PageRank, connecting a line of nodes with multiple complete graphs

Pitos Seleka Biganda

Department of Mathematics, University of Dar-es-Salaam

Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University

Second Network Meeting for Sida- and ISP-funded PhD Students in Mathematics

Stockholm 26–27 February 2018

LINKÖPING UNIVERSITY

UPPSALA UNIVERSITET

Sida

My Advisors

Sergei Silvestrov	Christopher Engström	Nyimvua Shaban
Main advisor	*Assistant advisor*	*Assistant advisor*
Mälardalen University	*Mälardalen University*	*University of Dar-es-Salaam*

MÄLARDALEN UNIVERSITY SWEDEN

Introduction

- PageRank is a method used to rank nodes in different link structure such as Internet webpages in order of importance (Brin and Page, 1998).
- However, other uses of PageRank or similar methods exist, e.g. the EigenTrust algorithm for reputation management to decrease distribution of unauthentic files in P2P networks (Kamvar & Schlosser, 2003).

Introduction

- PageRank is usually calculated using the Power method. The method has been proved to be efficient for any size of system (Haveliwala & Kamvar, 2003).
- A number of works has been done to improve the computation time of PageRank to accommodate the huge number of pages (nodes) on the web: Ishii *et al* (2009), Sepandar, Taher & Gene (2003), and Anderson & Silvestrov (2006).

Introduction

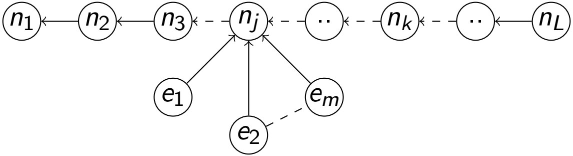
- The theory behind PageRank is built on Perron-Frobenius theory of non-negative irreducible matrices and the theory of Markov chains.
- Studies on how PageRank is affected from changes in the system is needed.
- We extended the work by Engström and Silvestrov (2016) by looking at how PageRank changes when a line of nodes is connected with a multiple number of outside nodes and complete graphs.

- **Definition**: PageRank $\vec{R}^{(2)}$ of a node when using uniform $\vec{u}$ can be written (Engström and Silvestrov, 2016):

$$\vec{R}_i^{(2)} = \left(\sum_{e_i \in S, e_i \neq e_j} w_j P(e_j \to e_i) + w_i \right) \left(\sum_{k=0}^{\infty} (P(e_i \to e_i))^k \right)$$

where $P(e_j \to e_i)$ is the probability to hit node e_i in a random walk starting in node e_j described as above. This can be seen as the expected number of visits to e_j if we do multiple random walks, starting in every node once.

A simple line with multiple links from m outside nodes to one node in a line

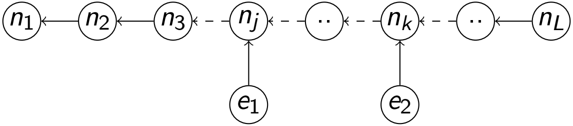


Theorem (1)

$$\begin{aligned} \vec{R}_i^{(2)} &= \sum_{k=0}^{n_L - i} c^k + b_{ij} = \frac{1 - c^{n_L - i + 1}}{1 - c} + b_{ij}, \\ b_{ij} &= \begin{cases} mc^{j-i+1}, & \textit{if } i \leq j \\ 0, & \textit{if } i > j \end{cases} \end{aligned} \tag{1}$$

where $m \geq 1$, n_L is the number of nodes in the line. The outer nodes each have rank 1.

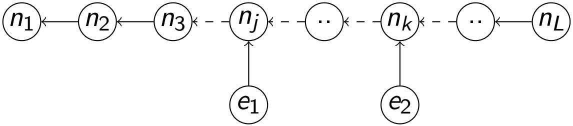
A simple line with two outside vertices linked to two nodes in the line



Theorem (2)

$$\vec{R}_i^{(2)} = \frac{1 - c^{n_L - i + 1}}{1 - c} + b_{ik} \quad \textit{for} \quad i \geq k$$
$$b_{ik} = \begin{cases} c, & \textit{if } i = k \\ 0, & \textit{if } i > k \end{cases}, \tag{2}$$

A simple line with two outside vertices linked to two nodes in the line



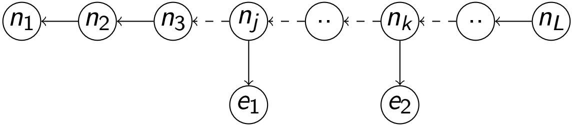
Theorem (2)

$$\vec{R}_i^{(2)} = \frac{1 - c^{n_L - i + 1}}{1 - c} + c^{k-i+1} + b_{ij} \quad \textit{for} \quad j \le i < k$$

$$b_{ij} = \begin{cases} c, & \textit{if } i = j \\ 0, & \textit{if } i > j \end{cases}, \tag{3}$$

$$\vec{R}_i^{(2)} = \frac{1 - c^{n_L - i + 1}}{1 - c} + c^{1-i}\left(c^j + c^k\right), \quad \textit{for} \quad i < j, \tag{4}$$

A simple line with two links from the line to two outside nodes

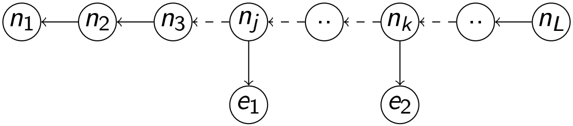


Theorem (3)

$$\vec{R}_i^{(2)} = \sum_{m=i}^{n_L} c^{m-i} = \frac{1 - c^{n_L - i + 1}}{1 - c}, \quad \textit{for} \quad i \geq k, \tag{5}$$

$$\begin{aligned} \vec{R}_i^{(2)} &= \sum_{m=i+1}^{k} c^{m-i-1} + \frac{1}{2} \sum_{m=k}^{n_L} c^{m-i} \\ &= \frac{2 - c^{k-i}\left(1 + c^{n_L - k + 1}\right)}{2(1-c)}, \quad \textit{for} \quad j \leq i < k, \end{aligned} \tag{6}$$

A simple line with two links from the line to two outside nodes

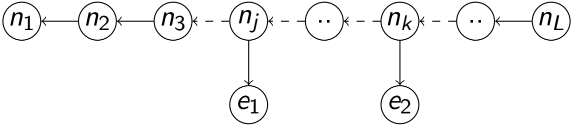


Theorem (3)

$$\vec{R}_i^{(2)} = \sum_{m=i+1}^{j} c^{m-i-1} + \frac{1}{2}\sum_{m=j}^{k-1} c^{m-i} + \frac{1}{4}\sum_{m=k}^{n_L} c^{m-i}, \quad \textit{for} \quad i < j$$
$$= \frac{1 - c^{k-i} - c^{j-i}\left(2 + c^{n_L-j+1}\right)}{4(1-c)}, \qquad (7)$$

where n_L is the number of nodes in the line.

A simple line with two links from the line to two outside nodes



Theorem (3)

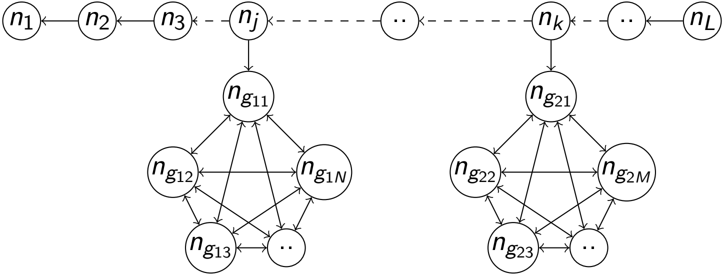
The PageRank of the new nodes e_1 and e_2 are respectively,

$$\vec{R}_{e_1}^{(2)} = 1 + \frac{1}{2}c\left(\frac{1-c^{k-j}}{1-c}\right) + \frac{1}{4}c^{k-j+1}\left(\frac{1-c^{n_L-k+1}}{1-c}\right) \quad (8)$$

and

$$\vec{R}_{e_2}^{(2)} = 1 + \frac{1}{2}c\left(\frac{1-c^{n_L-k+1}}{1-c}\right). \quad (9)$$

A simple line with two links from the line to two complete graphs



A simple line with two links from the line to two complete graphs

Theorem (4)

$$\vec{R}_{L,i} = \sum_{m=i}^{n_L} c^{m-i} = \frac{1 - c^{n_L - i + 1}}{1 - c}, \quad \textit{for} \quad i \geq k, \tag{10}$$

$$\begin{aligned} \vec{R}_{L,i} &= \sum_{m=i+1}^{k} c^{m-i-1} + \frac{1}{2} \sum_{m=k}^{n_L} c^{m-i}, \quad \textit{for} \quad j \leq i < k \\ &= \frac{2 - c^{k-i}\left(1 + c^{n_L - k + 1}\right)}{2(1 - c)}, \quad \textit{for} \quad j \leq i < k, \end{aligned} \tag{11}$$

A simple line with two links from the line to two complete graphs

Theorem (4)

$$\begin{aligned} \vec{R}_{L,i} &= \sum_{m=i+1}^{j} c^{m-i-1} + \frac{1}{2}\sum_{m=j}^{k-1} c^{m-i} + \frac{1}{4}\sum_{m=k}^{n_L} c^{m-i}, \quad for \quad i < j \\ &= \frac{4 - c^{k-i} - c^{j-i}\left(2 + c^{n_L-j+1}\right)}{4(1-c)}, \end{aligned} \tag{12}$$

A simple line with two links from the line to two complete graphs

Theorem (4)

The PageRank of the nodes in the complete graphs are given by

$$\vec{R}_{G_2,g_k} = \frac{c}{2}\left(\frac{1-c^{n_L-k+1}}{1-c}\right)\left(\frac{(n_{G_2}-1)-c(n_{G_2}-2)}{(n_{G_2}-1)-c(n_{G_2}-2)-c^2}\right)+\frac{1}{1-c}$$

$$\vec{R}_{G_2,i} = \left(\frac{c^2(1-c^{n_L-k+1})}{2(1-c)}\right)\left(\frac{1}{(n_{G_2}-1)-c(n_{G_2}-2)-c^2}\right)+\frac{1}{1-c}$$

$$\vec{R}_{G_1,g_j} = \left[\frac{2c-c^{k-j+1}-c^{n_L-j+2}}{4(1-c)}\right]\left[\frac{(n_{G_1}-1)-c(n_{G_1}-2)}{(n_{G_1}-1)-c(n_{G_1}-2)-c^2}\right]+\frac{1}{1-c}$$

$$\vec{R}_{G_1,i} = \left[\frac{2c^2-c^{k-j+2}-c^{n_L-j+3}}{4(1-c)}\right]\left[\frac{1}{(n_{G_1}-1)-c(n_{G_1}-2)-c^2}\right]+\frac{1}{1-c}$$

References

- Brin, S. and Page, L.: The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* **30**(1-7), 107-117 (1998). Proceedings of the Seventh International World Wide Web Conference.
- Kamvar, S.D. and Haveliwala, T.: The Condition Number of the PageRank Problem. *Technical Report*, Stanford InfoLab, (2003).
- Kamvar, S.D., Schlosser, M.T. and Garcia-Molina, H.: The eigentrust algorithm for reputation management in p2p networks. In: Proceedings of the 12th international conference on World Wide Web, WWW '03, pp. 640-651 (2003).
- Engström, C. and Silvestrov, S.: PageRank, connecting a line of nodes with a complete graph. In: Silvestrov S., Rančić M. (eds) Engineering Mathematics II. *Springer Proceedings in Mathematics & Statistics*, vol 179. Springer, Cham, (2016).

Tack så mycket!

Thank you!