Διάταξη κόμβων σε διεπίπεδα σύνθετα δίκτυα με τον Biplex PageRank

Σιδέρης Γεώργιος Νικόλαος 1622

Είναι γνωστός ο αλγόριθμος PageRank που χρησιμοποιείται απο την Google για κατάταξη των ιστοσελίδων στο διαδίκτυο, με βάση τις συνδέσεις μεταξύ τους.

Το άρθρο “*A biplex approach to PageRank centrality: From classic to multiplex networks”* των Francisco Pedroche, Miguel Romance και Regino Criado προτείνει τρόπους να επεκταθεί ο αλγόριθμος για πολυεπίπεδα δίκτυα και εύρεση των σημαντικότερων κόμβων για χρήση σε άλλες εφαρμογές.

Στο πλαίσιο της εργασίας υλοποιήθηκαν δύο αλγόριθμοι που αναλύονται στο άρθρο, ο ένας διεπίπεδη εκδοχή του κλασικού PageRank και ενας τρόπος υπολογισμού κατάταξης για πολυεπίπεδα δίκτυα.

Τα στάδια των δυο αλγορίθμων είναι ίδια, με μερικές διαφοροποιήσεις

**Διαδικασία Υπολογισμού**

1. Αρχικοποίηση πινάκων/παραμέτρων
2. Υπολογισμός Pa (πίνακας κατανομής γειτνίασης)
3. Υπολογισμός Πραγματικών και Μη Πραγματικών κόμβων κάθε επιπέδου (RealCounter, NoRealCounter)\*
4. Επέκταση για εισόδους με dangling nodes (χωρίς εξερχόμενες ακμές)

Αν υπάρχουν dangling nodes, τότε Pa = Pa + d \* uT

1. Υπολογισμός v (personalization vector)
2. Αρχικοποίηση “biplex pagerank” και σφάλματος
3. Επαναληπτική διαδικασία
4. Εμφάνιση των σημαντικότερων κόμβων

*\* Χρειάζεται μόνο στον Biplex v2*

**Ορισμός σημαντικών πινάκων / παραμέτρων**

Για την κατανόηση των αλγορίθμων χρειάζεται αποσαφήνιση κάποιων εννοιών και πινάκων/παραμέτρων και οι τιμές τους.

**Personalization Vector** : Το διάνυσμα που δίνει τις πιθανότητες “τηλεμεταφοράς” από τον ένα κόμβο στον άλλον (χωρίς σύνδεση στο δίκτυο)

**Dangling κόμβων** : Ένας κόμβος χωρίς εξερχόμενες συνδέσεις

**a** : Το damping ratio του αλγορίθμου, πχ αν είναι 0.5 υπάρχει ιση πιθανότητα ένας χρήστης να μεταφερθεί απο τον κόμβο που βρίσκεται σε έναν άλλον μέσω σύνδεσης, ή μέσω “τηλεμεταφοράς” (χρησιμοποιούμε την τιμή 0.85 όπως στον πραγματικο PageRank)

**e** : Tο διάνυσμα (1,....1)τ  με μήκος n (κόμβοι δικτύου)

Kout : Το διάνυσμα που υποδεικνύει πόσες εξερχόμενες ακμές έχει κάθε κόμβος

**Pa** : O πίνακας n\*n όπου Pa(i,j) ορίζεται ως

αν ο i είναι dangling Pa(i,j) = 0 για κάθε j ,

αλλιώς 1/Κout(i)

**u** : η πιθανοτική κατανομή των dangling nodes

**d** : διάνυσμα με μήκος n, όπου

d(i)=1, αν ο i είναι dangling,

0 , αλλιώς

**Pu** : Pagerank για το πραγματικό δίκτυο

**Pd** : Pagerank για το αντίστοιχο δίκτυο “τηλεμεταφοράς”

**Biplex Αpproach to Classic PageRank (v1)**

**Ορισμός**: Έστω ενα δίκτυο G με n ∊ ℕ κόμβους, και πίνακα γειτνίασης Pa, τότε:

p = pu + pd ∊ ℝn

**Υπολογισμός pu , pd**: Με επαναληπτική διαδικασία:

puT = puT \* a \* Pa + pdT\*a

pdT = (1-a) \* puT + (1-a) \* pdT \* e \* vT

Σε πρώτο στάδιο, υλοποιήθηκε ο αλγόριθμος που περιγράφεται στο άρθρο, ο οποίος είναι μια εναλλακτική περίπτωση του κλασικού PageRank, σε δυο επίπεδα. Το πρώτο επίπεδο θεωρείται το δίκτυο που εξετάζουμε στον PageRank με τις ακμές του, και το 2ο επίπεδο είναι το “δίκτυο τηλεμεταφοράς”, όπου υπάρχουν ακμές μεταξύ όλων των κόμβων και τα βάρη τους δίνονται από τον personalization vector **v**, και ο πίνακας γειτνίασης είναι ο e \* vT.

**Έλεγχος αποτελεσμάτων σε σύγκριση με τον PageRank**

Εκτελούμε το biplex.m με το αρχείο lesmis2.txt (το lesmis χωρις dangling nodes), ένα δίκτυο 77 κόμβων με 284 συνδέσεις. Μετά από 67 επαναλήψεις, οι παρακάτω είναι οι 10 κόμβοι με το μεγαλύτερο ranking:

**Value: Index:**

0.3319 77

0.3294 66

0.1653 67

0.0096 56

0.0072 76

0.0067 40

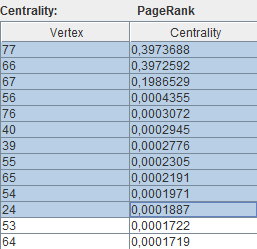
0.0064 39

0.0052 55

0.0050 65

0.0047 24

Έπειτα με το **CentiBiN**, εκτελούμε τον PageRank για το ίδιο δίκτυο και τα αποτελέσματα είναι τα εξής:



Συγκρίνοντας τα αποτελέσματα, βλέπουμε ότι η κατάταξη των κόμβων είναι ίδια , **επαληθεύοντας της ορθότητα του Biplex PageRank**

**Multilayer Approach for PageRank of Multiplex**

**Networks (v2)**

**Ορισμός**: Έστω ένα δίκτυο G με n ∊ ℕ κόμβους, και πίνακα γειτνίασης Pa, τότε:

p = ½ \* (pu + pu2 + pd + pd2) ∊ ℝn

**Yπολογισμός pu , pu2, pd, pd2** : Με επαναληπτική διαδικασία:

2 \* puT = puT \* a \* Pa + pu2T + 2 \* a \* pdT

2 \* pu2T = puT + pu2T \* a \* Pa2 + 2 \* a \* pd2T

2 \* pdT = (1-a) \* (pu2T + pdT \* e \* vT + pd2 \* e \* vT )

2 \* pd2T = (1-a) \* (pu2T + pdT \* e \* v2T + pd2 \* e \* v2T )

Υλοποιήθηκε ο αλγόριθμος που περιγράφεται στο άρθρο για 2 δίκτυα, ο οποίος μπορεί να επεκταθεί εύκολα και για περισσότερα επίπεδα.

Εκτελούμε το biplex\_v2.m, με το lesmis.txt ως 1ο και 2ο επίπεδο και περιμένουμε το αποτέλεσμα να είναι πανομοιότυπο με την εκτέλεση του biplex (v1) για το ίδιο δίκτυο.

|  |  |
| --- | --- |
| **Biplex v1** | **Biplex v2** |
| Value: Index:  0.0370 77.0000  0.0348 76.0000  0.0343 39.0000  0.0334 32.0000  0.0325 75.0000  0.0324 53.0000  0.0323 57.0000  0.0323 45.0000  0.0323 46.0000  0.0323 43.0000 | Value: Index:  0.0372 77.0000  0.0351 76.0000  0.0346 39.0000  0.0336 32.0000  0.0327 75.0000  0.0326 53.0000  0.0325 57.0000  0.0325 45.0000  0.0325 46.0000  0.0325 43.0000 |

Όντως, βλέπουμε ότι τα αποτελέσματα των 2 αλγόριθμων είναι ίδια, κάτι που επαληθεύει την ορθότητα της υπόθεσης που κάναμε πριν.

Επίσης, μπορούμε να κάνουμε μια γρήγορη αξιολόγηση της ορθότητας του αλγορίθμου, χρησιμοποιώντας ως πρώτο δίκτυο αυτό που χρησιμοποιήσαμε πριν και ως δεύτερο το παρακάτω:

77 1

76 1

39 1

Οι ακμές του κατευθύνονται από τους πιο σημαντικούς κόμβους στον 1ο, οπότε περιμένουμε ο κόμβος **1** να είναι πολύ σημαντικός. Και πράγματι:

**Value: Index:**

0.0301 1.0000

0.0258 77.0000

0.0215 32.0000

O 1 είναι σημαντικότερος κόμβος στο διεπίπεδο δίκτυο, κάτι που επαληθεύει εν μέρει την ορθότητα του αλγορίθμου.

Το συμπέρασμα είναι ότι ο δεύτερος αλγόριθμος μπορεί εύκολα να επεκταθεί για περισσότερα επίπεδα, έχοντας πολλές χρήσεις για κατάταξη κόμβων. Για παράδειγμα, για να βρεθούν οι πιο σημαντικοί σταθμοί σε ένα δίκτυο μετρό, με κάθε γραμμή να αποτελεί ένα επίπεδο ή για τους πιο σημαντικούς συγγραφείς επιστημονικών άρθρων με επίπεδα τα Paper to Paper Citation και Author to Author Citation και ούτω καθεξής.