

HY342 - Παράλληλος Προγραμματισμός

Άσκηση 1

Σε αυτή την άσκηση θα αναπτύξετε ένα παράλληλο πρόγραμμα για την επεξεργασία γράφων. Το πρόγραμμα πρέπει να διαβάσει ένα γράφο ως μια λίστα από ακμές αποθηκευμένες σε ένα αρχείο κειμένου, όπου κάθε γραμμή περιγράφει μια ακμή με τη μορφή "src dst" δηλαδή τον αριθμό του κόμβου απ' όπου ξεκινά η ακμή και τον αριθμό του κόμβου απ' όπου τελειώνει η ακμή. Οι αριθμοί είναι ακέραιοι 64bit, δηλαδή long σε μορφή κειμένου (διαβάζονται με `scanf("%ld")` ή `atol()`). Για τις μετρήσεις του speedup θα χρειαστεί να αγνοήσετε γραμμές σχολίων (που αρχίζουν με το χαρακτήρα #).

Το πρόγραμμά σας θα πρέπει να υπολογίζει παράλληλα το PageRank του κάθε κόμβου, εκτελώντας τον αλγόριθμο PageRank για 50 επαναλήψεις. Αρχικοποιήστε όλους τους κόμβους με 1.0.

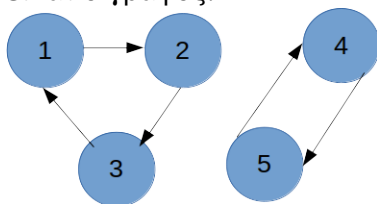
Χρησιμοποιήστε την απλοποιημένη μορφή του αλγορίθμου PageRank με damping factor 85%. Δηλαδή:

1. Κάθε κόμβος αρχικοποιείται με 1.0.
2. Το PageRank κάθε κόμβου μοιράζεται εξ'ίσου στους γείτονες μέσω των outgoing edges.
3. Το νέο PageRank κάθε κόμβου καθορίζεται ως 0.15 συν το 85% του αθροίσματος που έστειλαν οι γείτονες μέσω των incoming edges.

Το αποτέλεσμα θα πρέπει να παράγεται με τη μορφή αρχείου κειμένου `pagerank.csv`, που θα περιέχει ένα ζευγάρι αριθμών ανά γραμμή: "node, pagerank". Για παράδειγμα, ο γράφος που περιέχει τις παρακάτω ακμές:

```
1 2
2 3
3 1
4 5
5 4
```

είναι ο γράφος:



Η έξοδος για αυτόν τον γράφο θα είναι το αρχείο `pagerank.csv` με περιεχόμενα:

```
node,pagerank
1,1.0
2,1.0
3,1.0
```

4,1.0
5,1.0

Παραδώστε τον κώδικα του προγράμματός σας και μια αναφορά με τις μετρήσεις στο συνολικό χρόνο του προγράμματος για 1, 2, 3, και 4 threads.

Για τις μετρήσεις, χρησιμοποιήστε τα παρακάτω datasets του SNAP (<http://snap.stanford.edu/data/index.html>).

1. p2p-Gnutella24: <http://snap.stanford.edu/data/p2p-Gnutella24.txt.gz>
2. ego-Facebook: http://snap.stanford.edu/data/facebook_combined.txt.gz
3. email-Enron: <http://snap.stanford.edu/data/email-Enron.txt.gz>

Μετρήστε τουλάχιστον 5 runs για κάθε πειραματικό configuration, και αναφέρετε το average και standard deviation του χρόνου εκτέλεσης στην αναφορά σας, καθώς και το average speedup.