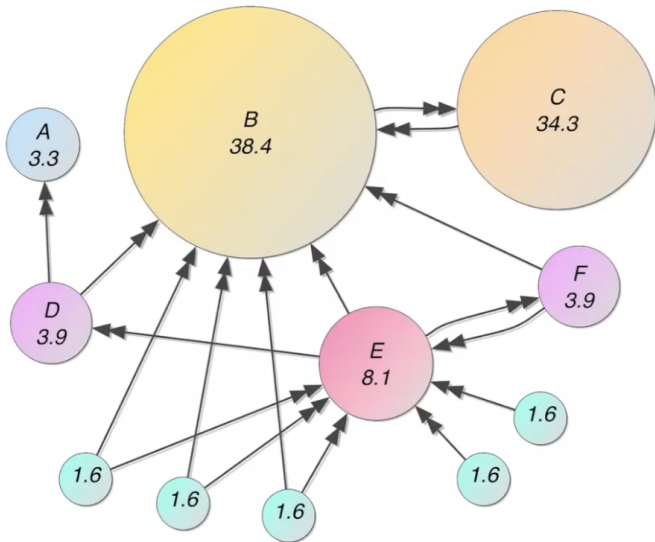


Image credit: [Wikipedia](#)

Βιβλιογραφία

1. Brin, S.; Page, L. (1998). "The anatomy of a large-scale hypertextual Web search engine" (PDF). Computer Networks and ISDN Systems. 30 (1–7): 107–117.
2. J. Leskovec, Machine Learning with Graphs, Stanford University, Fall 2023