Σύνθετα Δίκτυα

com+plex: with+ -fold (having parts)

Διδάσκων –
Δημήτριος Κατσαρός

Διάλεξη 08η

Μετρικές κεντρικότητας

Centrality measures

Περιεχόμενα

Παρουσιάσαμε

- Degree centrality (DC)
- Shortest-Path Betweenness Centrality (SPBC)
- Power Community Index (PCI)
- Closeness centrality (CC)
- Bridging centrality (BC)
- επεκτάσεις σε κατευθυνόμενα δίκτυα

Θα παρουσιάσουμε

- Φασματικές κεντρικότητες
  - PageRank vector
  - Katz status index

PageRank κεντρικότητα

- Φοιτητές του Stanford: Larry Page & Sergey Brin
- Google
  - PageRank για διάταξη ιστοσελίδων
  - Ανεξάρτητη του υποβαλλόμενου ερωτήματος στη μηχανή αναζήτησης

(12) **United States Patent**
Page

(10) **Patent No.: US 6,285,999 B1**
(45) **Date of Patent: Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221–231.*

L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometricka, vol. 18, pp. 39–43.

C.H. Hubbell, "An input–output approach to clique identification sociometry," 1965, pp. 377–399.

Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26–48.

E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471–479.

Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297–312.

Ανάλυση συνδέσμων: Κεντρική ιδέα

- Ένας σύνδεσμος από την ιστοσελίδα p προς την ιστοσελίδα q σηματοδοτεί επιδοκιμασία/έγκριση (endorsement)
  - Η ιστοσελίδα p θεωρεί την ιστοσελίδα q ως αυθεντία (authority) σε κάποιο ζήτημα
- Επεξεργασία του γραφήματος του Παγκοσμίου Ιστού για συστάσεις (recommendations)
- Ανάθεση μιας τιμής αυθεντίας (authority value) σε κάθε ιστοσελίδα

Η αρχική εξίσωση αθροίσματος

- Το PageRank μιας σελίδας είναι το άθροισμα του PageRank των σελίδων που δείχνουν σ' αυτή:

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|}$$

- Το πρόβλημα με τη εξίσωση αυτή είναι ότι δεν ξέρουμε το PageRank των σελίδων που "δείχνουν" στη P_i
- Το πρόβλημα επιλύθηκε με επαναληπτική διαδικασία
  - Αρχικά κάθε σελίδα έχει το ίδιο PageRank, ίσο με 1/n
  - Ακολουθούμε την παραπάνω εξίσωση επαναληπτικά

Η επαναληπτική διαδικασία (1/2)

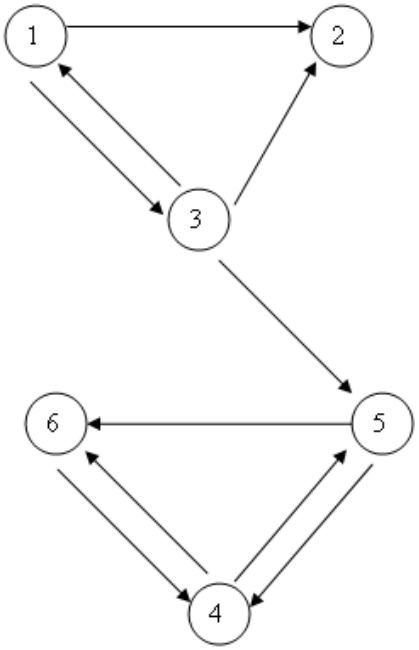
- Έστω ότι $r_{k+1}(P_i)$ είναι το PageRank της σελίδας P_i στην επανάληψη k+1:

$$r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \frac{r_k(P_j)}{|P_j|}$$

- Η διαδικασία ξεκινά με $r_0(P_i)=1/n$ για κάθε σελίδα
- Συνεχίζεται με την ελπίδα ότι τελικά θα συγκλίνει

Η επαναληπτική διαδικασία (2/2)

- Εφαρμόζοντας την επαναληπτική διαδικασία στο μικρό γράφημα αριστερά, μετά από μερικές επαναλήψεις έχουμε τον πίνακα δεξιά:



Iteration 0	Iteration 1	Iteration 2	Rank at Iter. 2
$r_0(P_1) = 1/6$	$r_1(P_1) = 1/18$	$r_2(P_1) = 1/36$	5
$r_0(P_2) = 1/6$	$r_1(P_2) = 5/36$	$r_2(P_2) = 1/18$	4
$r_0(P_3) = 1/6$	$r_1(P_3) = 1/12$	$r_2(P_3) = 1/36$	5
$r_0(P_4) = 1/6$	$r_1(P_4) = 1/4$	$r_2(P_4) = 17/72$	1
$r_0(P_5) = 1/6$	$r_1(P_5) = 5/36$	$r_2(P_5) = 11/72$	3
$r_0(P_6) = 1/6$	$r_1(P_6) = 1/6$	$r_2(P_6) = 14/72$	2

Αναπαράσταση της επανάληψης με πίνακα

- Η προηγούμενες εξισώσεις υπολογίζουν το PageRank των σελίδων μια σελίδα κάθε φορά
- Με χρήση πινάκων αντικαθιστούμε το σύμβολο Σ
- Εισαγάγουμε
  - τον πίνακα H, και
  - το 1 x n διάνυσμα π^T
- Ο H είναι ένας row-normalized πίνακας υπερσυνδέσεων με $H_{ij}=1/|P_i|$, εάν υπάρχει σύνδεσμος από τον κόμβο i στον j, αλλιώς $H_{ij}=0$
- Παρόλο που ο H έχει την ίδια μη-μηδενική δομή με τον δυαδικό πίνακα γειτνιάσεων, τα μη μηδενικά στοιχεία του H είναι πιθανότητες

Παράδειγμα αναπαράστασης με πίνακα

$$\mathbf{H} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

- Τα μη-μηδενικά στοιχεία της γραμμής i αναπαριστούν τους εξερχόμενους συνδέσμους της σελίδας i
- Τα μη-μηδενικά στοιχεία της στήλης i αναπαριστούν τους εισερχόμενους συνδέσμους στη σελίδα i
- Η προηγούμενη εξίσωση γίνεται τώρα:

$$\pi^{(k+1)T} = \pi^{(k)T}\mathbf{H}$$

Επίδοση της αναπαράστασης με πίνακα

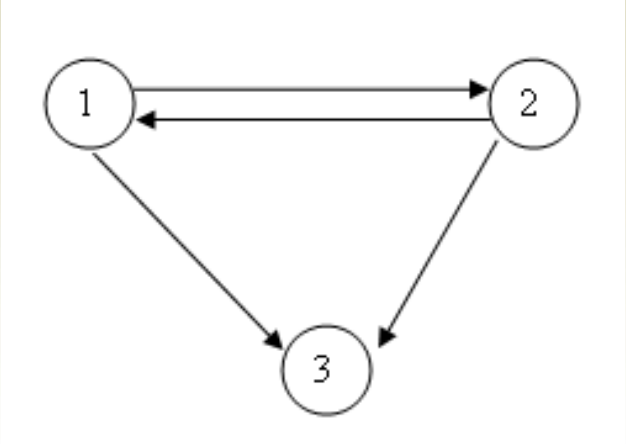
1. Κάθε επανάληψη της προηγούμενης εξίσωσης απαιτεί έναν πολλαπλασιασμό, άρα $O(n^2)$ πολυπλοκότητα
2. Ο H είναι γενικά πολύ αραιός (sparse), άρα
   - Απαιτεί μικρό αποθηκευτικό χώρο
   - Ο πολλαπλασιασμός είναι πιο οικονομικός σε σχέση με το $O(n^2)$
     - Απαιτεί μόνο O(nnz(H)), όπου nnz(H) είναι ο αριθμός των μη-μηδενικών
     - Μετρήσεις δείχνουν ότι το nnz(H) ~ 10n
     - Άρα υπολογιστικό κόστος της τάξης O(n)
3. Η επαναληπτική διαδικασία είναι απλά μια linear stationary process: είναι η κλασική power method πάνω στον H
4. Ο H μοιάζει με *στοχαστικό πίνακα πιθανοτήτων μετάβασης*, όμως είναι ***substochastic***, γιατί υπάρχουν **dangling nodes**, δηλ., χωρίς εξερχόμενους συνδέσμους

Προβλήματα της επαναληπτικής διαδικασίας

- Θα συγκλίνει;
- Κάτω από ποιες προϋποθέσεις ή ιδιότητες του Η θα συγκλίνει;
- Θα συγκλίνει σε κάτι που έχει "μαθηματικό"νόημα;
- Θα συγκλίνει σε ένα ή περισσότερα διανύσματα;
- Η σύγκλιση εξαρτάται από το αρχικό διάνυσμα $\pi^{(0)T}$;
- Πόσο γρήγορα θα συγκλίνει;

Προβλήματα της επαναληπτικής διαδικασίας

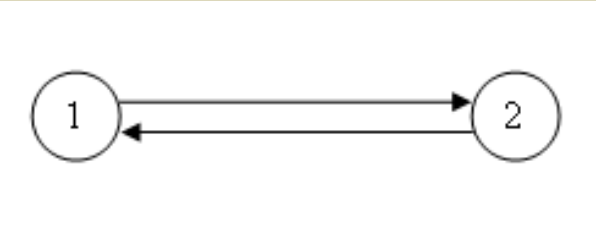
- Αρχικά, η επαναληπτική διαδικασία ξεκίνησε με $\pi^{(0)T}=1/n\mathbf{e}^T$ (όπου $\mathbf{e}^T$ είναι διάνυσμα-γραμμή με όλα 1)
- Προέκυψε το πρόβλημα της **καταβόθρας** (rank sinks)
  - σελίδες που αυξάνουν συνεχώς το PageRank τους
  - Στο παρακάτω παράδειγμα το κόμβος 3, ενώ στο προηγούμενο παράδειγμα η ομάδα των κόμβων 4, 5, και 6



- Μετά από 13 επαναλήψεις, $\pi^{(13)T}=(0\ \ 0\ \ 0\ \ 2/3\ \ 1/3\ \ 1/5)$

Προβλήματα της επαναληπτικής διαδικασίας

- Επίσης, καθώς οι κόμβοι αυξάνουν συνεχώς το PageRank τους, μερικοί δεν έχουν καθόλου
  - Τότε, ποιο είναι το νόημα της ταξινόμησης με βάση το PageRank, όταν η πλειονότητα έχει PageRank ίσο με 0;
- Υπάρχει το πρόβλημα των κύκλων



- Εάν, ξεκινήσουμε με $\pi^{(0)T}=(1\ \ 0)$, καταλήγουμε σε ατέρμονη διαδικασία
  - Στο διάνυσμα $\pi^{(k)T}=(1\ \ 0)$ για άρτιο k
  - Στο διάνυσμα $\pi^{(k)T}=(0\ \ 1)$ για περιττό k

Υπενθύμιση εννοιών Markov chains

- Με οποιοδήποτε διάνυσμα ξεκινήσουμε, όταν εφαρμοστεί η power method σε έναν Markov πίνακα P, συγκλίνει σε ένα μοναδικό θετικό διάνυσμα, το οποίο αποκαλείται *stationary vector*
- Προϋποθέσεις σύγκλισης
  - Ο P είναι stochastic: οι γραμμές αθροίζουν στο "1"
  - Ο P είναι irreducible: το υποκείμενο γράφημα είναι "strongly-connected"
  - Ο P είναι aperiodic: για οποιεσδήποτε σελίδες P_i και P_j υπάρχουν μονοπάτια από την P_i στην P_j (με οποιεσδήποτε επαναλήψεις) οποιουδήποτε μήκους, εκτός από ένα πεπερασμένο σύνολο μηκών
- Irreducible + aperiodic = primitive (πρωτογενής)
- Τα προβλήματα σύγκλισης του PageRank θα ξεπεραστούν εάν ο H τροποποιηθεί, ώστε να ικανοποιεί τις παραπάνω προϋποθέσεις

Πρώιμες προσαρμογές στο βασικό μοντέλο

- Οι Sergey Brin και Lawrence Page δεν χρησιμοποίησαν την έννοια της Markov chain, αλλά την έννοια του **random surfer**
- Μετά από "άπειρο χρόνο ταξιδιού", το ποσοστό του χρόνου που ο random surfer περνά σε μια σελίδα είναι ένα μέτρο της σημαντικότητας της σελίδας
- Δυστυχώς, υπάρχουν παγίδες για τον random surfer
  - pdf
  - image
  - data tables

Προσαρμογή στοχαστικότητας (1/2)

- Οι γραμμές $\mathbf{0}^T$ του H αντικαθίστανται με $1/n\mathbf{e}^T$
- Άρα ο random surfer, όταν συναντήσει έναν dangling node μπορεί από κει να μεταβεί σε οποιαδήποτε άλλη σελίδα
- Τον στοχαστικό πίνακα που προέκυψε από τον H τον συμβολίζουμε με S
- Για το γράφημα με τους 6 κόμβους είναι ο παρακάτω:

$$\mathbf{S} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Προσαρμογή στοχαστικότητας (2/2)

- Ο **S** παράγεται από μια *rank-one update* του **H**
- $\mathbf{S} = \mathbf{H} + \mathbf{a}(1/n\mathbf{e}^T)$
  - $a_i = 1$ εάν η σελίδα i είναι dangling node
  - $a_i = 0$ εάν η σελίδα i δεν είναι dangling node
- Ο **S** είναι συνδυασμός του αρχικού **H** με τον rank-one πίνακα $\mathbf{a}(1/n\mathbf{e}^T)$
- Η προσαρμογή αυτή εγγυάται ότι ο **S** είναι πίνακας μιας Markov chain
- Δεν εγγυάται όμως τη σύγκλιση

Προσαρμογή πρωτογένειας (1/2)

- Ο random surfer δεν ακολουθεί πάντα υπερσυνδέσμους
- Εγκαταλείπει την πλοήγηση και μεταβαίνει σε ένα "τυχαίο" URL
- "**Τηλεμεταφέρεται**" (**teleportation step**) και ξεκινά ξανά την πλοήγηση
- Προκύπτει ο πίνακας **G**, *Google matrix*

$$\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)1/n\mathbf{e}\mathbf{e}^T$$

- α (ελληνικό άλφα) έχει τιμή μεταξύ 0 και 1, και ελέγχει το ποσοστό του χρόνου που random surfer ακολουθεί υπερσυνδέσμους ή τηλεμεταφέρεται
- Η τελεμεταφορά είναι τυχαία, γιατί ο πίνακας τηλεμεταφοράς $E=1/n\mathbf{e}\mathbf{e}^T$ είναι ομοιόμορφος

Συνέπειες της προσαρμογής πρωτογένειας

- Ο **G** είναι *stochastic*: κυρτός συνδυασμός δυο στοχαστικών πινάκων **S** και **E**
- Ο **G** είναι *irreducible*: κάθε σελίδα συνδέεται άμεσα με κάθε άλλη
- Ο **G** είναι *aperiodic*: οι βρόχοι ($\mathbf{G_{ii}} > 0$ για κάθε i) δημιουργούν aperiodicity
- Ο **G** είναι *primitive*: επειδή $\mathbf{G^k} > 0$ για κάποιο k (για k=1)
  - Υπάρχει ένα μοναδικό π^T και όταν εφαρμόσουμε την power method στον G, θα συγκλίνει σ' αυτό

Συνέπειες της προσαρμογής πρωτογένειας

- Ο **G** είναι πολύ πυκνός, ευτυχώς μπορεί να γραφεί ως rank-one update του πολύ αραιού πίνακα υπερσυνδέσμων **H**

$$\begin{aligned}\mathbf{G} &= \alpha\mathbf{S} + (1-\alpha)1/n\mathbf{e}\mathbf{e}^T \\ &= \alpha(\mathbf{H} + 1/n\mathbf{a}\mathbf{e}^T) + (1-\alpha)1/n\mathbf{e}\mathbf{e}^T \\ &= \alpha\mathbf{H} + (\alpha\mathbf{a} + (1-\alpha)\mathbf{e})1/n\mathbf{e}^T\end{aligned}$$

- Ο **G** είναι τεχνητός
  - Το stationary vector δεν υπάρχει για τον **H**
  - Αλλά υπάρχει για τον **G**

Σύμβολα

- **H**: πολύ αραιός, substochastic πίνακας υπερσυνδέσμων
- **S**: αραιός, στοχαστικός, πιθανώς reducible πίνακας
- **G**: τελείως πυκνός, στοχαστικός, πρωτογενής πίνακας
- **E**: τελείως πυκνός, rank-one πίνακας τηλεμεταφοράς
- n: αριθμός σελίδων στη μηχανή της Google
- α: παράμετρος μεταξύ 0 και 1
- π^T: stationary row vector, PageRank διάνυσμα
- a^T: δυαδικό διάνυσμα dangling nodes

Η μέθοδος του PageRank

$$\pi^{(k+1)T} = \pi^{(k)T}\,\mathbf{G}$$

που είναι απλά η power method εφαρμοζόμενη στον **G**

Το παράδειγμα γραφήματος με 6 κόμβους

$$\mathbf{G} = .9\mathbf{H} + (.9 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + .1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}) 1/6 (1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1)$$

$$\mathbf{G} = \begin{pmatrix} 1/60 & 7/15 & 7/15 & 1/60 & 1/60 & 1/60 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 19/60 & 19/60 & 1/60 & 1/60 & 19/60 & 1/60 \\ 1/60 & 1/60 & 1/60 & 1/60 & 7/15 & 7/15 \\ 1/60 & 1/60 & 1/60 & 7/15 & 1/60 & 7/15 \\ 1/60 & 1/60 & 1/60 & 11/12 & 1/60 & 1/60 \end{pmatrix}$$

$$\pi^T = (.03721 \quad .05369 \quad .04151 \quad .3751 \quad .206 \quad .2862)$$

Υπολογισμός του διανύσματος PageRank

- Το πρόβλημα μπορεί να περιγραφεί με δυο τρόπους
  - Επίλυση του παρακάτω προβλήματος ιδιοδιανυσμάτων του π^T

$$\pi^T = \pi^T \mathbf{G}$$
$$\pi^T \mathbf{e} = 1$$

  - Επίλυση του γραμμικού ομογενούς συστήματος για το π^T

$$\pi^T (\mathbf{I} - \mathbf{G}) = \mathbf{0}^T$$
$$\pi^T \mathbf{e} = 1$$

Υπολογισμός του διανύσματος PageRank

- Στο πρώτο σύστημα, ο στόχος είναι να βρεθεί το κανονικοποιημένο κυρίαρχο αριστερό ιδιοδιάνυσμα που αντιστοιχεί στην κυρίαρχη ιδιοτιμή $\lambda_1=1$
- Στο δεύτερο σύστημα ο στόχος είναι να βρεθεί το κανονικοποιημένο αριστερό null vector του (**I-G**)
- Η εξίσωση κανονικοποίσης υπάρχει για να εγγυηθεί ότι το π^T είναι διάνυσμα πιθανοτήτων

Power method υπολογισμού του PageRank

- Είναι η παλιότερη και απλούστερη μέθοδος εύρεσης της κυρίαρχης (dominant) ιδιοτιμής και ιδιοδιανύσματος ενός πίνακα
- Άρα μπορεί να χρησιμοποιηθεί για εύρεση του stationary vector μιας Markov chain
  - Το stationary vector είναι απλά το κυρίαρχο αριστερό ιδιοδιάνυσμα
- Είναι εξαιρετικά αργή μέθοδος, μεταξύ των Gauss-Seidel, Jacobi, restarted GMRES
- Γιατί χρησιμοποιήθηκε;

Power method υπολογισμού του PageRank

- Είναι προγραμματιστικά απλή
- Εφαρμοζόμενη στον **G** μπορεί να γραφεί ως εφαρμογή στον πολύ αραιό **H**

$$\begin{aligned}\pi^{(k+1)T} &= \pi^{(k)T}\mathbf{G} \\ &= \alpha\pi^{(k)T}\mathbf{S} + \frac{1-\alpha}{n}\pi^{(k)T}\mathbf{e}\mathbf{e}^T \\ &= \alpha\pi^{(k)T}\mathbf{H} + (\alpha\pi^{(k)T}\mathbf{a} + 1 - \alpha)\mathbf{e}^T/n\end{aligned}$$

- Εκτελείται πάνω στον **H** και όχι πάνω στους **S** ή **G**
- Αποθηκεύονται μόνο οι **a**, **e**

Power method υπολογισμού του PageRank

- Οι άλλες μέθοδοι αναγκάζονται να προσπελάσουν τα στοιχεία του πίνακα, ενώ η power method μόνο διαμέσου του πολλαπλασιασμού διανύσματος-πίνακα
- Εκτός από την αποθήκευση του **H** και **a** απαιτεί μόνο την αποθήκευση του π^T και όχι πολλαπλά διανύσματα όπως οι άλλες μέθοδοι
- Απαιτεί πολύ λίγες επαναλήψεις για να επιτευχθεί η σύγκλιση
  - 50-100
- Το ερώτημα που προκύπτει είναι από ποιο/ποιους παράγοντες εξαρτάται/καθορίζεται η σύγκλιση

Ρυθμός σύγκλισης (1/2)

- Ο ασυμπτωτικός ρυθμός σύγκλισης της power method όταν εφαρμόζεται σε κάποιο Markov πίνακα εξαρτάται από το κλάσμα των δυο ιδιοτιμών που έχουν το μεγαλύτερο μέγεθος, λ_1, λ_2
- Για τους στοχαστικούς πίνακες, όπως ο G, ισχύει ότι $\lambda_1 = 1$
- Άρα η σύγκλιση εξαρτάται από την τιμή του λ_2
- Επειδή ο **G** είναι πρωτογενής, ισχύει ότι $|\lambda_2| < 1$
- Η εύρεση του είναι χρονοβόρα, οπότε δεν είναι φρόνιμο να σπαταλήσουμε πόρους για να έχουμε μια εκτίμηση του ρυθμού σύγκλισης

Ρυθμός σύγκλισης (2/2)

- Στις επόμενες διαφάνειες θα δείξουμε ότι εάν οι ιδιοτιμές του **S** είναι $\sigma(\mathbf{S})=\{1,\mu_2,\mu_3,\mu_n\}$ και του **G** είναι $\sigma(\mathbf{G})=\{1,\lambda_2,\lambda_3,\lambda_n\}$, τότε

$$\lambda_k = \alpha\mu_k \quad k=2,3,\ldots n$$

- Η δομή του Παγκοσμίου Ιστού είναι τέτοια που καθιστά πολύ πιθανό να ισχύει ότι $|\mu_2| =1$ (ή $|\mu_2| \approx 1$)
- Άρα $\lambda_2(\mathbf{G})=\alpha$ (ή $\lambda_2(\mathbf{G}) \approx \alpha$)
- Με α=.85, σημαίνει ότι μετά από 50 επαναλήψεις $\alpha^{50}=.85^{50} \approx .000296$, δηλ., 2-3 θέσεις ακρίβειας που είναι αρκετά ικανοποιητικές όταν το ranking συνδυάζεται με το περιεχόμενο

Απρόσμενη εφαρμογή του PageRank

Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions?

Stefano Allesina[1]*, Mercedes Pascual[2,3,4]

1 National Center for Ecological Analysis and Synthesis, Santa Barbara, California, United States of America, **2** Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Santa Fe Institute, Santa Fe, New Mexico, United States of America, **4** Howard Hughes Medical Institute

Abstract

A major challenge in ecology is forecasting the effects of species' extinctions, a pressing problem given current human impacts on the planet. Consequences of species losses such as secondary extinctions are difficult to forecast because species are not isolated, but interact instead in a complex network of ecological relationships. Because of their mutual dependence, the loss of a single species can cascade in multiple coextinctions. Here we show that an algorithm adapted from the one Google uses to rank web-pages can order species according to their importance for coextinctions, providing the sequence of losses that results in the fastest collapse of the network. Moreover, we use the algorithm to bridge the gap between qualitative (who eats whom) and quantitative (at what rate) descriptions of food webs. We show that our simple algorithm finds the best possible solution for the problem of assigning importance from the perspective of secondary extinctions in all analyzed networks. Our approach relies on network structure, but applies regardless of the specific dynamical model of species' interactions, because it identifies the subset of coextinctions common to all possible models, those that will happen with certainty given the complete loss of prey of a given predator. Results show that previous measures of importance based on the concept of "hubs" or number of connections, as well as centrality measures, do not identify the most effective extinction sequence. The proposed algorithm provides a basis for further developments in the analysis of extinction risk in ecosystems.

Μέτρηση μονοπατιών – Katz status index

- Η σημαντικότητα ενός κόμβου μετριέται με το weighted άθροισμα των μονοπατιών που οδηγούν σ' αυτόν τον κόμβο
- $A^m[i,j]$ = αριθμός μονοπατιών με μήκος m από τον κόμβο i στον j
- Υπολογισμός

$$P = bA + b^2A^2 + \cdots + b^mA^m + \cdots = (I - bA)^{-1} - I$$

- Συγκλίνει όταν $b < \lambda_1(A)$
- Διατάσσουμε τους κόμβους σύμφωνα με τα αθροίσματα στήλης του πίνακα P

Κεντρικότητα Katz

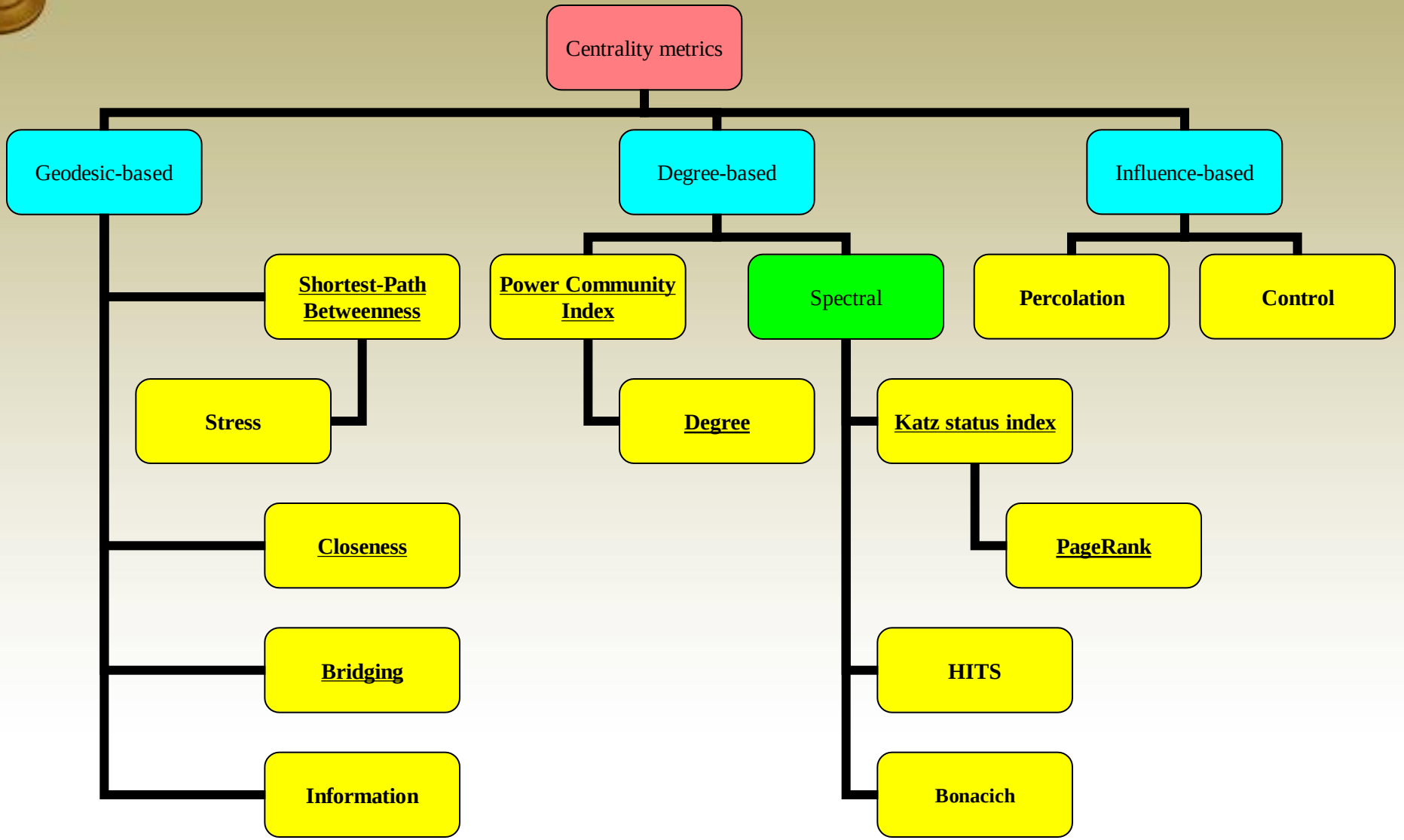
- Για τον κόμβο i:

$$C_{\text{Katz}}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)_{ji}$$

- Ως διάνυσμα για όλους τους κόμβους:

$$\vec{C}_{\text{Katz}} = ((I - \alpha A^T)^{-1} - I)\vec{I}$$

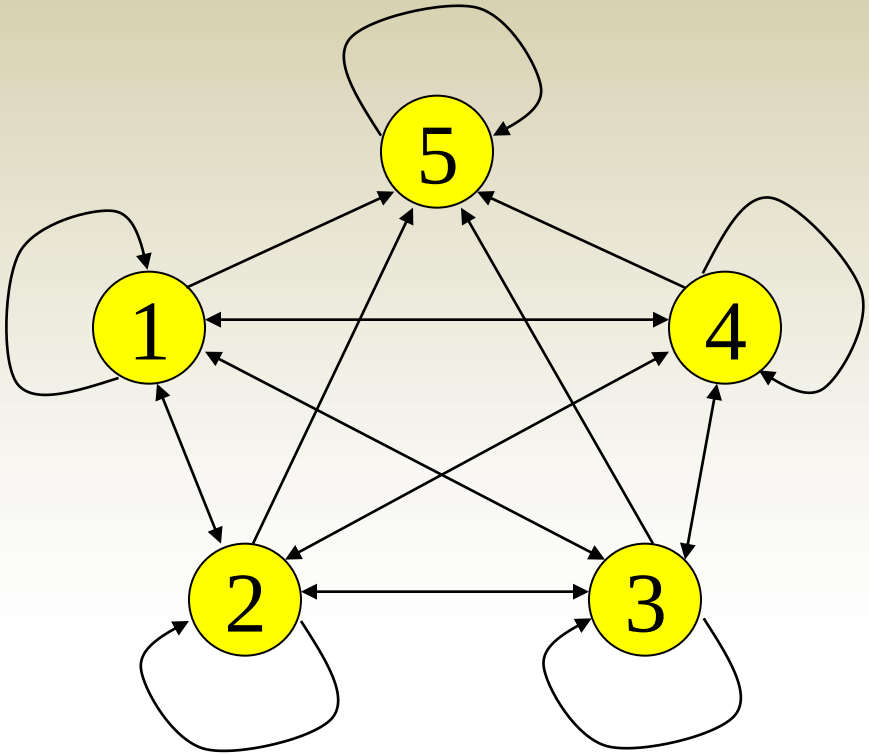
Οικογένειες μετρικών κεντρικότητας
(ενδεικτικά)



Ασκήσεις

- **Άσκηση 1.**

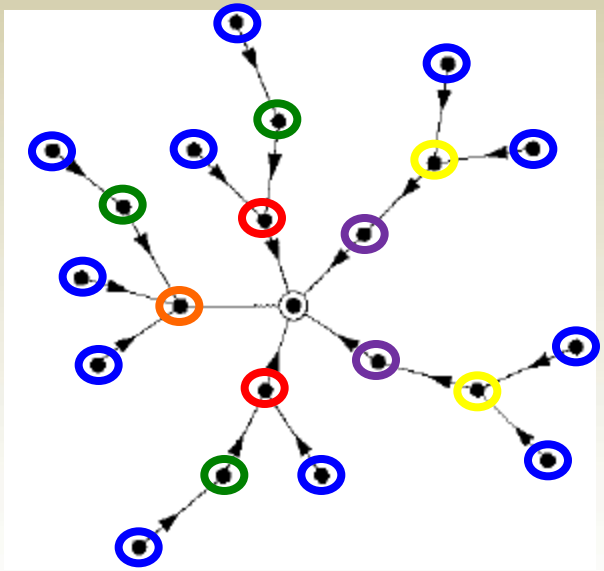
  Να βρεθεί η PageRank τιμή (με τηλεμεταφορά) των κόμβων του κάτωθι γραφήματος.



Ασκήσεις

Άσκηση 2.

Να βρεθεί η τιμή PageRank του κεντρικού κόμβου συναρτήσει του damping factor α και της γεωδεσικής απόστασης γ_i των κόμβων από τον κεντρικό κόμβο.



Λύση.

PageRank κόμβων με μπλε κύκλο (χωρίς incoming link): $x=1-\alpha$

PageRank κόμβων με πράσινο κύκλο: $y= 1-\alpha^2$

PageRank κόμβων με κόκκινο κύκλο: $z= (1-\alpha)(1+\alpha)^2$

PageRank κόμβων με κίτρινο κύκλο: $w= (1-\alpha)(2\alpha+1)$

PageRank κόμβων με μωβ κύκλο: $v= (1-\alpha)(2\alpha^2+\alpha+1)$

PageRank κόμβων με πορτοκαλί κύκλο: $u= (1-\alpha)(\alpha^2+3\alpha+1)$

PageRank κεντρικού κόμβου= $\alpha(2v+2z+u)+(1-\alpha)= \ldots.= (1-\alpha)(7\alpha^3+9\alpha^2+5\alpha+1)$

Παρατηρήστε ότι υπάρχουν:

- 7 κόμβοι σε απόσταση γ_3=3 hops (7 μπλε)
- 9 κόμβοι σε απόσταση γ_2=2 hops (οι 2 κίτρινοι, οι 3 πράσινοι, και 4 μπλε)
- 5 κόμβοι σε απόσταση γ_1=1 hop (οι 2 μωβ, οι 2 κόκκινοι, και ο πορτοκαλί)
- 1 κόμβος σε απόσταση γ_0=0 hops (ο κεντρικός)

PageRank(κεντρικού_κόμβου)= $(1-\alpha)(7\alpha^{\gamma 3}+9\alpha^{\gamma 2}+5\alpha^{\gamma 1}+\alpha^{\gamma 0})$