On Computing Prestige in a Network with Negative Relations

Allen H. Tai
The Hong Kong Polytechnic University

Wai-Ki Ching
The University of Hong Kong

Wai-Shun Cheung
The University of Hong Kong

International Journal of Applied Mathematical Sciences,
ISSN 0973-0176 Vol.2 No.1(2005), pp. 97-104

http://www.gbspublisher.com/ijams.htm

On Computing Prestige in a Network with Negative Relations

Allen H. Tai, Wai-Ki Ching*, Wai-Shun Cheung[1]

Department of Mathematics, University of Hong Kong,
Pokfulam Road, Hong Kong, China.
Email:h0150695@graduate.hku.hk, wkc@maths.hku.hk, wshun@graduate.hku.hk

Abstract

The computation of network prestige is an important issue in studying networks such as the Internet and social networks. A number of iterative methods have been proposed for the measurement of prestige of symmetric or asymmetric network. The PageRank algorithm has been successfully applied in the computation of ranking of webpages and data mining in the Internet. In this paper, we propose a revised PageRank algorithm for the computation of prestige of a general networks and extend it to handle the case of negative relations.

Key Words. Markov chain, Networks, Prestige, Negative Relations.
Subject Classifications. 65C20, 65F10.

1. Introduction

The measurement of prestige in a network is an important issue [4, 15] and it has many applications such as the study of ranking of webpages [11, 12], social networks [4, 15] and disease transmission [3]. A number of methods based on the computation of eigenvectors have been proposed in the literatures [4, 15]. In a network, being chosen or nominated by a popular or powerful person would add one's popularity [4]. In the following, we give an example of a social network of five persons [4].

In the network, 3 is chosen (or supported) by 1 and 2, 4 is chosen by 3 and 5 is chosen by 4. The relations can be represented by using an 5×5 adjacency

[1]Research supported by RGC Grant No. HKU 7126/02P and HKU CRCG Grant No. 10204436, 10203919.

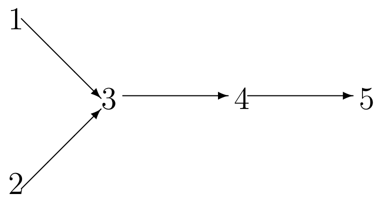

Figure 1: The Social Network.

matrix A, where A_{ij} is 1 if i is chosen (supported) by j otherwise it is zero. In this case, the adjacency matrix is given by

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

In this paper, we propose a PageRank type algorithm to compute the prestige of general social networks and extend it to handle the case of negative relations.

In World-Wide-Web, very often surfers use search engines (e.g. Google, Yahoo) to find the related webpages satisfying their queries. Therefore a proper list of the webpages in certain order of importance is very useful and this can be achieved by using the famous PageRank algorithm. PageRank has been proposed by Page et al. (1998) [11, 13] to reflect the importance of the webpages. In fact, a similar idea was first proposed by Garfield (1955, 1972) [8, 9] as a measure of standing for journals, which is called the *impact factor*. The impact factor of a journal is defined as the average number of citations per recently published papers in that journal. In the PageRank algorithm, we let N be the total number of webpages in the web and we define a matrix Q called the *hyperlink matrix* where

$$Q_{ij} = \begin{cases} 1/k & \text{if webpage } i \text{ is an outgoing link of webpage } j; \\ 0 & \text{otherwise;} \end{cases}$$

and k is the total number of outgoing links of webpage j, see for instance [6, 13]. Hence Q can be regarded as a transition probability matrix of a Markov chain of a random walk. Assuming that this underlying Markov chain is irreducible, then the steady-state probability distribution $(p_1, p_2, \ldots, p_N)^T$ of the states (webpages) exists. Here p_i is the proportion of time that the random walker (surfer) visiting state (webpage) i. The higher the value of p_i is, the more important the webpage i will be. Thus the PageRank of webpage i can be defined as p_i. If the Markov chain is reducible then one can follow the treatment in Section 2. Since the size of the Markov chain is huge, direct method for solving the steady-state probability is not desirable. Iterative methods proposed by Baldi et ak. (2003)

[2] and decomposition methods proposed by Avrachenkov and Litvak (2004) [1] have been applied to solve the problem. Other iterative methods such as the hybrid iterative methods have been proposed in Yuen et al. (2004) [16] and Ching et al. (2004) [6].

The remainder of this paper is organized as follows. In Section 2, we give a brief mathematical discussion on the PageRank algorithm. In Section 3, we describe our revised PageRank algorithm for the computation of prestige with negative relations. Finally concluding remarks are given in Section 4.

2. The PageRank Algorithm

In this section, we briefly review the PageRank Algorithm. The PageRank algorithm has been successfully applied in ranking the importance of webpages by a lot of search engines such as Google [11]. We consider a web of N webpages with Q being the hyperlink matrix. Since the matrix Q can be reducible, very often one would consider the following revised matrix P ([6, 13]):

$$P = \alpha \begin{pmatrix} 0 & Q_{12} & \cdots & Q_{1N} \\ Q_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Q_{N-1N} \\ Q_{N1} & \cdots & Q_{NN-1} & 0 \end{pmatrix} + \frac{(1-\alpha)}{N} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \quad (1)$$

where $0 < \alpha < 1$. In this case, the matrix P is irreducible and aperiodic, therefore the steady state probability distribution exists and is unique [14]. Some typical values for α are 0.85 and $(1 - 1/N)$, see for instance [2, 11, 12]. In fact, One can interpret (1) as follows. For a network of N webpages, each webpage has an inherent importance of $(1 - \alpha)/N$ (depending on N only). If a webpage P_i has an importance of p_i, then it will contribute an importance of αp_i which is shared among the webpages that it points to. Therefore the importance of webpage P_i can be obtained by solving the following linear system of equations subject to the normalization constraint:

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} = \alpha \begin{pmatrix} 0 & Q_{12} & \cdots & Q_{1N} \\ Q_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Q_{N-1N} \\ Q_{N1} & \cdots & Q_{N,N-1} & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1/N \\ 1/N \\ \vdots \\ \vdots \\ 1/N \end{pmatrix}. \quad (2)$$

Since

$$\sum_{i=1}^{N} p_i = 1,$$

(2) can be re-written as

$$(p_1, p_2, \ldots, p_N)^T = P(p_1, p_2, \ldots, p_N)^T.$$

In fact, $(p_1, p_2, \ldots, p_N)^T$ is a right hand eigenvector of the matrix P with eigenvalue 1.

Example 1 *If the PageRank algorithm is applied to the social network in Figure 1 of Section 1, we have*

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \alpha \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{pmatrix}.$$

We then obtain the following results

$$(p_1, p_2, p_3, p_4, p_5) = (\frac{1-\alpha}{5}, \frac{1-\alpha}{5}, \frac{1+\alpha-2\alpha^2}{5}, \frac{1+\alpha^2-2\alpha^3}{5}, \frac{1+\alpha^3-2\alpha^4}{5}).$$

If $\alpha = 0.85$ *or* $1 - 1/5 = 0.8$, *we have the following ranking of prestige*

$$1 = 2 < 3 < 4 < 5.$$

The PageRank algorithm here does not take into account the variation of inherent importance of individual members. One may expect that the importance of members 1 and 2 are zero. Moreover, the current setting cannot handle the case when there are negative relations among the members [5]. A negative relation from member 1 to member 2 means member 1 is against member 2. We remark that an algorithm has been proposed for computing the centrality of a social network with negative relations in [5].

3. A Revised PageRank Algorithm for Networks with Negative Relations

In this section, we present the idea of the revised PageRank algorithm for a network of N members. We will first consider the case of the variation of inherent importance of individual member. We then take into account the negative relation by creating an artificial member.

Each member has an inherent importance of $(1-\alpha)I_i$ (instead of a constant in PageRank algorithm). If a member P_i has an importance of p_i, then it will contribute an importance of αp_i which is shared among the members that it pointed to. Therefore the importance of member P_i can be obtained by solving the following linear system of equations $\mathbf{p} = \alpha Q \mathbf{p} + (1-\alpha)\mathbf{b}$:

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} = \alpha \begin{pmatrix} 0 & Q_{12} & \cdots & Q_{1N} \\ Q_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Q_{N-1N} \\ Q_{N1} & \cdots & Q_{NN-1} & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} + (1-\alpha) \begin{pmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{pmatrix} \quad (3)$$

where α is a weighting factor such that $\alpha \in (0,1)$ and $\mathbf{b} = (I_1, I_2, \ldots, I_N)^T$. Moreover, we assume that the inherent popularity of member i is proportional to the proportion of all the connections pointing to i, i.e.

$$I_i = \begin{cases} \dfrac{\sum\limits_{j=1}^{N} Q_{ij}}{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} Q_{ij}} & \text{if } \sum\limits_{j=1}^{N} Q_{ij} \neq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 1 *By applying the Gershgorin disc theorem [10], the matrix $(I - \alpha Q)$ is invertible for $\alpha \in (0,1)$. Hence equation (3) has a unique solution.*

We then apply the revised PageRank algorithm to the network in Section 1 and get

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \alpha \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 0 \\ 0 \\ 0.5 \\ 0.25 \\ 0.25 \end{pmatrix}.$$

Solving the linear system, we obtain

$$(p_1, p_2, p_3, p_4, p_5) = (0, 0, \frac{1-\alpha}{2}, \frac{1+\alpha-2\alpha^2}{4}, \frac{1+\alpha^2-2\alpha^3}{4})$$

and $p_1 = p_2 = 0$ for all values of $\alpha \in (0,1)$. If $\alpha < 0.5$, the ranking is given by

$$3 > 4 > 5 > 1 = 2$$

and if $\alpha \geq 0.5$, and the new ranking is given by

$$5 > 4 > 3 > 1 = 2.$$

This is due to the fact that member 3 has the highest in-degree (inherent importance). Therefore if the value of α is not large, member 3 is dominant.

We then extend the revised PageRank algorithm to the computation of prestige when there are negative relations [5]. The idea here is to introduce an extra artificial member $(N+1)$ representing the aggregated effect of negative relations. In the $(N+1) \times (N+1)$ matrix, for $i < N+1$, Q_{iN+1} is the proportion of negative relations that member i got in the network. For $i < N+1$, Q_{N+1i} is the proportion of negative relations given out by member i. Assuming the effect

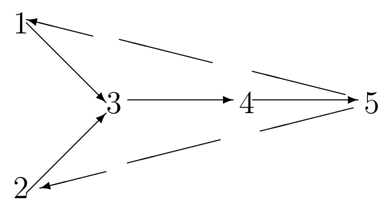

Figure 2: The Network with Negative Relations.

of a positive relation and a negative relation are the same, one may set up the following equations:

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \\ p_{N+1} \end{pmatrix} = \alpha \begin{pmatrix} 0 & Q_{12} & \cdots & Q_{1N+1} \\ Q_{21} & 0 & \cdots & Q_{2N+1} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{N1} & Q_{N2} & \cdots & Q_{NN+1} \\ Q_{N+11} & \cdots & Q_{N+1N} & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \\ p_{N+1} \end{pmatrix} + (1-\alpha) \begin{pmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \\ I_{N+1} \end{pmatrix} \quad (4)$$

Proposition 2 *By the Gershgorin disc theorem again, the matrix $(I - \alpha Q)$ in (4) is invertible for $\alpha \in (0, 1)$. Hence equation (4) has a unique solution.*

Example 2 *Suppose there are negative relations pointing from 5 to 1 and 2 (See Figure 2), then one can add an artificial member 6.*

The matrix system is then given by

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{pmatrix} = \alpha \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 0 \\ 0 \\ 1/3 \\ 1/6 \\ 1/6 \\ 1/3 \end{pmatrix}.$$

Solving the linear system, we have

$$\begin{cases} p_1 = \dfrac{-2\alpha^5 + \alpha^4 - \alpha^2 + 2\alpha}{12(1-\alpha^5)}, \\ p_2 = \dfrac{-2\alpha^5 + \alpha^4 - \alpha^2 + 2\alpha}{12(1-\alpha^5)}, \\ p_3 = \dfrac{-\alpha^5 - \alpha^3 + 2\alpha^2 - 2\alpha + 2}{6(1-\alpha^5)}, \\ p_4 = \dfrac{-\alpha^5 - \alpha^4 + 2\alpha^3 - 2\alpha^2 + \alpha + 1}{6(1-\alpha^5)}, \\ p_5 = \dfrac{-2\alpha^5 + 2\alpha^4 - 2\alpha^3 + \alpha^2 + 1}{6(1-\alpha^5)}, \\ p_6 = \dfrac{-2\alpha^4 + \alpha^3 - \alpha + 2}{6(1-\alpha^5)} \end{cases}$$

and for $0 < \alpha < 1$*, it can be shown that we have the following prestige ranking*

$$6 > 3 > 4 > 5 > 2 = 1.$$

The importance of members 1 *and* 2 *are due to the negative relation from* 5 *and they share the importance of the artificial member* 6 *which is the highest in this network. With the support of both members* 1 *and* 2*, member* 3 *has the highest ranking among all the real members follows by member* 4 *and then member* 5*.*

4. Concluding Remarks

In this paper, we present a revised PageRank algorithm for the computation of the prestige of a general network. Extension to the case of negative relations is also discussed. Finally we remark that if the network is very large, the matrix Q can be huge and one may have the problem of solving the linear system of equations. Here we present the following iterative scheme for solving the vector of prestige $\mathbf{p}$:

$$\mathbf{p}_{n+1} = \alpha Q \mathbf{p}_n + (1-\alpha)\mathbf{b}$$

where $\mathbf{b} = (I_1, I_2, \ldots, I_N)^T$. For any initial guess $\mathbf{p}_0$, we have

$$\mathbf{p}_{n+1} = (\alpha Q)^{n+1}\mathbf{p}_0 + (1-\alpha)\sum_{i=0}^{n}(\alpha Q)^i \mathbf{b}.$$

Since the spectral radius of αQ is $\alpha < 1$, we have

$$\lim_{n\to\infty}(\alpha Q)^n = 0$$

and

$$\lim_{n\to\infty}\mathbf{p}_n = (1-\alpha)\sum_{i=0}^{\infty}(\alpha Q)^i\mathbf{b} = (1-\alpha)(I-\alpha Q)^{-1}\mathbf{b} = \mathbf{p}$$

where

$$\mathbf{p} = \alpha Q\mathbf{p} + (1-\alpha)\mathbf{b},$$

see for instance [10].

References

[1] Avrachenkov, L. and Litvak, N., Decomposition of the Google PageRank and Optimal Linking Strategy, Research Report, INRIA, Sophia Antipolis, 2004.

[2] Baldi, P., Frasconi, P. and Smyth, P., Modeling the Internet and the Web, Wiley, England, 2003.

[3] Bell David C., Atkinson John S. and Carlson Jerry W., Centrality Measures for Disease Transmission Networks, Social Networks, 21 (1999) 1-21.

[4] Bonacich, P. and Lloyd, P., Eigenvector-like Measures of Centrality for Asymmetric Relations, Social Networks, 23 (2001) 191-201.

[5] Bonacich, P. and Lloyd, P., Calculating Status with Negative Relations, Social Networks, 26 (2004) 331-338.

[6] Ching, W., Ng, M. and Yuen, W., A Hybrid Algorithm for Solving the PageRank, to appear in Proceedings of HPCA2004, LNCS, Springer, 2005.

[7] Douglas, R. and Borgatti, P., Betweenness Centrality Measures for Directed Graphs Social Networks, 16 (1994) 335-346.

[8] Garfield, E., Citation Indexes for Science: A New Dimension in Documentation Through Association of Ideas, Scicence , 122 (1955)108-111.

[9] Garfield, E., Citation Analysis as a Tool in Journal Evaluation, Scicence, 178 (1972) 471-479.

[10] Golub, G. and van Loan, C., Matrix Computations, The John Hopkins University Press, Baltimore, 1993.

[11] http://www.search-engine-marketing-sem.com/Google/GooglePageRank.html.

[12] Haveliwala, T. and Kamvar, S., The Second Eigenvalue of the Google Matrix, Stanford University, Technical Report, 2003.

[13] Page, L., Brin, S., Motwani, R. and Winograd T., The PageRank Citation Ranking: Bring Order to the Web, Technical Report, Stanford University, 1998.

[14] Ross, S., Introduction to Probability Models, (7th Edition) San Diego, Calif.: Academic Press, 1997.

[15] Wasserman, S. and Faust, K.: Social Network Analysis: Methods and Applications, Cambridge Univeristy Press, Cambridge, 1994.

[16] Yuen, W., Ching, W. and Ng, M., A Hybrid Algorithm for Queueing Systems, CALCOLO, 41 (2004) 139-151.