

# Αλγόριθμος Pagerank με παράλληλη υλοποίηση της μεθόδου Gauss-Seidel

Νικόλαος Κατωμέρης

AEM: 8551

ngkatomer@auth.gr

Παράλληλα και διανεμημένα συστήματα.

Τμήμα ηλεκτρολόγων μηχανικών και μηχανικών υπολογιστών, ΑΠΘ

## Περίληψη

Στην παρούσα αναφορά περιγράφεται η διαδικασία παραλληλοποίησης του αλγορίθμου Gauss-Seidel για τον υπολογισμό του διανύσματος Pagerank γράφων του διαδικτύου. Επίσης, παρουσιάζονται αποτελέσματα σε ενδεικτικούς γράφους καθώς και αποδείξεις ορθότητας του υλοποιημένου προγράμματος.

## Pagerank

Η βασική ιδέα του αλγορίθμου Pagerank είναι η αξιολόγηση της κάθε σελίδας του διαδικτύου με βάση τον αριθμό των συνδέσμων προς αυτήν και την αξιολόγηση των σελίδων που τους περιέχουν.

Το αποτέλεσμα της παραπάνω αξιολόγησης ταυτίζεται με το ποσοστό του χρόνου στον οποίο ένας τυχαίος περιπατητής θα βρισκόταν στην εκάστοτε σελίδα, αν ξεκινώντας από κάποια τυχαία σελίδα, ακολουθούσε τυχαία κάποιον από τους συνδέσμούς της, στη συνέχεια της επόμενης κι ούτω καθ' εξής.

Η παραπάνω διαδικασία μπορεί να προσομοιωθεί με μία αλυσίδα Markov.

Έστω  $A$ , ο  $N \times N$  πίνακας μεταβάσεων της αλυσίδας για  $N$  σελίδες του διαδικτύου. Αν  $j \rightarrow i$  σημαίνει ότι η σελίδα  $j$  περιέχει σύνδεσμο προς τη σελίδα  $i$ , ο  $A$  ορίζεται αρχικά ως εξής:

$$A(i, j) = \begin{cases} 1, & \text{αν } j \rightarrow i \\ 0, & \text{αλλιώς} \end{cases}$$

Επειδή όμως ο παραπάνω πίνακας θέλουμε να εκφράζει την πιθανότητα ο περιπατητής να πάει από τη σελίδα  $j$  στη σελίδα  $i$  σε ένα βήμα, είναι απαραίτητο το άθροισμα των στοιχείων

κάθε στήλης του πίνακα να ισούται με 1.

Οπότε, αφού λάβουμε τον πίνακα με τους συνδέσμούς των σελίδων, διαιρούμε όλα του τα στοιχεία με το άθροισμα των στοιχείων της εκάστοτε στήλης  $j$ , έστω  $L_j$ , και παίρνουμε τον «στοχαστικοποιημένο» πίνακα  $A_s$ :

$$A_s(i, j) = \begin{cases} \frac{1}{L_j}, & \text{αν } j \rightarrow i \\ 0, & \text{αλλιώς} \end{cases}$$

Επιπλέον, καθώς ο περιπατητής ξεκινά από μία σελίδα, είναι πολύ πιθανό, ακολουθώντας συνεχώς συνδέσμούς να είναι αδύνατο να μεταβεί σε όλες τις σελίδες του γράφου. Το γεγονός αυτό, συν το γεγονός ότι ένας χρήστης του διαδικτύου δεν ακολουθεί μόνο συνδέσμούς, αλλά «τηλεμεταφέρεται» κιόλας σε άλλες σελίδες, καθιστά τον παραπάνω πίνακα ανεπαρκή για την προσομοίωση της τυχαίας διαδρομής ενός χρήστη του διαδικτύου.

Έτσι, εισάγεται στον αλγόριθμο μια σταθερά «τηλεμεταφοράς» που δηλώνει την πιθανότητα ο περιπατητής να μεταβεί σε κάποια τυχαία σελίδα του διαδικτύου χωρίς να υπάρχει κάποιος σύνδεσμος προς αυτήν στη σελίδα που βρίσκεται.

Η σταθερά αυτή, έστω  $b$ , μπορεί να επιλε-

γείως όρισμα στο πρόγραμμα υπολογισμού του διανύσματος pagerank που υλοποιήθηκε. Στην περίπτωση που δεν δοθεί, θα έχει την τιμή 0.15 που έχει δειχθεί ότι αποτελεί καλή επιλογή [1].

Με την εισαγωγή της παραπάνω σταθεράς, ο πίνακας μεταβάσεων πλέον θα είναι:

$$A_{final} = \frac{b}{N} \cdot \mathbf{1}_{N \times N} + (1 - b)A_s$$

Στην παραπάνω σχέση, η σταθερά  $b$  διαιρείται με τον αριθμό των γραμμών  $N$  έτσι ώστε οι στήλες και του τελικού πίνακα  $A_{final}$  να έχουν στοιχεία με άθροισμα 1.

Με τη χρήση του παραπάνω πίνακα είναι δυνατό να υπολογιστεί το διάνυσμα  $N \times 1$  με την αξιολόγηση της κάθε σελίδας, πολλαπλασιάζοντάς τον με ένα αρχικό διάνυσμα -εδώ ένα ομοιόμορφο- αρκετές φορές, έως ότου το αποτέλεσμα συγκλίνει. Το διάνυσμα αυτό αποτελεί ιδιοδιάνυσμα του πίνακα  $A$  για την ιδιοτιμή 1.

Για να συγκλίνει με βεβαιότητα το διάνυσμα απαιτείται ακόμη ένα βήμα, θα πρέπει, αν ο περιπατητής βρεθεί σε κάποια σελίδα που δεν περιέχει καμία σύνδεση, να τηλεμεταφερθεί σε μία τυχαία σελίδα με βάση ένα διάνυσμα πιθανοτήτων. Στην υλοποίηση αυτή, ο περιπατητής έχει την ίδια πιθανότητα να τηλεμεταφερθεί σε οποιαδήποτε σελίδα.

Με όλα τα παραπάνω δεδομένα, αποδεικνύεται ότι ο τελικός πίνακας θα έχει μία ιδιοτιμή ένα και το πρόβλημα πλέον έγκειται στην εύρεση του ιδιοδιανύσματος της.

Για την εύρεση του ιδιοδιανύσματος, όπως ζητείται, χρησιμοποιείται η μέθοδος Gauss-Seidel.

## Gauss Seidel

Η μέθοδος Gauss-Seidel είναι μια αναδρομική μέθοδος επίλυσης γραμμικών συστημάτων. Ο αναδρομικός αλγόριθμός της μεθόδου για την επίλυση ενός γραμμικού συστήματος

της μορφής:

$$A\vec{x} = \vec{b}$$

είναι για το  $i$ -στο στοιχείο στο  $k + 1$  βήμα:

$$x_i(k+1) = \frac{1}{A_{ii}} \left( b_i - \sum_{j=1}^{i-1} A_{ij}x_j(k+1) - \sum_{j=i+1}^N A_{ij}x_j(k) \right)$$

Όπως φαίνεται, ο αλγόριθμος αυτός μοιάζει καθαρά σειριακός αφού ο υπολογισμός του κάθε στοιχείου σε κάθε βήμα εξαρτάται από τον υπολογισμό των νέων τιμών των στοιχείων πριν απ' αυτό. Για περισσότερες λεπτομέρειες σχετικά με τη μέθοδο αυτή και τα κριτήρια σύγκλισης της ανατρέξτε στον Saad [2].

Στην περίπτωσή μας, για την εύρεση της pagerank με τη μέθοδο Gauss-Seidel, θα πρέπει πρώτα να το εκφράσουμε σαν σύστημα γραμμικών εξισώσεων. Όπως προαναφέρθηκε το διάνυσμα Pagerank αποτελεί ιδιοδιάνυσμα του πίνακα  $A$  για την ιδιοτιμή 1. Επομένως, ισχύει ότι:

$$\begin{aligned} A_{final}\vec{x} &= \vec{x} \Rightarrow (I - A_{final})\vec{x} = 0 \\ \Rightarrow \left( I - \frac{b}{N} \cdot \mathbf{1}_{N \times N} - (1 - b)A_s \right) \vec{x} &= 0 \end{aligned}$$

Όμως, το  $\mathbf{1}_{N \times N} \cdot \vec{x}$  είναι πάντα ίσο με  $\mathbf{1}_{N \times 1}$  επειδή το άθροισμα των στοιχείων του  $\vec{x}$ , ως άθροισμα όλων των πιθανοτήτων, εκφράζει την πιθανότητα ο περιπατητής να βρίσκεται σε μια οποιαδήποτε σελίδα. Έτσι, το σύστημά μας παίρνει τη μορφή:

$$(I - (1 - b)A_s)\vec{x} = \frac{b}{N}\mathbf{1}_{N \times 1}$$

που μπορεί να λυθεί αναδρομικά με τη μέθοδο Gauss-Seidel, αρκεί διασφαλιστεί ότι σε κάθε βήμα  $\mathbf{1}_{N \times N} \cdot \vec{x}_k = \mathbf{1}_{N \times 1}$ .

**Hypothesis 1** *A hypothesis is likely to be false if it is longer than this line*

Lorem ipsum dolor sit amet, consectetur

adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

| text | text | text | text | text | text |
|------|------|------|------|------|------|
| 1    | 84   | 46   | 4    | 3    | 5    |
| 2    | 67   | 24   | 5    | 1    | 5    |
| 3    | 54   | 26   | 4    | 2    | 5    |
| 4    | 44   | 16   | 5    | 2    | 5    |
| 5    | 53   | 21   | 6    | 2    | 5    |

Πίνακας 1: An example single column table

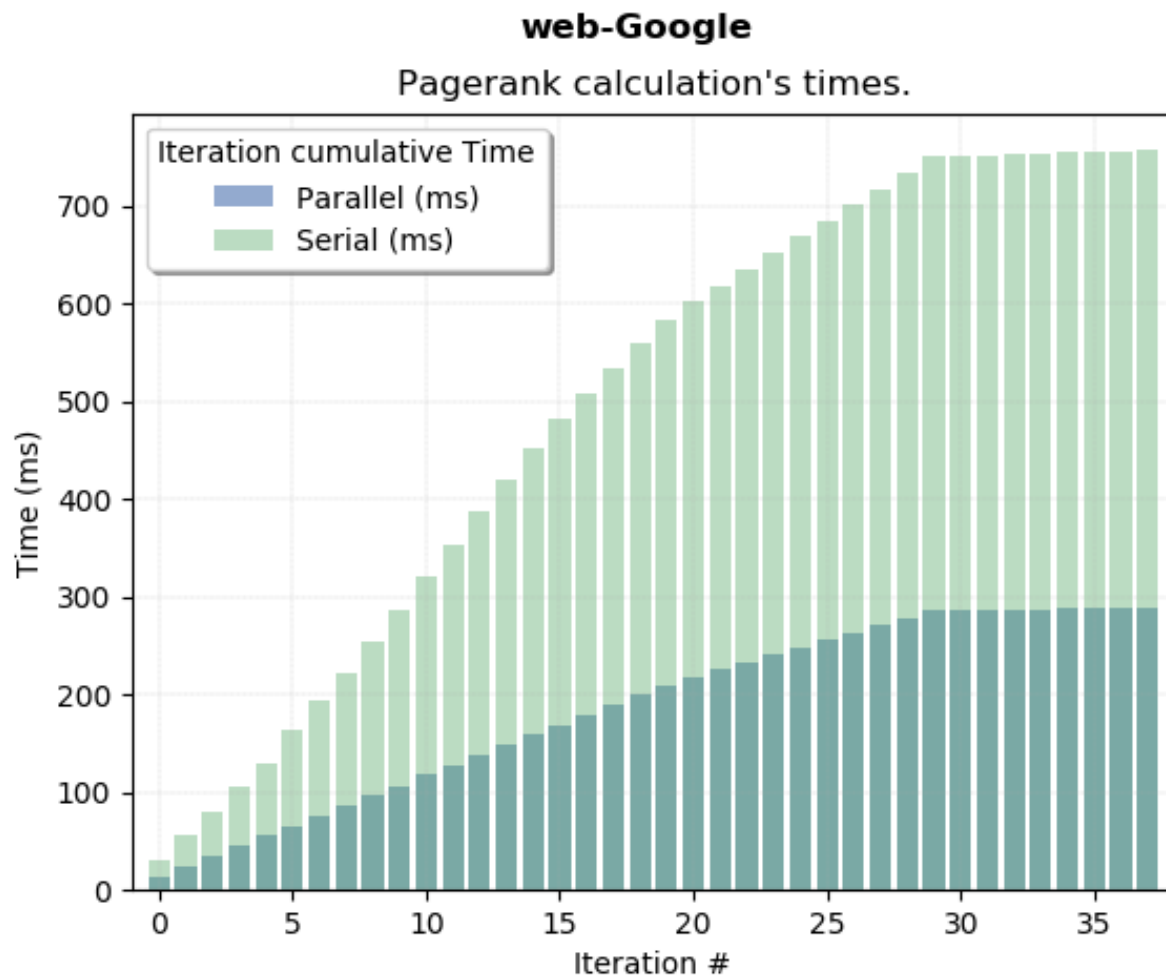
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean

faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

## Αναφορές

- [1] S. Brin και L. Page, «The anatomy of a large-scale hypertextual web search engine,» *Computer networks and ISDN systems*, τόμ. 30, αρθμ. 1-7, σσ. 107–117, 1998.
- [2] Y. Saad, *Iterative methods for sparse linear systems*. siam, 2003, τόμ. 82, σσ. 103–105.



Σχήμα 1: A two column figure (indicate with figure\*)