# PageRank Convergence

**Komal Krishna Mogilipalepu**

# PageRank Formulation

- Web is considered as a directed graph with nodes as the pages and edges as the links.
- Connections between the nodes are modelled as a probability transition matrix $\mathbf{X}$.

$\mathbf{X}_{ij}$ - Transition probability from node $j$ to $i$

$\mathbf{p}^{(k)}$- distribution vector on the nodes which is same as the rank of the page

$$\mathbf{p}^{(k+1)} = \mathbf{X}\mathbf{p}^{(k)}$$

**Limitations:**

- Dead ends

  - no outgoing links from a node

- Spider traps

  - periodic graph. Starting from a node we can visit the same node in a fixed time interval.

**Solution:** - teleportation - makes the Markov chain aperiodic and irreducible

$$\mathbf{p}^{(k+1)} = (\beta\mathbf{M} + (1-\beta)\mathbf{t}\mathbf{e}^T)\mathbf{p}^{(k)}$$

# Stationary Distribution

- A stochastic, aperiodic, and irreducible Markov chain has a stationary distribution.

$$\mathbf{p} = (\beta\mathbf{M} + (1-\beta)\mathbf{t}\mathbf{e}^T)\mathbf{p}$$

-here, stationary distribution $\mathbf{p}$ is an eigen vector corresponding to an eigen value 1 of the matrix $(\beta\mathbf{M} + (1-\beta)\mathbf{t}\mathbf{e}^T)$

$$(\mathbf{I} - \beta\mathbf{M})\mathbf{p} = (1-\beta)\mathbf{t}$$

$(\mathbf{I} - \beta\mathbf{M})$ - strictly diagonally dominant $\implies$ *invertible*

- Iterative approach: $\mathbf{p}^{(k+1)} = (\beta\mathbf{M} + (1-\beta)\mathbf{t}\mathbf{e}^T)\mathbf{p}^{(k)}$ where $\mathbf{p}^{(0)} = \mathbf{t}$ or $\mathbf{p}^{(0)} = \mathbf{0}$

- Error: $\mathbf{p} - \mathbf{p}^{(k+1)} = \beta\mathbf{M}(\mathbf{p} - \mathbf{p}^{(k)})$

# Convergence

- If $\mathbf{p}^{(0)} = \mathbf{t}$ then $||\mathbf{p} - \mathbf{p}^{(k)}||_1 = \beta^k ||\mathbf{M}^k(\mathbf{p} - \mathbf{t})||_1 \leq ||\mathbf{p} - \mathbf{t}||_1 \beta^k \leq 2\beta^k$

  **Proof:** uses the following ideas

  - product of stochastic matrices are stochastic
  - triangle inequality
  - for any real value: $|x - y| \leq |x| + |y|$

- If $\mathbf{p}^{(0)} = \mathbf{0}$ then $\mathbf{p} - \mathbf{p}^{(k)} \geq 0 \quad \forall k$ and $||\mathbf{p} - \mathbf{p}^{(k)}||_1 = \beta^k$

**Remark:** It shows that zero initialization giving less error compared to random initialization, however practical experience suggests that random initialization results in faster convergence.

- it could be confirmed by computing the error by bounding it by residual.

# pseudo-PageRank

- Let $\bar{\mathbf{M}}$ be a column-substochastic matrix with $\bar{\mathbf{M}}_{ij} \geq 0$ and $\mathbf{e}^T \bar{\mathbf{M}} \leq \mathbf{e}^T$ element-wise.
- Let $\mathbf{f}$ be a nonnegative vector, and let $0 < \beta < 1$ be a teleportation parameter.
- Then the pseudo-PageRank problem is to find the solution of the linear system

  $(\mathbf{I} - \beta \bar{\mathbf{M}})\mathbf{y} = \mathbf{f}$ where $\mathbf{y}$ pseudo-PageRank vector.

- Let $\mathbf{t} = \mathbf{f}/(\mathbf{e}^T \mathbf{f})$ then $\mathbf{p} = \mathbf{y}/(\mathbf{e}^T \mathbf{y})$ is the solution of a PageRank system with $\beta$, $\mathbf{M} = \bar{\mathbf{M}} + \mathbf{t}\mathbf{c}^T$, and $\mathbf{t}$, where $\mathbf{c}^T = \mathbf{e}^T - \mathbf{e}^T \bar{\mathbf{M}} \geq 0$ is a correction vector.

- Similar to PageRank, we can show the convergence guarantees for pseudo-PageRank.

**References:**


- https://www.cs.purdue.edu/homes/dgleich/publications/Gleich%202015%20-%20prbeyond.pdf
- https://www.youtube.com/watch?v=UZePPh340sU&t=329s