Markov Chains and Web Ranking: a Multilevel Adaptive Aggregation Method

Hans De Sterck

Department of Applied Mathematics, University of Waterloo

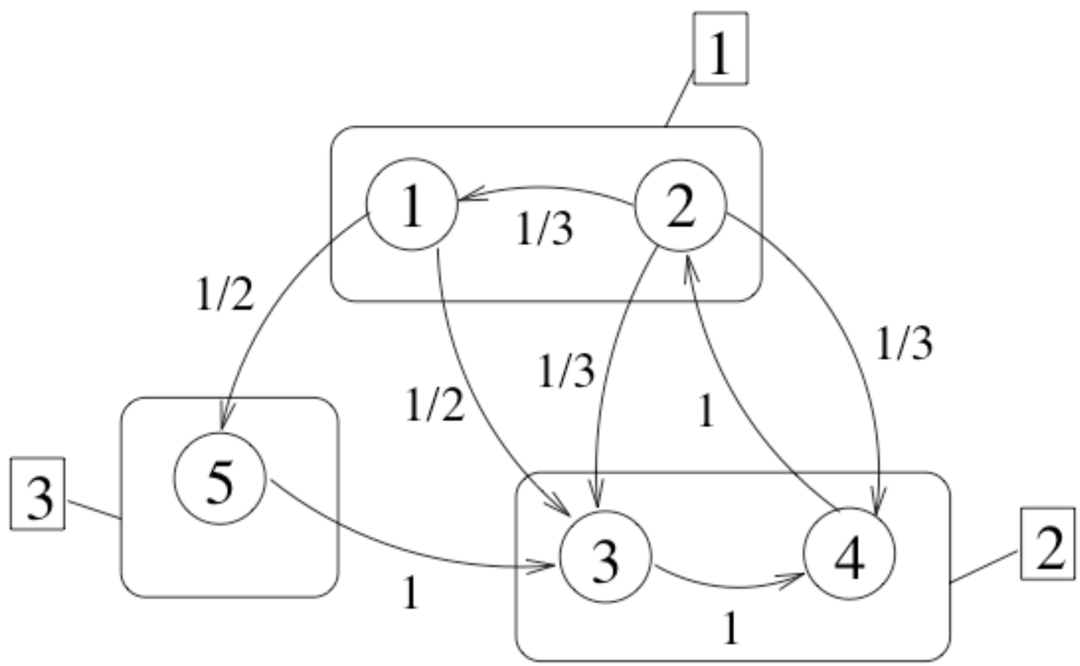
Quoc Nguyen; Steve McCormick, John Ruge, Tom Manteuffel (Boulder, Colorado)

University of Waterloo

UBC CS May 2007

1. Simple Markov Chain Example

- 5 states
- each outgoing edge same probability (random walk on directed graph)



Simple Markov Chain Example

- start in one state with probability 1: what is the stationary probability vector after ∞ number of steps?

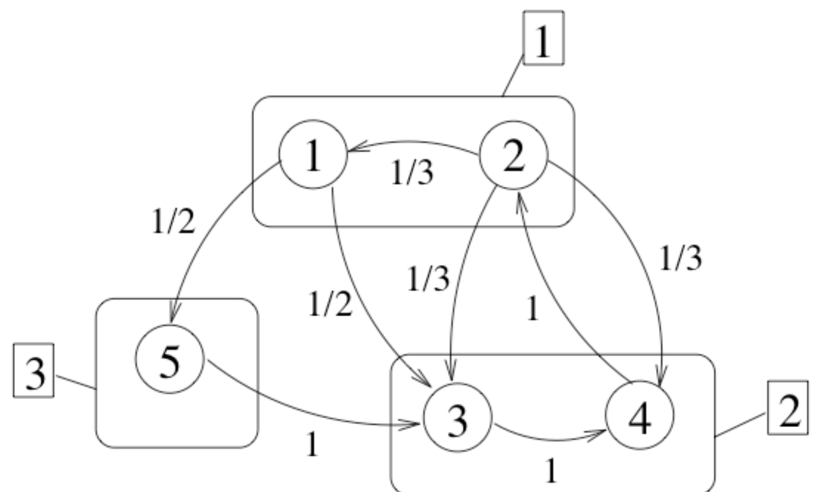
$$\mathbf{x}_{i+1} = B\,\mathbf{x}_i$$

- stationary probability:

$$B\,\mathbf{x} = \mathbf{x} \qquad \|\mathbf{x}\|_1 = 1$$

$$\mathbf{x}^T = [2/19 \;\; 6/19 \;\; 4/19 \;\; 6/19 \;\; 1/19]$$

$$B = \begin{bmatrix} 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1/2 & 1/3 & 0 & 0 & 1 \\ 0 & 1/3 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix}$$



UBC CS May 2007
hdesterck@uwaterloo.ca

2. Problem Statement

$$B\,\mathbf{x} = \mathbf{x} \qquad \|\mathbf{x}\|_1 = 1$$

- *B* is column-stochastic

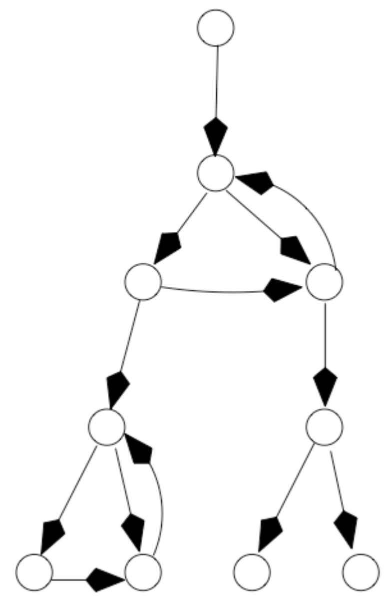
$$b_{i,j} \geq 0, \qquad \sum_i b_{i,j} = 1 \;\forall j$$

- *B* is irreducible (every state can be reached from every other state in the directed graph)

$$\Longrightarrow$$

$$\exists!\; \mathbf{x} : B\,\mathbf{x} = \mathbf{x} \quad \text{and} \quad \|\mathbf{x}\|_1 = 1, \qquad x_i > 0 \;\forall i$$

(no probability sinks!)



UBC CS May 2007
hdesterck@uwaterloo.ca

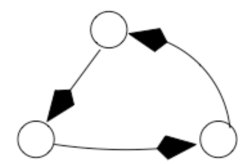
Problem Statement

$$B\,\mathbf{x} = \mathbf{x} \qquad \|\mathbf{x}\|_1 = 1$$

- B is column-stochastic
- B is irreducible
- B is a-periodic

$\Longrightarrow$

$$\mathbf{x} = \lim_{n\to\infty} B^n\,\mathbf{x}_0 \qquad \text{for any } \mathbf{x}_0$$



- largest eigenvalue of B: $\lambda_1 = 1$

University of Waterloo

UBC CS May 2007
hdesterck@uwaterloo.ca

3. Traditional, One-Level Iterative Methods

- Power Method

$$\mathbf{x}_{i+1} = B\mathbf{x}_i$$

- Weighted Jacobi Method (WJAC)

$$B\,\mathbf{x} = \mathbf{x}$$

$$(B - I)\,\mathbf{x} = 0 \qquad A\mathbf{x} = 0$$

$$A = L + D + U$$

$$\mathbf{x}_{i+1} = N((1-w)\,\mathbf{x}_i + w\,D^{-1}(L+U)\mathbf{x}_i)$$

normalization: $$N(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|_1} \qquad (\mathbf{x} \neq 0)$$

UBC CS May 2007
hdesterck@uwaterloo.ca

Traditional, One-Level Iterative Methods

- Power Method
  - convergence factor: $|\lambda_2|$
  - convergence is very slow when $|\lambda_2| \approx 1$

- WJAC: similar

4. Aggregation

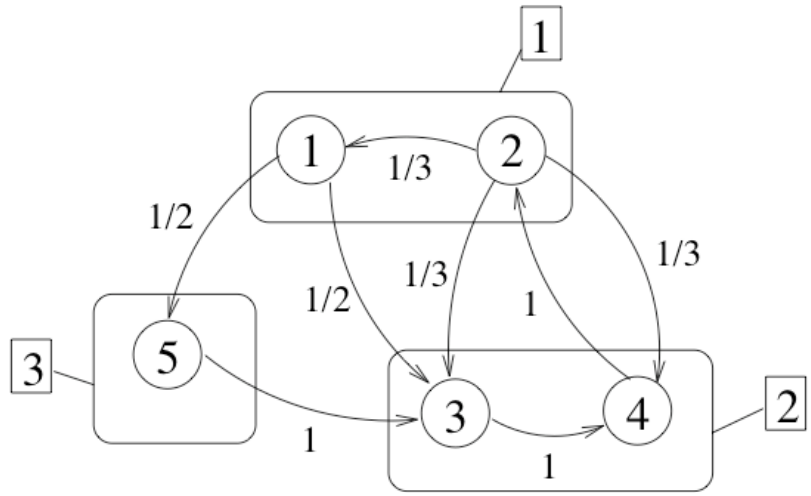
- form three coarse, aggregated states

$$x_{c,I} = \sum_{i \in I} x_i$$

$$\mathbf{x}_c^T = [8/19 \;\; 10/19 \;\; 1/19]$$

$$B_c \, \mathbf{x}_c = \mathbf{x}_c$$

$$b_{c,IJ} = \frac{\displaystyle\sum_{j \in J} x_j \left(\sum_{i \in I} b_{ij} \right)}{\displaystyle\sum_{j \in J} x_j}$$



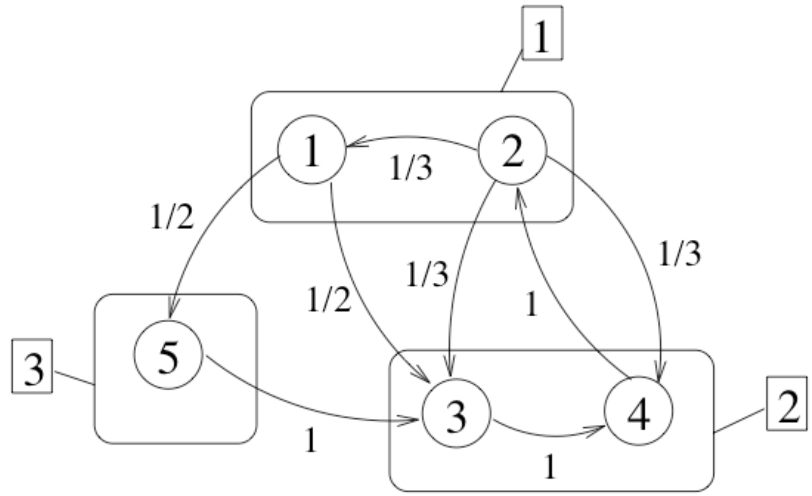
$$B_c = \begin{bmatrix} 1/4 & 3/5 & 0 \\ 5/8 & 2/5 & 1 \\ 1/8 & 0 & 0 \end{bmatrix}$$

(Simon and Ando, 1961)

Aggregation

$$B_c\,\mathbf{x}_c = \mathbf{x}_c$$

$$b_{c,IJ} = \frac{\sum_{j\in J} x_j \left(\sum_{i\in I} b_{ij}\right)}{\sum_{j\in J} x_j}$$



$$\boxed{\mathbf{x}_c = P^T\,\mathbf{x}}$$

$$\boxed{B_c = P^T\,B\,\mathrm{diag}(\mathbf{x})\,P\,\mathrm{diag}(P^T\,\mathbf{x})^{-1}}$$

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(Krieger, Horton, ...)

University of Waterloo

5. Two-Level Acceleration by Aggregation

Algorithm: Two-level acceleration by agglomeration

choose initial guess $\mathbf{x}$

repeat

$\mathbf{x} \leftarrow N(\text{Relax}(A, \mathbf{x})) \qquad \nu$ times

$A_c = P^T A \, \text{diag}(\mathbf{x}) \, P \, \text{diag}(P^T \mathbf{x})^{-1}$

$\mathbf{x}_c \leftarrow$ solve $A_c \mathbf{x}_c = 0$, $\|x_c\|_1 = 1$ (coarse-level solve)

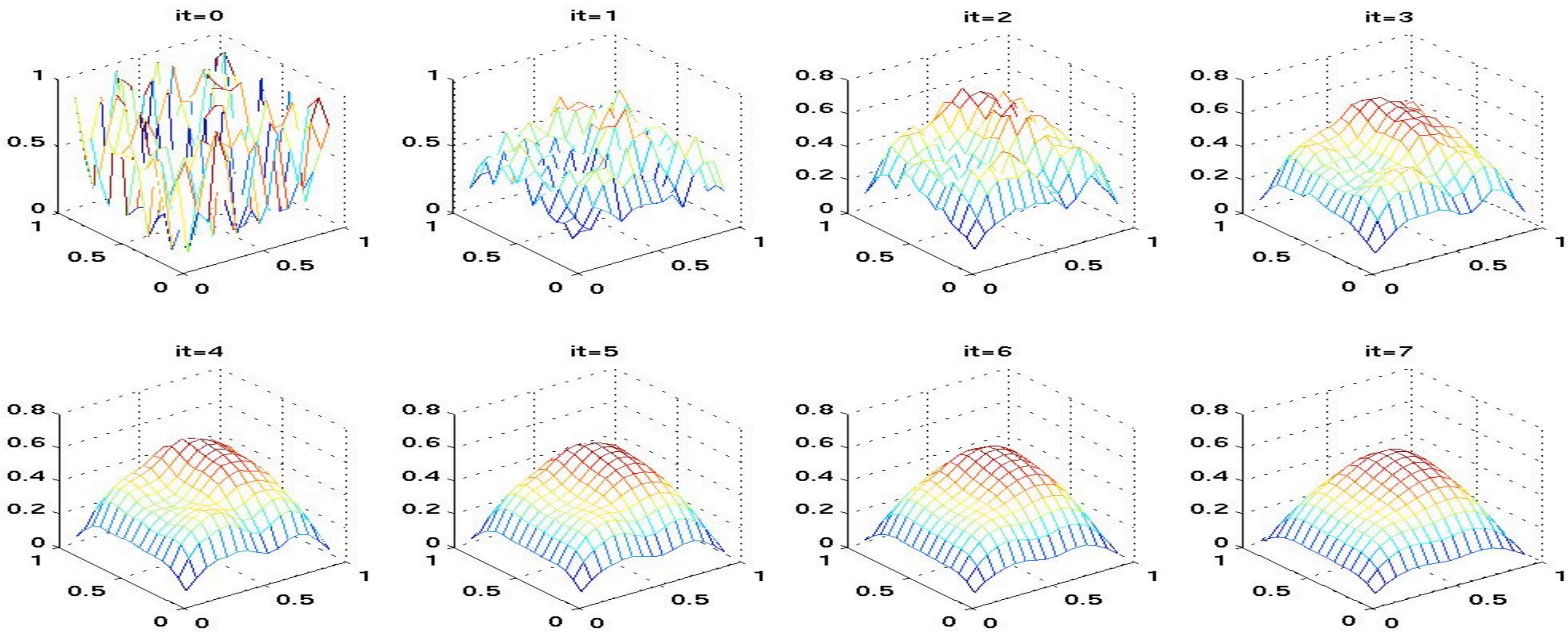
$\mathbf{x} = N(\text{diag}(P \, \text{diag}(P^T \mathbf{x})^{-1} \mathbf{x}_c) \, \mathbf{x})$ (coarse-level correction)

end

$\approx$ Iterative Aggregation/Disaggregation Algorithm (IAD) (Takahashi 1975, Koury et al 1984, Schweitzer and Kindle 1986, Krieger 1990, Marek and Mayer 1998, ...)

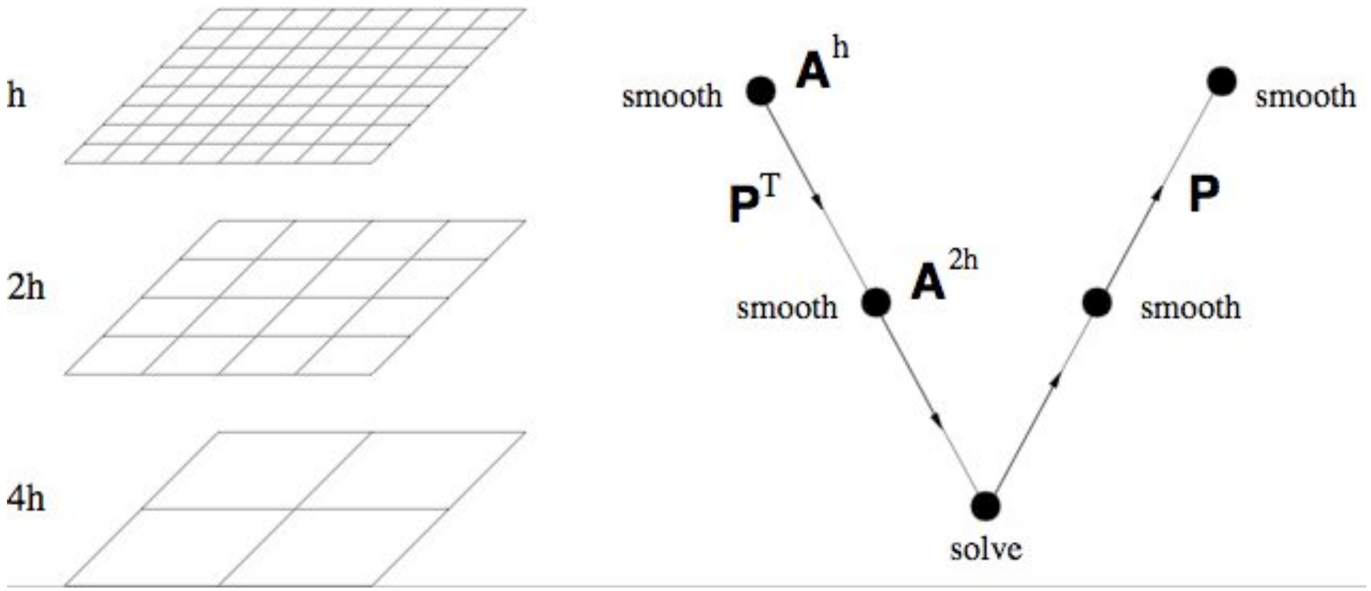
6. Relation with Multigrid for PDEs

$$-u_{xx} - u_{yy} = f(x, y)$$



- high-frequency error is removed by relaxation
- low-frequency-error needs to be removed by coarse-grid correction

multigrid hierarchy: V-cycle



- multigrid V-cycle:
  - relax (=smooth) on various grids
  - transfer error using restriction (P^T) and interpolation (P)

University of Waterloo

UBC CS May 2007
hdesterck@uwaterloo.ca

7. Multilevel Adaptive Aggregation (MAA) Method

Algorithm: Multilevel Adaptive Aggregation method (V-cycle)

$\mathbf{x} = \text{MAA\_V}(A, \mathbf{x}, \nu_1, \nu_2)$

begin

$\mathbf{x} \leftarrow N(\text{Relax}(A, \mathbf{x}))$ ν_1 times

build P based on $\mathbf{x}$ and A (P is rebuilt every level and cycle)

$A_c = P^T A \,\text{diag}(\mathbf{x})\, P \,\text{diag}(P^T \mathbf{x})^{-1}$

$\mathbf{x}_c = \text{MAA\_V}(A_c, N(P^T \mathbf{x}), \nu_1, \nu_2)$ (coarse-level solve)

$\mathbf{x} = N(\text{diag}(P \,\text{diag}(P^T \mathbf{x})^{-1} \mathbf{x}_c)\, \mathbf{x})$ (coarse-level correction)

$\mathbf{x} \leftarrow N(\text{Relax}(A, \mathbf{x}))$ ν_2 times

end

(Krieger, Horton 1994, but no satisfactory way to build P)

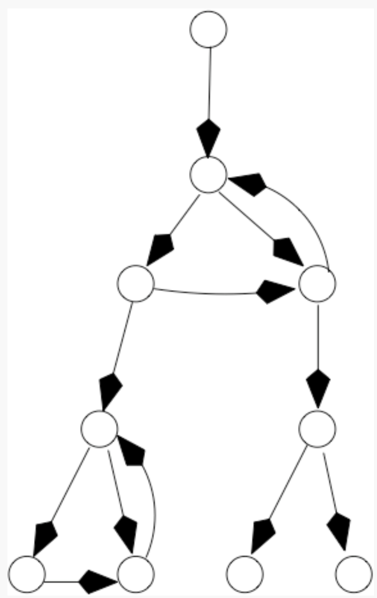
8. Choosing Aggregates Based on Strength

- error equation: $A\,\mathrm{diag}(\mathbf{x}_i)\,\mathbf{e}_i = 0$
- use strength of connection in $A\,\mathrm{diag}(\mathbf{x}_i)$
- define row-based strength (determine all states that strongly influence a row's state)
- state that has largest value in $\mathbf{x}_i$ is seed point for new aggregate, and all unassigned states influenced by it join its aggregate
- repeat

(similar to Brandt, McCormick and Ruge, 1983)

9. Web Matrix Regularization

- web adjacency matrix G, then B=N(G)
- needs to be made irreducible, a-periodic
- need to add in extra links, with 'coupling factor' α



University of Waterloo

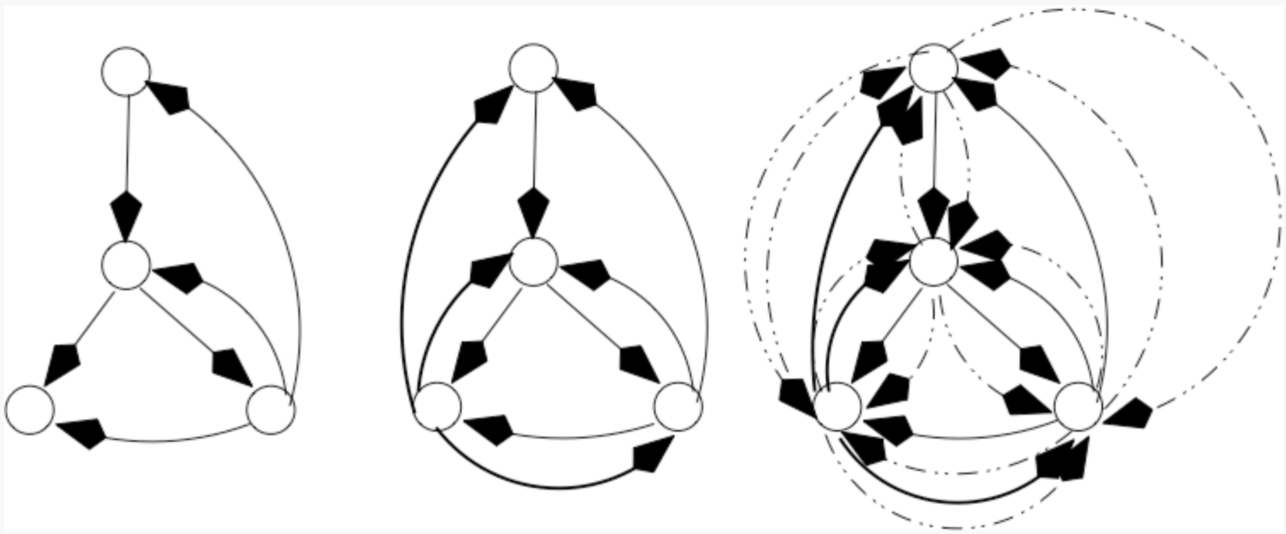
hdesterck@uwaterloo.ca

Web Matrix Regularization

- PageRank (used by Google):

$$B_{PR} = (1-\alpha)\, N(G + \mathbf{e}\,\mathbf{d}^T) + \alpha\, N(\mathbf{e}\,\mathbf{e}^T)$$

$(\alpha = 0.15)$



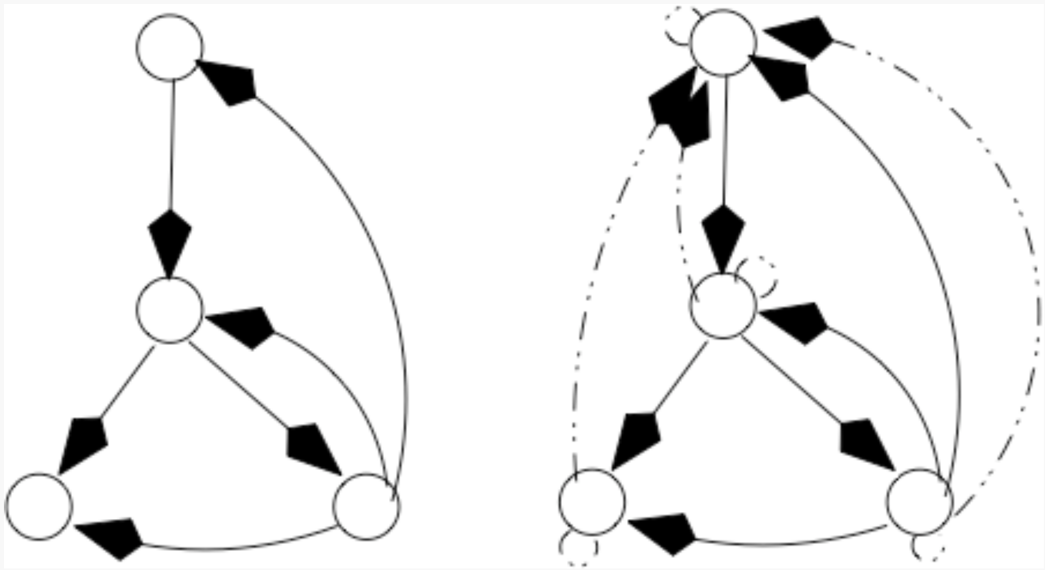
University of Waterloo

UBC CS May 2007
hdesterck@uwaterloo.ca

Web Matrix Regularization

- BackLink (to root page):

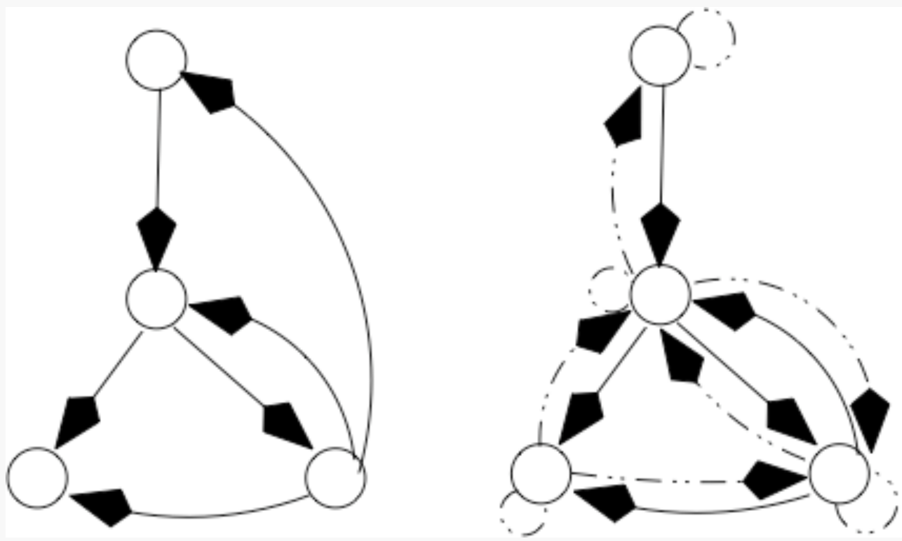
$$B_{BL} = N((1-\alpha-\epsilon)\, N(G) + \alpha\, \mathbf{e}^{(1)}\, \mathbf{e}^T + \epsilon\, I)$$



Web Matrix Regularization

- BackButton (add reverse of each link):

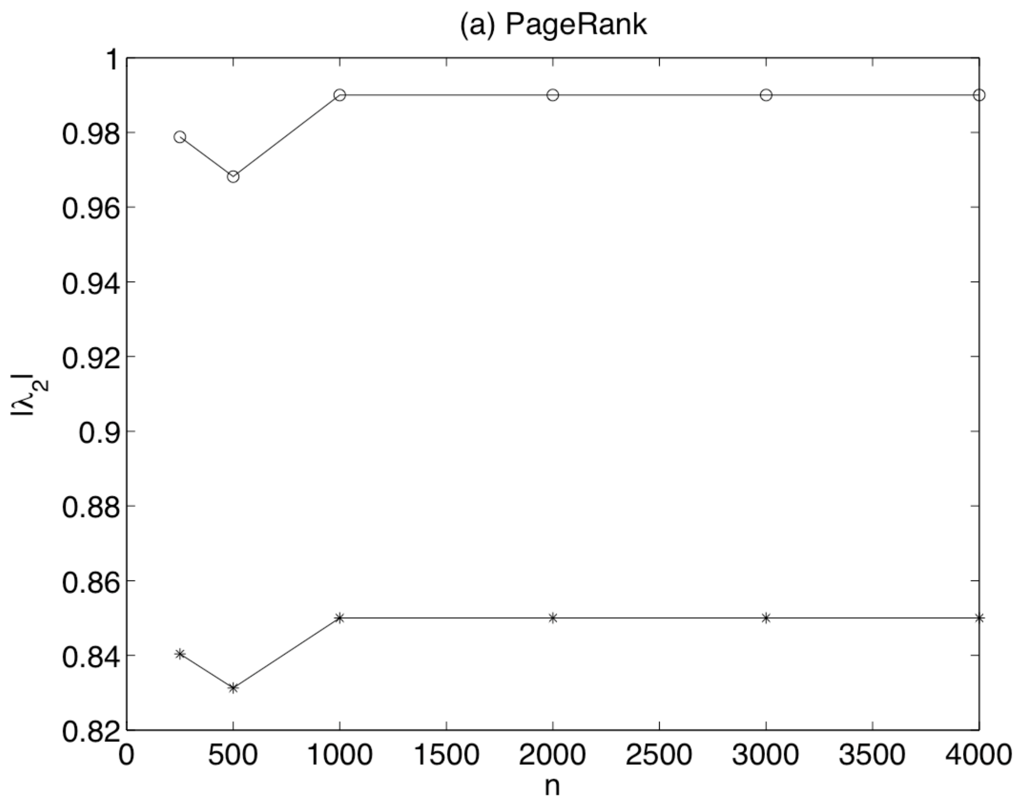
$$B_{BB} = N((1-\alpha-\epsilon)\,N(G) + \alpha\,N(G^T) + \epsilon\,I)$$



University of Waterloo

Web Matrix Regularization

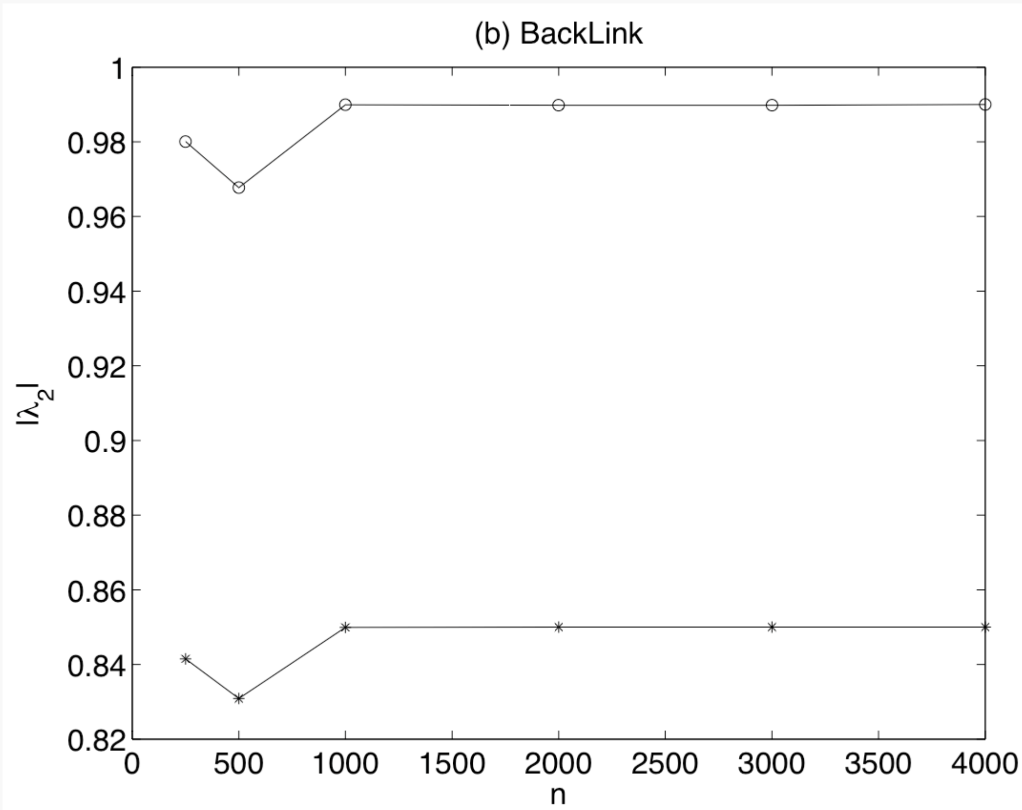
- second eigenvalue for PageRank: ≤ 1-α (α=0.15 and α=0.01)



UBC CS May 2007
hdesterck@uwaterloo.ca

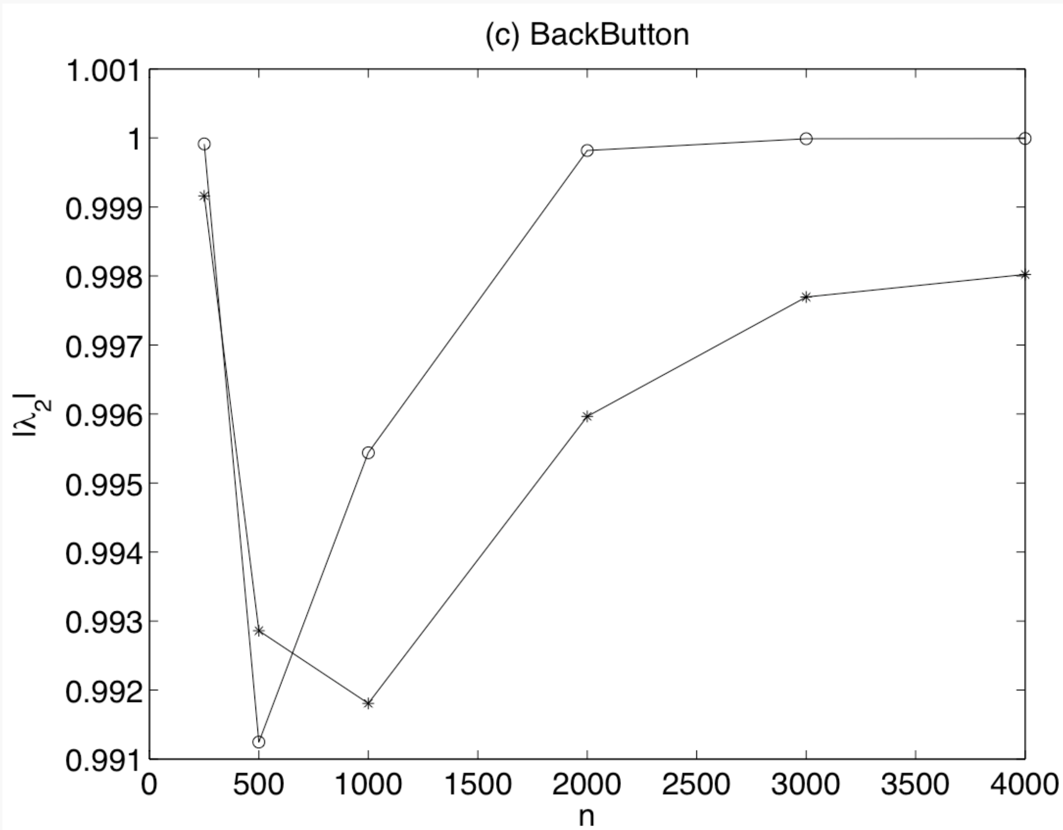
Web Matrix Regularization

- second eigenvalue for BackLink (α=0.15 and α =0.01):



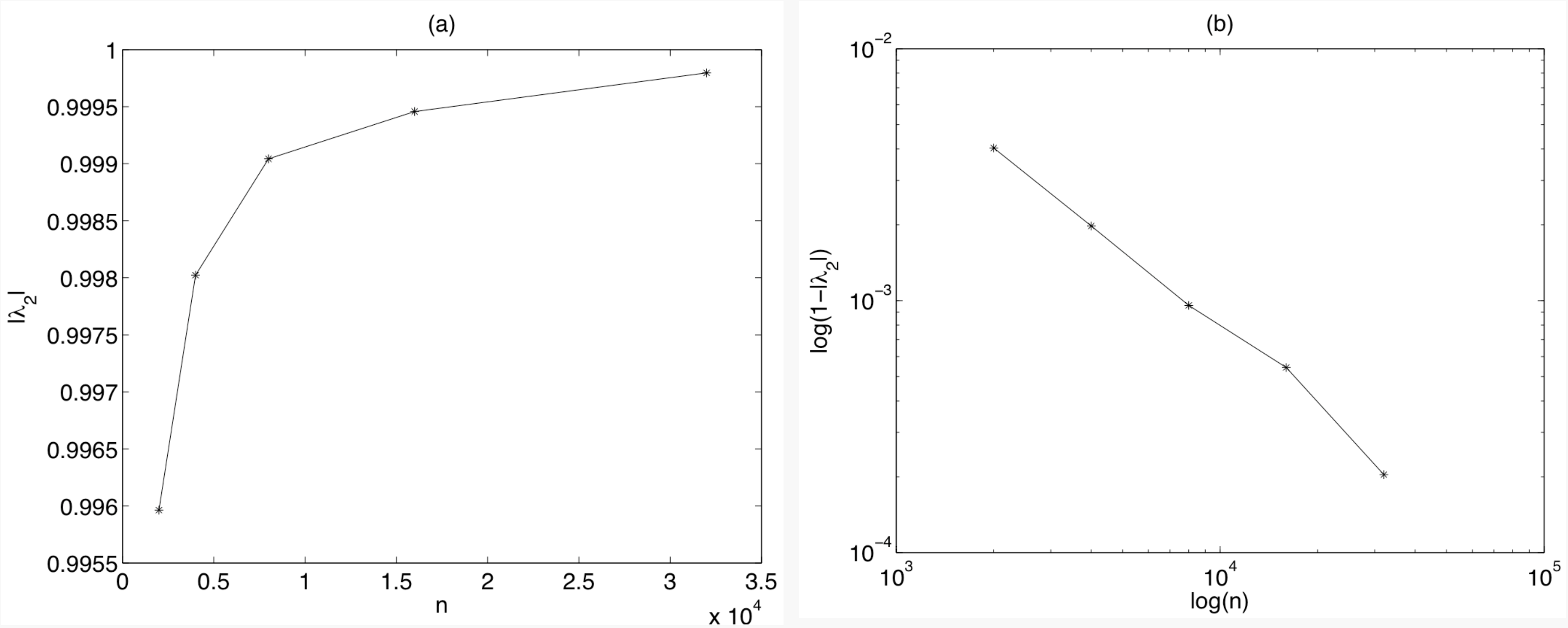
Web Matrix Regularization

- second eigenvalue for BackButton (α=0.15 and α =0.01):



Web Matrix Regularization

- second eigenvalue for BackButton:



$$1-|\lambda_2| \approx O(1/n)$$

University of Waterloo

9. Performance of MAA

- total efficiency factor of MAA relative to WJAC:

$$f^{(tot)}_{MAA-WJAC} = \frac{\log(r_{MAA})/t_{MAA}}{\log(r_{WJAC})/t_{WJAC}},$$

- $f^{(tot)} = 2$: MAA 2 times more efficient than WJAC

- $f^{(tot)} = 1/2$: MAA 2 times *less* efficient than WJAC

UBC CS May 2007
hdesterck@uwaterloo.ca

Performance of MAA

n	γ_{MAA}	it_{MAA}	$c_{grid,MAA}$	γ_{WJAC}	$f^{(tot)}_{MAA-WJAC}$	$f^{(as)}_{MAA-WJAC}$
PageRank, $\alpha = 0.15$						
2000	0.355124	10	1.67	0.815142	1/2.74	1/3.13
4000	0.335889	9	1.67	0.805653	1/2.52	1/3.59
8000	0.387411	9	1.65	0.821903	1/2.79	1/4.14
16000	0.554686	12	1.78	0.836429	1/4.07	1/6.89
32000	0.502008	11	1.83	0.833367	1/3.94	1/6.20
64000	0.508482	11	1.75	0.829696	1/3.86	1/6.21
128000	0.532518	12	1.75	0.829419	1/4.31	1/7.01
PageRank, $\alpha = 0.01$						
2000	0.321062	10	1.77	0.956362	3.42	1.32
4000	0.658754	20	1.75	0.980665	2.16	1.03
8000	0.758825	22	1.65	0.976889	1.88	1/1.65
16000	0.815774	27	1.77	0.979592	1.45	1/2.31
32000	0.797182	29	1.82	0.979881	1.35	1/2.09
64000	0.786973	33	1.79	0.980040	1.19	1/1.96
128000	0.854340	38	1.72	0.980502	1.05	1/2.88

hdesterck@uwaterloo.ca

Performance of MAA

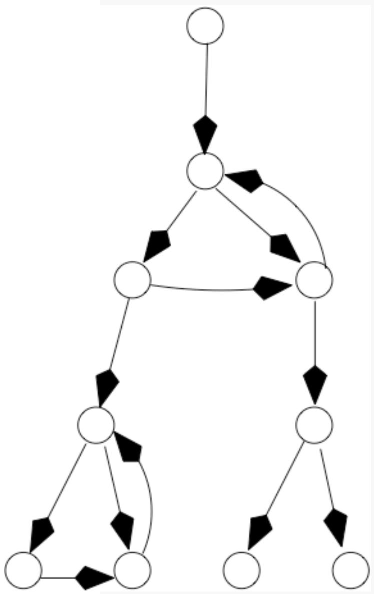
n	γ_{MAA}	it_{MAA}	$c_{grid,MAA}$	γ_{WJAC}	$f^{(tot)}_{MAA-WJAC}$	$f^{(as)}_{MAA-WJAC}$
BackLink, $\alpha = 0.15$						
2000	0.331226	11	1.67	0.839540	1/3.11	1/3.04
4000	0.344225	11	1.75	0.851397	1/3.30	1/3.18
8000	0.361255	11	1.69	0.858532	1/3.04	1/3.24
16000	0.358282	11	2.03	0.866344	1/3.75	1/4.11
32000	0.369351	11	2.26	0.868116	1/3.99	1/4.39
64000	0.368789	11	1.88	0.868889	1/3.30	1/3.53
128000	0.369744	11	1.78	0.871525	1/3.07	1/3.22
BackLink, $\alpha = 0.01$						
2000	0.452383	16	1.89	0.952865	2.01	1/1.21
4000	0.778003	28	1.76	0.953782	1.41	1/4.23
8000	0.749847	20	1.72	0.970096	2.23	1/2.23
16000	0.745776	23	1.96	0.976919	1.87	1/2.11
32000	0.855323	28	1.93	0.981223	1.66	1/3.04
64000	0.868049	32	1.96	0.983076	1.45	1/3.15
128000	0.837747	31	1.83	0.985161	1.65	1/2.09

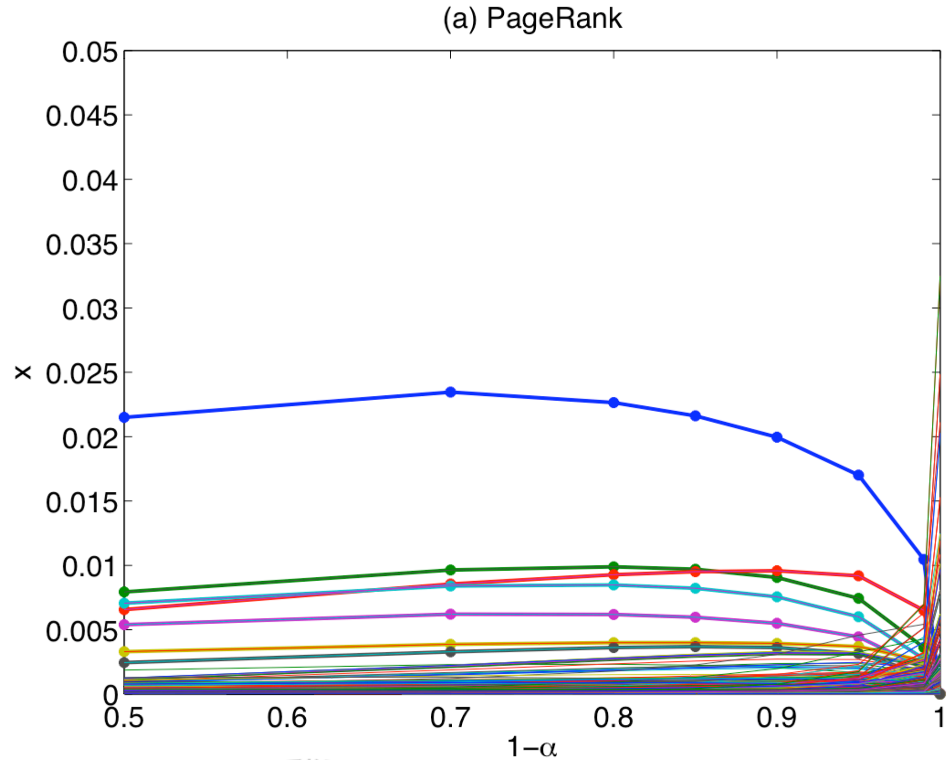
Performance of MAA

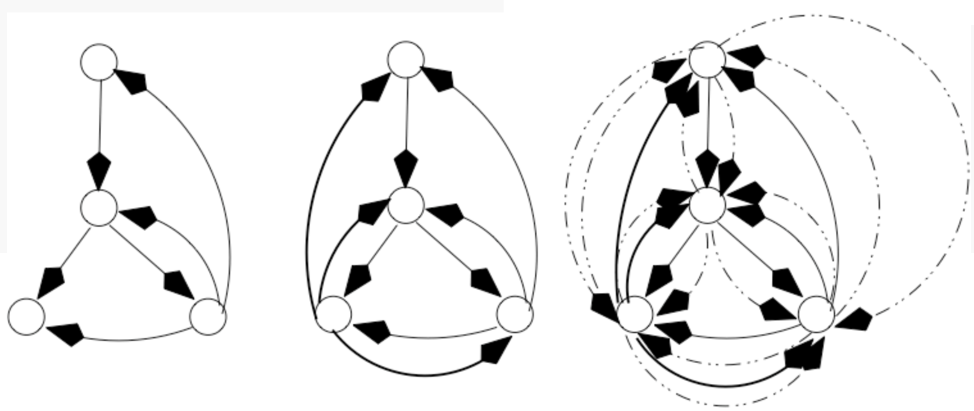
n	γ_{MAA}	it_{MAA}	$c_{grid,MAA}$	γ_{WJAC}	$f^{(tot)}_{MAA-WJAC}$	$f^{(as)}_{MAA-WJAC}$
BackButton, $\alpha = 0.15$						
2000	0.746000	35	1.74	0.981331	2.36	1/1.41
4000	0.800454	39	1.64	0.982828	2.70	1/1.36
8000	0.786758	40	1.53	0.992129	3.15	1.17
16000	0.851671	50	1.62	0.992330	3.00	1/1.38
32000	0.988423	214	1.64	0.998366	4.92	1/2.88
64000	0.973611	185	1.59	0.999013	9.95	1.40
128000	0.943160	116	1.55	0.999693	34.64	9.90
BackButton, $\alpha = 0.01$						
2000	0.658032	23	1.68	0.999563	106.02	46.05
4000	0.794123	29	1.71	0.999345	73.02	19.78
8000	0.841182	39	1.70	0.997624	23.49	2.64
16000	0.835592	44	1.78	0.998696	19.72	4.42
32000	0.845457	56	1.83	0.999114	39.58	8.22
64000	0.959561	81	1.75	0.999660	75.05	5.74
128000	0.921870	42	1.70	0.999963	816.62	103.79

10. Web Matrix Regularizations as a Function of Coupling Factor α

- PageRank

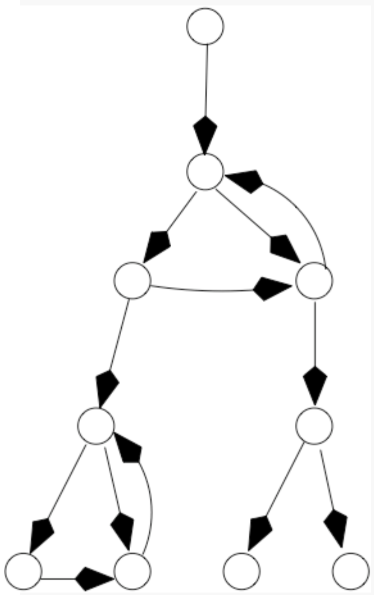


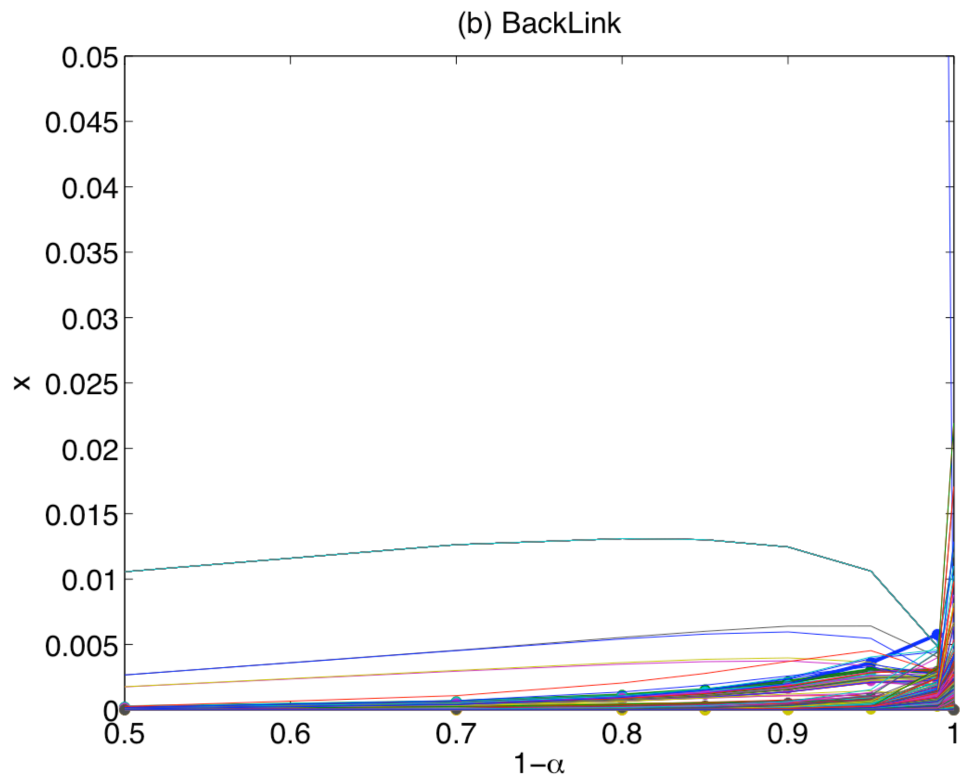


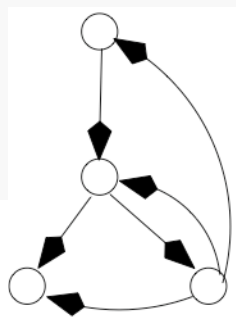


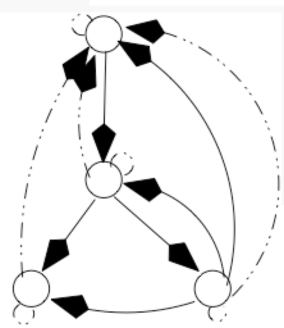
Web Matrix Regularizations as a Function of Coupling Factor α

- BackLink



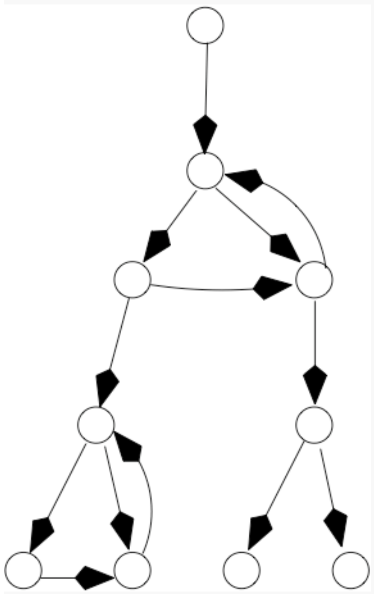


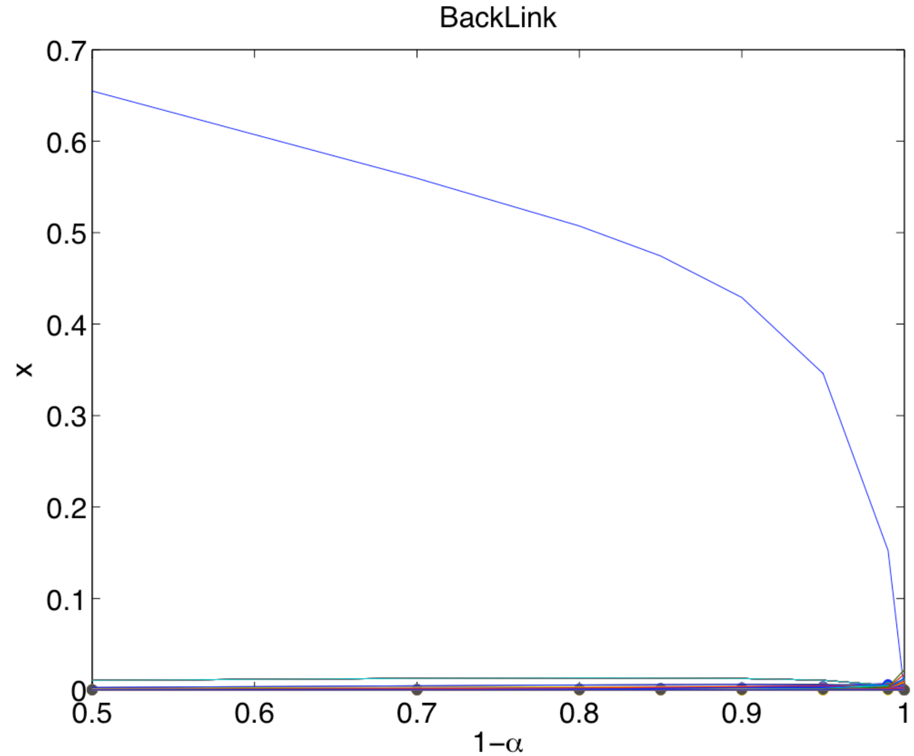


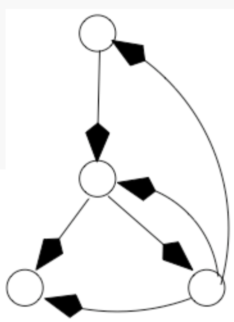


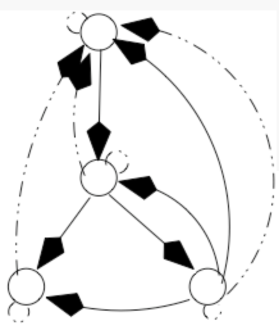
Web Matrix Regularizations as a Function of Coupling Factor α

- BackLink





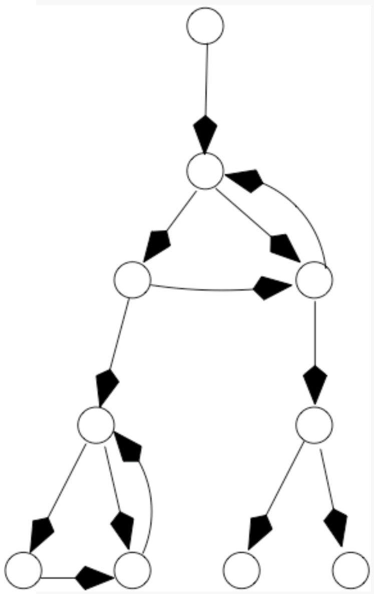


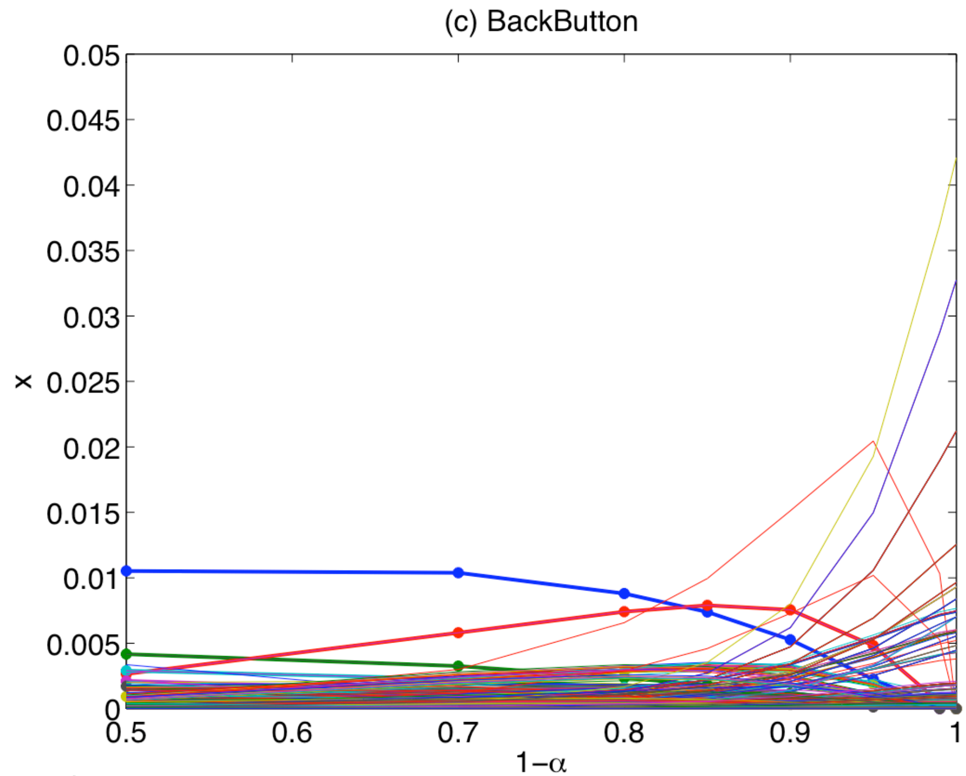


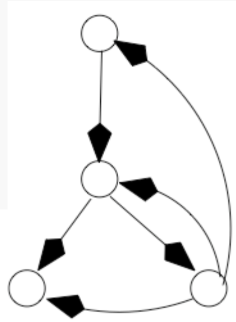
UBC CS May 2007
hdesterck@uwaterloo.ca

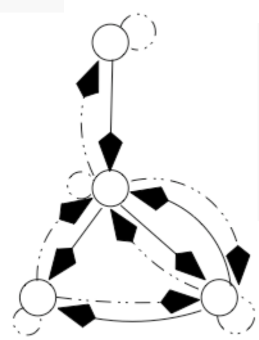
Web Matrix Regularizations as a Function of Coupling Factor α

- BackButton









University of Waterloo

UBC CS May 2007
hdesterck@uwaterloo.ca

11. Conclusions

- PageRank regularization of webgraph with α=0.15 seems to be a pretty good model for web ranking

UBC CS May 2007
hdesterck@uwaterloo.ca

Conclusions

- MAA cannot beat Power Method speed for PageRank

  reasons:

  - λ_2=0.85, independent of n
  - there are global connections on all scales
  - Power Method scales perfectly (also in parallel)
  - $0.85^{10}=0.2$ $0.85^{20}=0.04$

still:

- MAA provides information about how web pages are clustered, at all levels

University of Waterloo

hdesterck@uwaterloo.ca

Conclusions

- MAA can dramatically outperform Power Method for Markov Chains for which $|\lambda_2(n)| \rightarrow 1$ for large n

  reason:

  - multilevel nature of MAA allows to bridge scales (Markov Chain has only local connections in this case)

University of Waterloo

Comparison: Strength-based Aggregation

- major differences with previously considered multilevel methods for Markov chains (Horton, Krieger,...):
  - we use AMG strength of connection based on scaled problem matrix $A\,\mathrm{diag}(\mathbf{x}_i)$
  - our aggregates based on columns of strength matrix (~ AMG coarsening)

- IAD methods
  - normally only 2-level
  - aggregates fixed, based on previously known, regular structure of Markov chain

hdesterck@uwaterloo.ca

Two-Level Acceleration by Aggregation

- multiplicative correction: error equation

$$\mathbf{x} = \text{diag}(\mathbf{x}_i)\,\mathbf{e}_i$$

$$A\,\text{diag}(\mathbf{x}_i)\,\mathbf{e}_i = 0$$

$$P^T\,A\,\text{diag}(\mathbf{x}_i)\,P\,\mathbf{e}_c = 0$$

$$\mathbf{x}_c = \text{diag}(P^T\,\mathbf{x}_i)\,\mathbf{e}_c$$

$$A_c\,\mathbf{x}_c = 0$$

$$P^T\,A\,\text{diag}(\mathbf{x})\,P\,\text{diag}(P^T\,\mathbf{x})^{-1}\,\mathbf{x}_c = 0$$

UBC CS May 2007
hdesterck@uwaterloo.ca