Ανάκληση Πληροφορίας

Διδάσκων –
Δημήτριος Κατσαρός

Διάλεξη 14η

Το πρόβλημα PageRank ως γραμμικό σύστημα

PageRank ως γραμμικό σύστημα

- Το πρόβλημα του PageRank μπορεί να γραφεί είτε ως
  - Πρόβλημα ιδιοδιανύσματος: $\boldsymbol{\pi}^T(\alpha\mathbf{S}+(1-\alpha)\mathbf{e}\mathbf{v}^T) = \boldsymbol{\pi}^T$
  - Πρόβλημα Γραμμικού συστήματος: $\boldsymbol{\pi}^T(\mathbf{I}-\alpha\mathbf{S})= (1-\alpha)\mathbf{v}^T$
- Ποια από τις δυο μορφές είναι προτιμότερες;
- Υπάρχει κάποια διαφορά;

Ιδιότητες του (**I**-α**S**)

1. (**I**-α**S**) είναι ένας **M** πίνακας
2. Ο (**I**-α**S**) δεν είναι ιδιόμορφος
3. Τα αθροίσματα γραμμών του (**I**-α**S**) είναι ίσα προς 1-α
4. $||(\mathbf{I}-\alpha\mathbf{S})||_\infty = 1 + \alpha$
5. Αφού ο (**I**-α**S**) είναι ένας **M** πίνακας, $(\mathbf{I}-\alpha\mathbf{S})^{-1} \geq 0$
6. Τα αθροίσματα γραμμών του $(\mathbf{I}-\alpha\mathbf{S})^{-1}$ είναι $(1-\alpha)^{-1}$. Επομένως, $||(\mathbf{I}-\alpha\mathbf{S})^{-1}||_\infty = (1 - \alpha)^{-1}$
7. Ο *condition number* είναι $\kappa_\infty(\mathbf{I}-\alpha\mathbf{S}) = (1+\alpha)/(1-\alpha)$

Επειδή ο (**I**-α**S**) είναι αρκετά πυκνός, θα θέλαμε να ελέγξουμε εάν παρόμοιες ιδιότητες ισχύουν για τον (**I**-α**H**)

Ιδιότητες του ($\mathbf{I}$-α$\mathbf{H}$)

- Χρησιμοποιώντας το διάνυσμα για τους dangling κόμβους $\mathbf{a}\mathbf{v}^T$
- Το γραμμικό σύστημα: $\boldsymbol{\pi}^T(\mathbf{I}-\alpha\mathbf{H}-\alpha\mathbf{a}\mathbf{v}^T) = (1-\alpha)\mathbf{v}^T$
- Εάν δέσουμε $\boldsymbol{\pi}^T\mathbf{a} = \gamma$, τότε το γραμμικό σύστημα γίνεται: $\boldsymbol{\pi}^T(\mathbf{I}-\alpha\mathbf{H}) = (1-\alpha+\alpha\gamma)\mathbf{v}^T$
- Η scalar μεταβλητή γ κρατά το συνολικό PageRank των dangling κόμβων
- Αφού στο τέλος θα εφαρμόσουμε την εξίσωση κανονικοποίησης $\boldsymbol{\pi}^T\mathbf{e} = 1$, διαλέγουμε αυθαίρετα μια τιμή για το γ, π.χ., γ = 1

Το PageRank ως γραμμικό σύστημα

- **<u>ΘΕΩΡΗΜΑ</u>**. Επιλύοντας το γραμμικό σύστημα $\mathbf{x}^T(\mathbf{I}-\alpha\mathbf{H}) = \mathbf{v}^T$ και θέτοντας $\boldsymbol{\pi}^T=\mathbf{x}^T/\mathbf{x}^T\mathbf{e}$ έχουμε ως αποτέλεσμα το διάνυσμα PageRank

  Επιπλέον, ο $(\mathbf{I}-\alpha\mathbf{H})$ έχει πολλές από τις ιδιότητες του $(\mathbf{I}-\alpha\mathbf{S})$

Ιδιότητες του (**I**-α**H**)

- Ο (**I**-α**H**) είναι ένας **M** πίνακας
- Ο (**I**-α**H**) δεν είναι ιδιόμορφος
- Τα αθροίσματα γραμμών του (**I**-α**H**) είναι είτε ίσα προς 1-α για τους μη-dangling κόμβους ή 1 για τους dangling κόμβους
- $||(\mathbf{I}-\alpha\mathbf{H})||_{\infty} = 1 + \alpha$
- Αφού ο (**I**-α**H**) είναι ένας **M** πίνακας, $(\mathbf{I}-\alpha\mathbf{H})^{-1} \geq 0$
- Τα αθροίσματα γραμμών του $(\mathbf{I}-\alpha\mathbf{H})^{-1}$ είναι 1 για τους dangling κόμβους και μικρότερο ή ίσο με $(1-\alpha)^{-1}$ για τους μη dangling κόμβους
- Ο *condition number* είναι $\kappa_{\infty}(\mathbf{I}-\alpha\mathbf{H}) \leq (1+\alpha)/(1-\alpha)$
- Η γραμμή του $(\mathbf{I}-\alpha\mathbf{H})^{-1}$ που αντιστοιχεί στον dangling κόμβο i είναι το $\mathbf{e}_i^T$, όπου $\mathbf{e}^i$ είναι η i-οστή στήλη του μοναδιαίου πίνακα

Σχόλια

- Για μικρά προβλήματα, η προσέγγιση αυτή είναι πολύ πιο γρήγορη, π.χ., company Intranet
- Επειδή καθώς το α τείνει στο 1 η power μέθοδος αργεί να συγκλίνει
- Νέοι ερευνητικοί ορίζοντες
- Φυσικά, καθώς το α τείνει στο 1 τα ζητήματα ευαισθησίας παραμένουν και για το γραμμικό σύστημα

Απόδειξη 1: PageRank ως Γραμμικό Σύστημα

- Το $\boldsymbol{\pi}^T$ είναι το PageRank διάνυσμα εάν ικανοποιεί $\boldsymbol{\pi}^T\mathbf{G}=\boldsymbol{\pi}^T$ και $\boldsymbol{\pi}^T\mathbf{e}=1$
- Προφανώς, $\boldsymbol{\pi}^T\mathbf{e} = 1$
- Το να δείξουμε ότι ισχύει $\boldsymbol{\pi}^T\mathbf{G}=\boldsymbol{\pi}^T$ είναι ισοδύναμο με το να δείξουμε ότι $\boldsymbol{\pi}^T(\mathbf{I}-\mathbf{G}) = \mathbf{0}^T$, το οποίο είναι ισοδύναμο με το να δείξουμε ότι $\mathbf{x}^T(\mathbf{I}-\mathbf{G}) = \mathbf{0}^T$

$$\begin{aligned}\mathbf{x}^T(\mathbf{I}-\mathbf{G}) &= \mathbf{x}^T(\mathbf{I}-\alpha\mathbf{H}-\alpha\mathbf{a}\mathbf{v}^T-(1-\alpha)\mathbf{e}\mathbf{v}^T)\\ &= \mathbf{x}^T(\mathbf{I}-\alpha\mathbf{H}-\mathbf{x}^T(\alpha\mathbf{a}+(1-\alpha)\mathbf{e})\mathbf{v}^T)\\ &= \mathbf{v}^T-\mathbf{v}^T=\mathbf{0}^T\end{aligned}$$

Απόδειξη 1: PageRank ως Γραμμικό Σύστημα

- Η προηγούμενη γραμμή προκύπτει από το γεγονός ότι $\mathbf{x}^T(\alpha\mathrm{a}+(1-\alpha)\mathbf{e})\mathbf{v}^T$, επειδή:

$$
\begin{aligned}
1 &= \mathbf{v}^T\mathbf{e} \\
&= \mathbf{x}^T(\mathbf{I}-\alpha\mathbf{H})\mathbf{e} \\
&= \mathbf{x}^T\mathbf{e}-\alpha\mathbf{x}^T\mathbf{H}\mathbf{e} \\
&= \mathbf{x}^T\mathbf{e}-\alpha\mathbf{x}^T(\mathbf{e}-\mathbf{a}) \\
&= (1-\alpha)\mathbf{x}^T\mathbf{e}+\alpha\mathbf{x}^T\mathbf{a}
\end{aligned}
$$

Απόδειξη 2: PageRank ως Γραμμικό Σύστημα

- Ας δούμε μια εναλλακτική θεώρηση του PageRank ως γραμμικό σύστημα. Έχουμε δει ότι:

$$\mathbf{S} = \mathbf{H} + \mathbf{a}(\frac{1}{n}\mathbf{e}^T) \quad \text{or} \quad \mathbf{S} = \mathbf{H} + \mathbf{a}\mathbf{v}^T$$

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

- Ξεκινώντας από τον ορισμό του PageRank ως πρόβλημα ιδιοδιανύσματος (eigenvector), έχουμε ότι:

$$\begin{aligned}
\pi^T\mathbf{G} &= \pi^T && \Longrightarrow \\
\pi^T(\alpha\mathbf{S} + (1-\alpha)\mathbf{e}\mathbf{v}^T) &= \pi^T && \Longrightarrow \\
(1-\alpha)\pi^T\mathbf{e}\mathbf{v}^T &= \pi^T(\mathbf{I} - \alpha\mathbf{S}) && \Longrightarrow \\
(1-\alpha)(\pi^T\mathbf{e})\mathbf{v}^T &= \pi^T(\mathbf{I} - \alpha\mathbf{S}) && \overset{\pi^T\mathbf{e}=1}{\Longrightarrow} \\
(1-\alpha)\mathbf{v}^T &= \pi^T(\mathbf{I} - \alpha\mathbf{S}) && \overset{transposing}{\Longrightarrow} \\
(1-\alpha)\mathbf{v} &= (\mathbf{I} - \alpha\mathbf{S})^T\pi && \Longrightarrow \\
(1-\alpha)\mathbf{v} &= (\mathbf{I} - \alpha\mathbf{H}^T - \alpha\mathbf{v}\mathbf{a}^T)\pi && \Longrightarrow \\
(\mathbf{I} - \alpha\mathbf{H}^T - \alpha\mathbf{v}\mathbf{a}^T)\pi &= (1-\alpha)\mathbf{v} && \text{Linear System Formulation}
\end{aligned}$$

Απόδειξη 2: PageRank ως Γραμμικό Σύστημα

- Έστω ότι: $\mathbf{R} = \mathbf{I} - \alpha \mathbf{H}^T$
- Από το θεώρημα Sherman-Morrison, γνωρίζουμε ότι ο αντίστροφος μιας rank-one update u⊗w πάνω σε έναν πίνακα A μπορεί να γραφεί ως συνάρτηση του αντιστρόφου του A ως εξής:

$$\left(\mathbf{A} + \mathbf{u} \otimes \mathbf{w}\right)^{-1} = \mathbf{A}^{-1} - \frac{(\mathbf{A}^{-1}\mathbf{u}) \otimes (\mathbf{w}\mathbf{A}^{-1})}{1 + \mathbf{w}\mathbf{A}^{-1}\mathbf{u}}$$

- Συνεπώς, στη δική μας περίπτωση για τον **R** έχουμε ότι:

$$(\mathbf{R} - \alpha \mathbf{v}\mathbf{a}^T)^{-1} = \mathbf{R}^{-1} + \frac{\mathbf{R}^{-1}\mathbf{v}\mathbf{a}^T\mathbf{R}^{-1}}{\frac{1}{\alpha} + \mathbf{a}^T\mathbf{R}^{-1}\mathbf{v}}$$

Απόδειξη 2: PageRank ως Γραμμικό Σύστημα

- Από την έκφραση του PageRank ως γραμμικού συστήματος: $(\mathbf{I} - \alpha\mathbf{H}^T - \alpha\mathbf{v}\mathbf{a}^T)\pi = (1-\alpha)\mathbf{v}$

- με $\mathbf{R} = \mathbf{I} - \alpha\mathbf{H}^T$

- έχουμε $(\mathbf{R} - \alpha\mathbf{v}\mathbf{a}^T)\pi = (1-\alpha)\mathbf{v}$

- άρα $(\mathbf{R} - \alpha\mathbf{v}\mathbf{a}^T)^{-1}(\mathbf{R} - \alpha\mathbf{v}\mathbf{a}^T)\pi = (1-\alpha)(\mathbf{R} - \alpha\mathbf{v}\mathbf{a}^T)^{-1}\mathbf{v}$

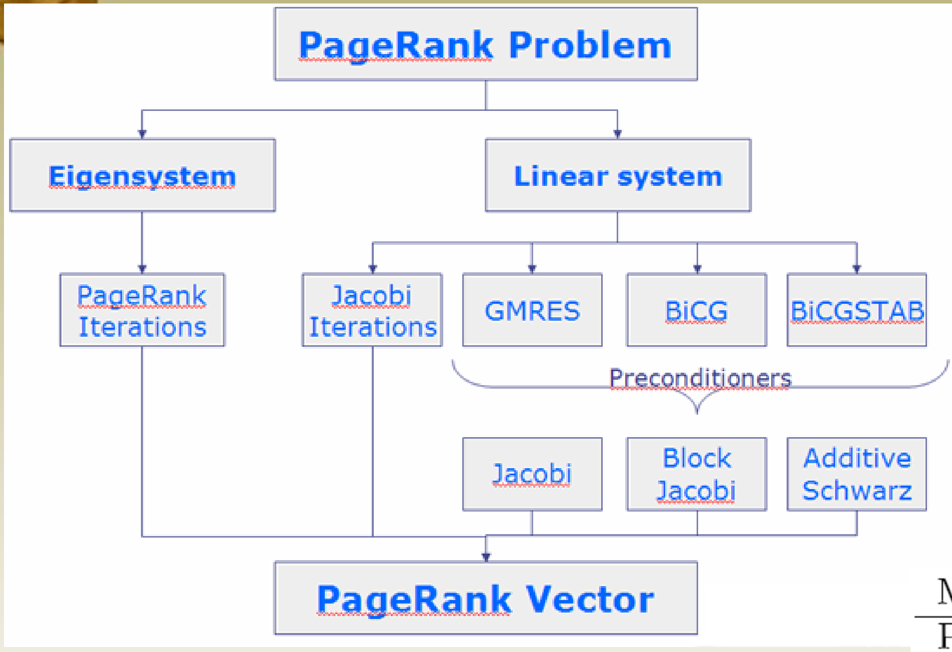
- Συνεπώς:

$$\begin{aligned}\pi &= (1-\alpha)\Big(\mathbf{R}^{-1} + \frac{\mathbf{R}^{-1}\mathbf{v}\mathbf{a}^T\mathbf{R}^{-1}}{\frac{1}{\alpha} + \mathbf{a}^T\mathbf{R}^{-1}\mathbf{v}}\Big)\mathbf{v} \\ &= (1-\alpha)\Big(\mathbf{R}^{-1}\mathbf{v} + \frac{\mathbf{R}^{-1}\mathbf{v}\mathbf{a}^T\mathbf{R}^{-1}\mathbf{v}}{\frac{1}{\alpha} + \mathbf{a}^T\mathbf{R}^{-1}\mathbf{v}}\Big)\end{aligned}$$

Απόδειξη 2: PageRank ως Γραμμικό Σύστημα

- Έστω: $$\mathbf{R}\mathbf{y} = \mathbf{v} \Longrightarrow \mathbf{y} = \mathbf{R}^{-1}\mathbf{v}$$
- τότε
$$\pi = (1-\alpha)\Big(\mathbf{y} + \frac{\mathbf{y}\mathbf{a}^T\mathbf{y}}{\frac{1}{\alpha} + \mathbf{a}^T\mathbf{y}}\Big) \Longrightarrow$$
$$\pi = (1-\alpha)\Big(1 + \frac{\mathbf{a}^T\mathbf{y}}{\frac{1}{\alpha} + \mathbf{a}^T\mathbf{y}}\Big)\mathbf{y} \Longrightarrow$$
$$\pi = \gamma\mathbf{y}$$
- όπου $$\gamma = (1-\alpha)\Big(1 + \frac{\mathbf{a}^T\mathbf{y}}{\frac{1}{\alpha} + \mathbf{a}^T\mathbf{y}}\Big)$$
- Αυτό σημαίνει ότι αφού υπολογίσω το **y** με γνωστούς τρόπους, μπορώ να βρω το **π** εύκολα, αφού εφαρμόσω μια κατάλληλη κανονικοποίηση στο **y**, ώστε $|\boldsymbol{\pi}|_1=1$

Παράλληλες υλοποιήσεις του PageRank



Method	IP	SAXPY	MV	Storage
PAGERANK		1	1	$M + 3v$
JACOBI		1	1	$M + 3v$
GMRES	$i + 1$	$i + 1$	1	$M + (i + 5)v$
BiCG	2	5	2	$M + 10v$
BiCGSTAB	4	6	2	$M + 10v$

Table 1: Computational Requirements. Operations per iteration: IP counts inner products, SAXPY counts AXPY operations, MV counts matrix vector multiplications, and Storage counts the number of matrices and vectors required for the method.

Απαιτήσεις από τη μέθοδο

- Work with nonsymmetric matrices
- Be easily parallelizable
- Μέθοδοι
  - Power iterations
  - Jacobi iterations
  - Krylov subspace methods
    - Generalize Minimum Residual (GMRES)
    - Biconjugate Gradient (BiCG)
    - Quasi-Minimal Residual (QMR)
    - Conjugate Gradient Squared (CGS)
    - Biconjugate Gradient Stabilized (BiCGSTAB)
    - Chebyshev Iterations

Δεδομένα πειραματισμού

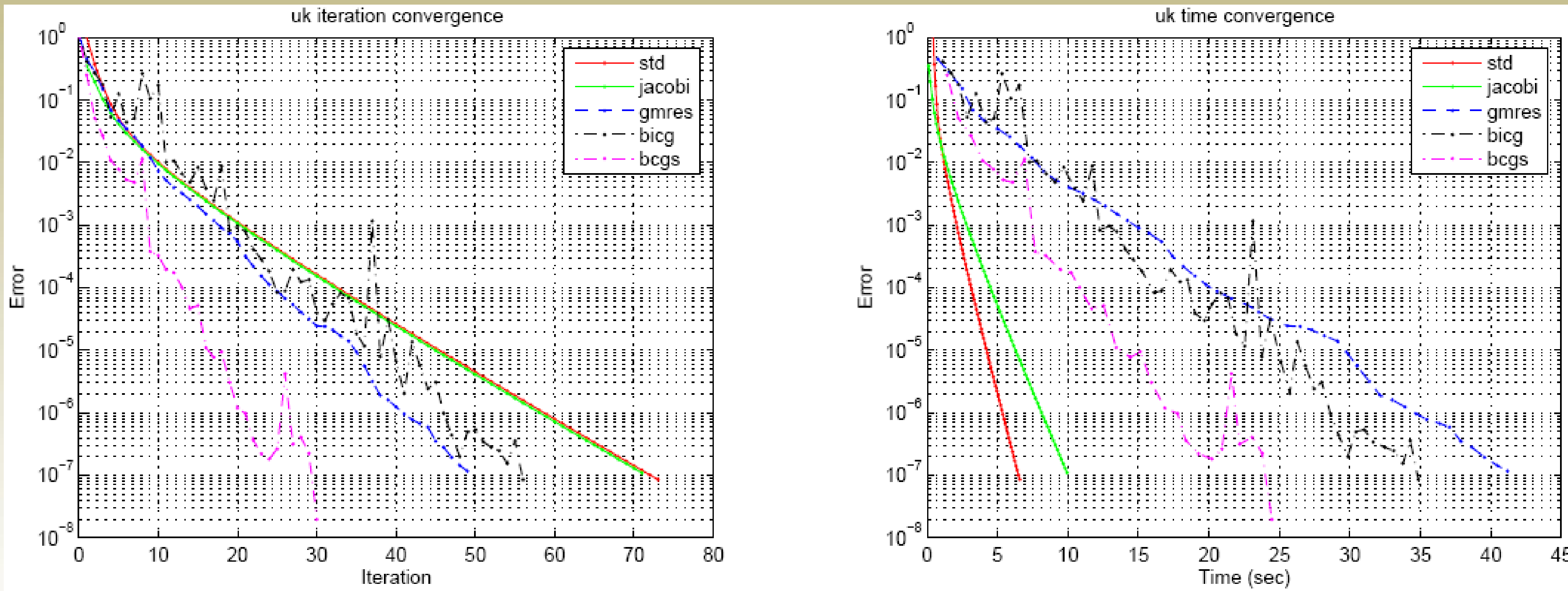
Name	Nodes	Links	Storage Size
edu	2M	14M	176 MB
yahoo-r2	14M	266M	3.25 GB
uk	18.5M	300M	3.67 GB
yahoo-r3	60M	850M	10.4 GB
db	70M	1B	12.3 GB
av	1.4B	6.6B	80 GB

Σύγκλιση 1/4

Graph	PR	Jacobi	GMRES	BiCG	BCGS
edu 20 procs	84 0.06 s	84 0.06 s	$21^{\dagger}$ 0.6 s	44^{*} 0.85 s	21^{*} 0.41 s
yahoo-r2	71	65	12	20	10
uk 60 procs	73 0.08 s	71 0.14 s	22^{*} 0.8 s	25^{*} 0.78 s	11^{*} 1.05 s
yahoo-r3 60 procs	76 2.4 s	75 2.2 s			
db 60 procs	62 13 s	58 12 s	29 22 s	45 21 s	15^{*} 21 s
av 140 procs	72 30 s	76 30 s			26 60 s

In the table, the first line for each graph denotes the number of iterations required for each method to converge to an absolute residual value of 10−7. The second line shows the mean time per iteration at the given number of processors. For the Krylov subspace methods, no superscript means we used no preconditioner, whereas * denotes a block Jacobi preconditioner and † denotes an additive Schwartz preconditioner

Σύγκλιση 2/4

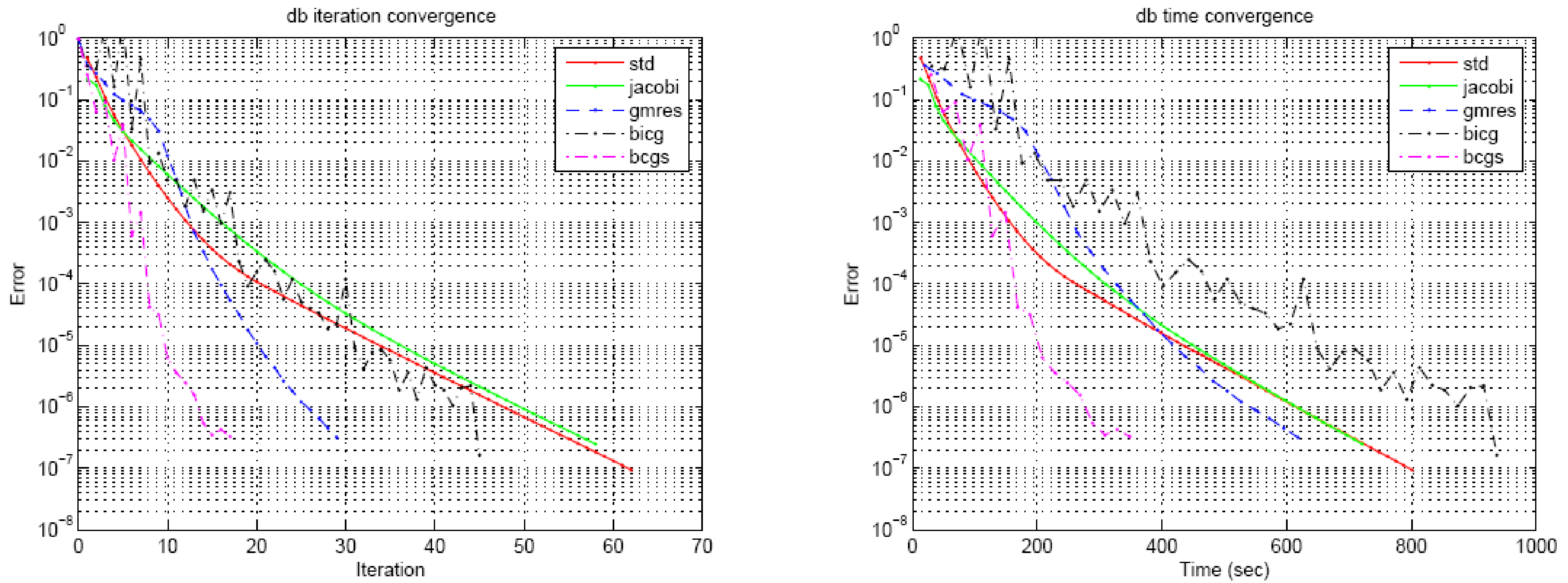


(a) Convergence Iterations

(b) Convergence Time

Figure 4: Convergence of iterative methods on the "uk" Web graph.

Σύγκλιση 3/4

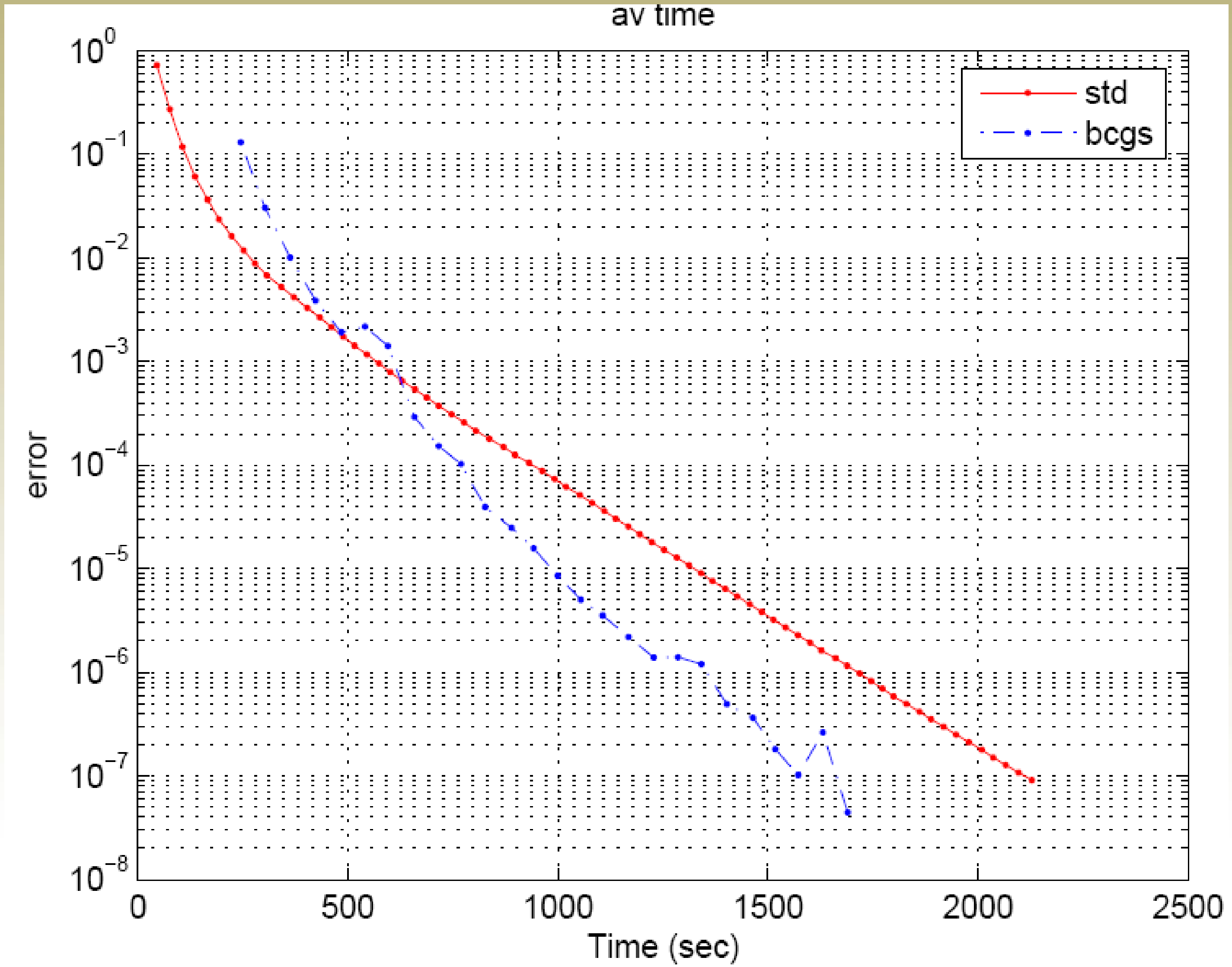


(a) Convergence Iterations

(b) Convergence Time

Figure 5: Convergence of iterative methods on the "db" Host graph.

Σύγκλιση 4/4



av time
std
bcgs
error
Time (sec)
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
0
500
1000
1500
2000
2500

Σύγκλιση και τηλεμεταφορά