Beyond Google's PageRank: A Novel Link Analysis Algorithm without a Damping Factor

INSNA 2020 Keita Sugihara, Nanzan University

1. Introduction

PageRank adopts a damping factor to gives a high score to a page with (i) an in-link from a high score page, (ii) many in-links, and, (3) selected in-links. This factor causes the problem of inconsistent ranking. HermitianStatus is an alternative to it.

2. PageRank

- Equation (1) of the adjacency matrix of the network: to derive page scores reflecting (i) and (ii).
- However, a solution is not assured because the network may not be strongly connected.

$$A^T R = |\lambda|_1 R, \quad \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} = \frac{1}{|\lambda|_1} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{N1} \\ a_{12} & a_{22} & \cdots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & a_{2N} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} \quad (1)$$

- Equation (2) of the matrix created from A^T: derives node scores reflecting (i), (ii), and, (iii).
- The damping factor d factor in (2) realizes the strongly connected graph and a solution is assured.
- The factor causes the problem of inconsistent ranking with a changing damping factor value, as shown in Fig. 2 .

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} = \frac{1}{|\lambda|_1} \left(d \times \left(\begin{bmatrix} a_{11} & 0 & \dots & a_{N1} \\ a_{12} & 0 & \dots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & 0 & \dots & a_{NN} \end{bmatrix} \begin{bmatrix} 1/\Sigma_j a_{1j} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 1/\Sigma_j a_{Nj} \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{N} & \dots & 0 \\ 0 & \frac{1}{N} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{1}{N} & 0 & 0 \end{bmatrix} \right) + (1-d) \times \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \dots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \dots & \frac{1}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N} & \frac{1}{N} & \dots & \frac{1}{N} \end{bmatrix} \right) \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} \quad (2)$$

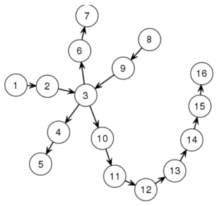

Fig. 1 Weakly Connected Network

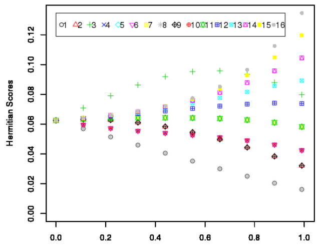

Fig. 2: Ranking of the nodes of the network with a changing damping a factor value in Fig. 1

3. HermitianStatus

- Equation (3) of the Hermitian adjacency matrix : to derive node scores reflecting (i) and (ii).
- In (3), a solution is experimentally assured if the network is weakly connected.
- However, (i) and (ii) are not assured when a network has a lot of nodes.
- HermitianStatus employs parameters s and t, as well as k_1, k_2, k_3, and, k_4 to realize (iii), and ,correspond the above problem.
- The algorithm realizes the consistent Ranking in Fig 3. Also, Hermitian Status reproduces the PageRank ranking in Fig. 4.

$$H X = |\lambda|_1 X \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \frac{1}{|\lambda|_1} \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1N} \\ h_{21} & h_{22} & \dots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \dots & h_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad (3)$$

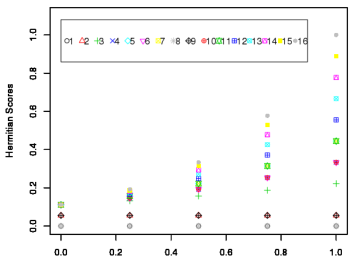

Fig. 3: Ranking of the nodes of the network with a changing k1 with fixed k2 and k3 in Fig. 1

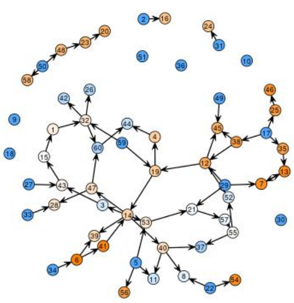

Fig. 4: Network nodes ranking (high: orange, low: blue). Spearman's correlation coefficient between PageRank scores at the damping factor 0.85 and HermitianStatus scores with specific *k1*, *k2*, *k3*, and, *k4*. is 0.9268808.

Patent: JP6502592B1(PCT/JP2018/02656)
Patent application: 2019-230822

References

K. J. Guo, 2015. Simple eigenvalues of graphs and digraphs. Dissertation, the Department of Mathematics of the Faculty of Science of Simon Fraser University.

K., Sugihara, 2019. Using Complex Numbers in Website Ranking Calculations: A Non-ad hoc Alternative to Google's PageRank. Journal of Software 14, 58-64.

4. Discussion

- The future research will focus on calculation cost efficiency of HermitianStatus over PageRank.