APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
*Appl. Stochastic Models Bus. Ind.*, 2006; **22**:431–444
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.656

# Damping factor in Google page ranking

Hwai-Hui Fu*,†, Dennis K. J. Lin and Hsien-Tang Tsai

*Department of Business Administration, Shu-Te University, 59 Hun Shan Road, Yen Chau, Kaohsiung 82445, Taiwan*

## SUMMARY

Google, the largest search engine worldwide, adopts PageRank technology to determine the rank of website listings. This paper describes how damping factor is a critical factor in changing a website's ranking in traditional Google PageRank technology. A modified algorithm based on input–output ratio concept is proposed to substitute for the damping factor. Besides there is no need to choose an optimal damping factor value, the modified algorithm has an equivalent effect on computation as the traditional Google's PageRank algorithm. Copyright © 2006 John Wiley & Sons, Ltd.

Received 30 June 2005; Revised 28 April 2006; Accepted 29 June 2006

KEY WORDS: input–output ratio; search engine

## 1. INTRODUCTION

Search engines are commonly used to search for information on the Internet. High-ranking websites generate the most traffic, thereby creating the largest commercial opportunities [1–4]. Notess [5] and Ward [6] found that Google is unique in its focus on developing the perfect search engine that understands accurately what users mean and gives them back precisely the desired information. Google combines 'PageRank technology' with complex 'text-matching techniques' to ensure its search quality [5, 6].

PageRank Technology is adopted to examine the entire link structure of the Internet and locate the most important pages [7]. PageRank Technology objectively measures the importance of web pages by solving an equation of more than 500 million variables and 2 billion terms. Google's technology uses the collective intelligence of the Internet to determine the importance of a page. Neither human involvement nor manipulation of results occurs, explaining why users have confidence in Google as a source of objective information untainted by paid placement.

Recently, however, the behaviour of PageRank with respect to changes in *d,* the so-called *damping factor*, was found to be useful in link-spam detection [8]. PageRank is an important

*Correspondence to: Hwai-Hui Fu, Department of Business Administration, Shu-Te University, 59 Hun Shan Road, Yen Chau, Kaohsiung 82445, Taiwan.
† E-mail: qwing@ms39.hinet.net

ranking technique used in search engines. As a simple, robust and reliable means of evaluating the importance of web pages [1, 9, 10], PageRank is also computationally advantageous with respect to other ranking methods in that it is query independent and content independent. The selection of $d$ is empirical, and in most cases, the suggestion of $d = 0.85$ by Brin and Page [8] is used. Some studies have indicated that for real-world graphs, values of $d$ close to 1 do not provide a meaningful ranking [11, 12]. Furthermore, the value chosen for $d = 0.85$ has not been justified analytically. To our knowledge, no attempt has been made to prove that $d = 0.85$ is indeed the optimal solution.

The rest of this paper is organized as follows. Section 2 briefly introduces the evolution of Google Page Rank algorithm. Section 3 then presents the atypical effect of the damping factor $d$ on PageRank. Section 4 describes Google's enhanced PageRank formula. Concluding remarks are finally drawn in Section 5.

## 2. GOOGLE'S PAGE RANK ALGORITHM

PageRank is Google's unique quantitative algorithm for determining the significances of a page. PageRank is not the only factor that Google adopts to rank pages, but it is a major one. Page and Brin have published two different versions of Google's PageRank algorithm in several papers [1, 8, 13]. In what follows, Section 2.1 shows the Original Google PageRank algorithm and Section 2.2 illustrates the second version of Google's PageRank algorithm.

### 2.1. *The Original Google's PageRank algorithm*

The Original PageRank algorithm, as described by Page and Brin [8], is given by

$$\mathrm{PR}_{\mathrm{ori}}(\mathrm{A}) = (1-d) + d \times \left(\frac{\mathrm{PR}_{\mathrm{ori}}(T_1)}{C(T_1)} + \frac{\mathrm{PR}_{\mathrm{ori}}(T_2)}{C(T_2)} + \cdots + \frac{\mathrm{PR}_{\mathrm{ori}}(T_n)}{C(T_n)}\right)$$

where $\mathrm{PR}_{\mathrm{ori}}(A)$ is the Original Google PageRank of page $A$; $\mathrm{PR}_{\mathrm{ori}}(T_i)$ is the Original Google PageRank of pages $T_i$ that link to page $A$; $C(T_i)$ is the number of outbound links on page $T_i$; $d$ is a damping factor which can be set between 0 and 1; $n$ is the total number of all pages that link to page $A$.

Within the Original PageRank algorithm, the PageRank of a page $T$ is constantly weighted by the number of outbound links $C(T)$ on page $T$. Restated, the more outbound links a page $T$ has, the less page $A$ benefits from a link to it on page $T$. The additional inbound link for page $A$ increases page $A$'s PageRank. Finally, the sum of the weighted PageRanks of all pages $T_i$ is multiplied by a damping factor $d$, generally set to 0.85.

### 2.2. *The Second Google's PageRank algorithm*

In the second version of the algorithm, the PageRank of page $A$ is given as

$$\mathrm{PR}_{\mathrm{sec}}(A) = \frac{(1-d)}{N} + d \times \left(\frac{\mathrm{PR}_{\mathrm{sec}}(T_1)}{C(T_1)} + \frac{\mathrm{PR}_{\mathrm{sec}}(T_2)}{C(T_2)} + \cdots + \frac{\mathrm{PR}_{\mathrm{sec}}(T_n)}{C(T_n)}\right)$$

where $\mathrm{PR}_{\mathrm{sec}}(A)$ is the second version PageRank of page $A$; $\mathrm{PR}_{\mathrm{sec}}(T_i)$ is the second version PageRank of pages $T_i$ that link to page $A$; $C(T_i)$ is the number of outbound links on page $T_i$; $d$ is a damping factor that can be set between 0 and 1; $n$ is the total number of all pages that link to page $A$; $N$ is the total number of all pages on the web.

As mentioned above, the two versions of the algorithm do not differ fundamentally from each other. The second algorithm merely adapts $(1-d)/N$ to replace $(1-d)$. As for the Random Surfer Model, the second version PageRank of a page denotes the real probability of a random surfer reaching that page after clicking on many links. The PageRanks form a probability distribution over web pages, explaining why the sum of PageRanks of all pages is 1.

## 3. IMPACT OF THE DAMPING FACTOR

This section numerically analyses PageRank when $d$ changes. The impact of the damping factor on two versions of the algorithm of PageRank is analysed by studying all typical linkage examples in Rogers [14]. The degree of impact on the damping factor can be classified into four main groups, as described in Section 3.1. Inconsistent ranking under both the Original and Second PageRank algorithms, as revealed in Sections 3.2 and 3.3, will be considered.

### *3.1. Four categories of PageRank under diverse damping factor*

PageRank is divided into four categories based on the diverse damping factor $d$. Figure 1 shows the four PageRank categories. In Figure 1(a), all of the pages (page $H$, $A$, $P$ and $M$) have the same number of incoming links, all pages are of equal importance to each other, and all pages obtain the same PR of 1.0, indicating that all pages have equal PageRank value irrespective of the value of $d$. The hyperlink type of pages $H$, $A$, $P$ and $M$ in Figure 1(b) is different from that in Figure 1(a). Figure 1(b) indicates that gaps between Pages' PageRank increase for higher damping factor $d$. Figure 1(c) has six pages, named page $H$, $A$, $P$, $M$, SA and SB, in this network. The gaps between pages' PageRank are largest for $d$ in the range [0, 1]. In Figure 1(d), lines $H$ and $A$ cross line $P$, demonstrating that hyperlink ranking is not consistent with different values of $d$.

### *3.2. Inconsistent ranking under Original PageRank algorithm*

The hyperlink diagram in Figure 2 shows inconsistent ranking under the Original PageRank algorithm. The hyperlink example comprises six pages, namely Site $A$, Site $B$, Home, About, Product and More. The hyperlink includes six equations, respectively.

$$\mathrm{PR}_{\mathrm{ori}}(\mathrm{SA}) = (1-d) \tag{1}$$

$$\mathrm{PR}_{\mathrm{ori}}(\mathrm{SB}) = (1-d) + d\left(\frac{\mathrm{PR}_{\mathrm{ori}}(P)}{2}\right) \tag{2}$$

$$\mathrm{PR}_{\mathrm{ori}}(H) = (1-d) + d\left(\frac{\mathrm{PR}_{\mathrm{ori}}(M)}{1} + \frac{\mathrm{PR}_{\mathrm{ori}}(\mathrm{SA})}{1}\right) \tag{3}$$

$$\mathrm{PR}_{\mathrm{ori}}(A) = (1-d) + d\left(\frac{\mathrm{PR}_{\mathrm{ori}}(H)}{1}\right) \tag{4}$$

DOI: 10.1002/asmb

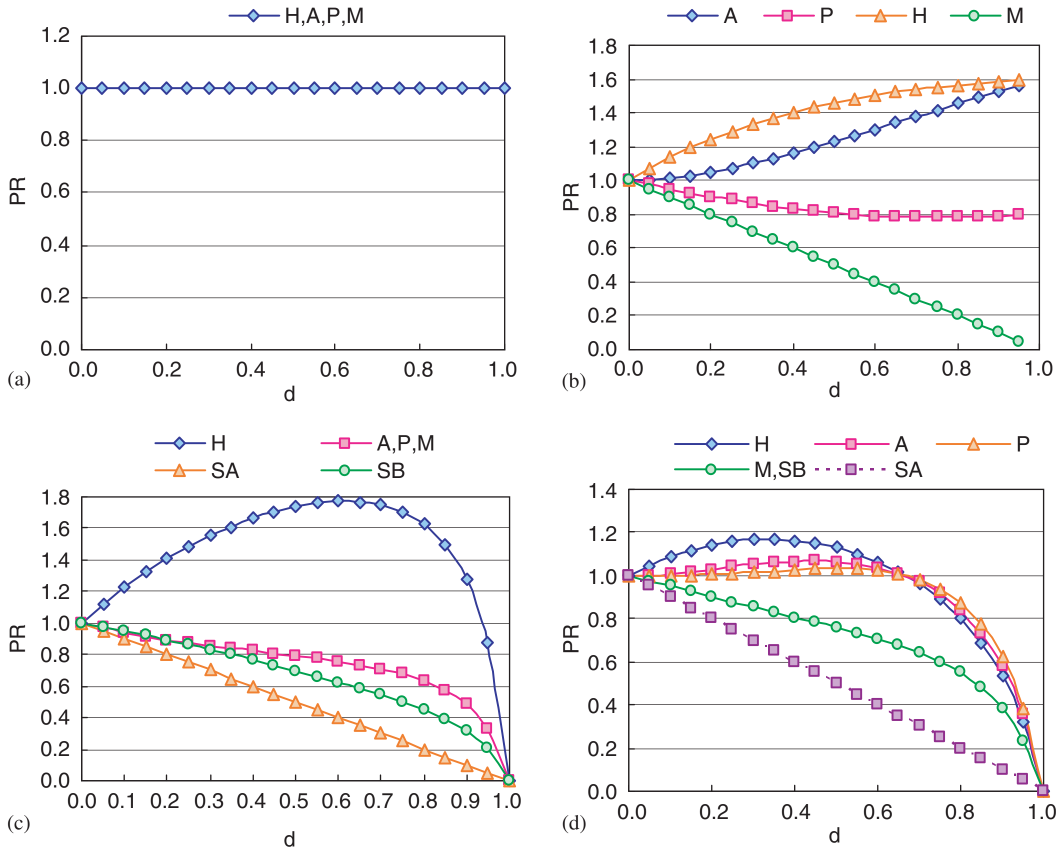

Figure 1. The four PageRank categories under diverse damping factors.

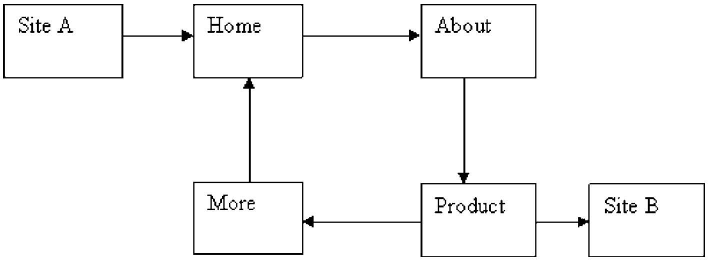

Figure 2. The hyperlink of six pages.

$$\mathrm{PR}_{\mathrm{ori}}(P) = (1-d) + d\left(\frac{\mathrm{PR}_{\mathrm{ori}}(A)}{1}\right) \tag{5}$$

$$\mathrm{PR}_{\mathrm{ori}}(M) = (1-d) + d\left(\frac{\mathrm{PR}_{\mathrm{ori}}(P)}{2}\right) \tag{6}$$

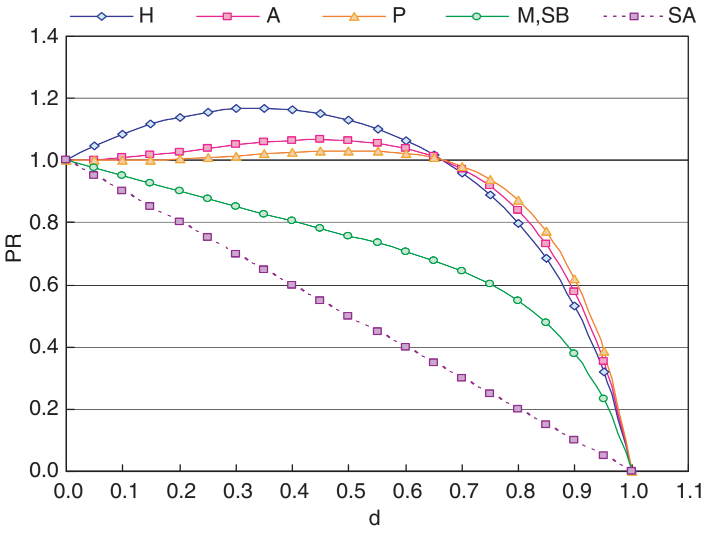

Figure 3. Inconsistent ranking under Original Page Rank algorithm.

Each page's PageRank can be calculated by solving simultaneous Equations (1)–(6). Exactly, how PageRank and damping factor $d$ are related is addressed next. Figure 3 shows how PageRank and damping factor $d$ are related, indicating that lines $H$ and $A$ intersect line $P$. In other words, this hyperlink diagram does not have a consistent ranking under diverse damping factor $d$. The damping factor is evidently the key element to change the page ranking.

### 3.3. Inconsistent ranking under the Second PageRank algorithm

The same hyperlink diagram is adopted to determine the pages' PageRank under second version's PageRank algorithm. The hyperlink includes six equations, respectively.

$$\mathrm{PR}_{\mathrm{sec}}(\mathrm{SA}) = \frac{(1-d)}{N} \tag{7}$$

$$\mathrm{PR}_{\mathrm{sec}}(\mathrm{SB}) = \frac{(1-d)}{N} + d\left(\frac{\mathrm{PR}_{\mathrm{sec}}(P)}{2}\right) \tag{8}$$

$$\mathrm{PR}_{\mathrm{sec}}(H) = \frac{(1-d)}{N} + d\left(\frac{\mathrm{PR}_{\mathrm{sec}}(M)}{1} + \frac{\mathrm{PR}_{\mathrm{sec}}(\mathrm{SA})}{1}\right) \tag{9}$$

$$\mathrm{PR}_{\mathrm{sec}}(A) = \frac{(1-d)}{N} + d\left(\frac{\mathrm{PR}_{\mathrm{sec}}(H)}{1}\right) \tag{10}$$

$$\mathrm{PR}_{\mathrm{sec}}(P) = \frac{(1-d)}{N} + d\left(\frac{\mathrm{PR}_{\mathrm{sec}}(A)}{1}\right) \tag{11}$$

$$\mathrm{PR}_{\mathrm{sec}}(M) = \frac{(1-d)}{N} + d\left(\frac{\mathrm{PR}_{\mathrm{sec}}(P)}{2}\right) \tag{12}$$

DOI: 10.1002/asmb

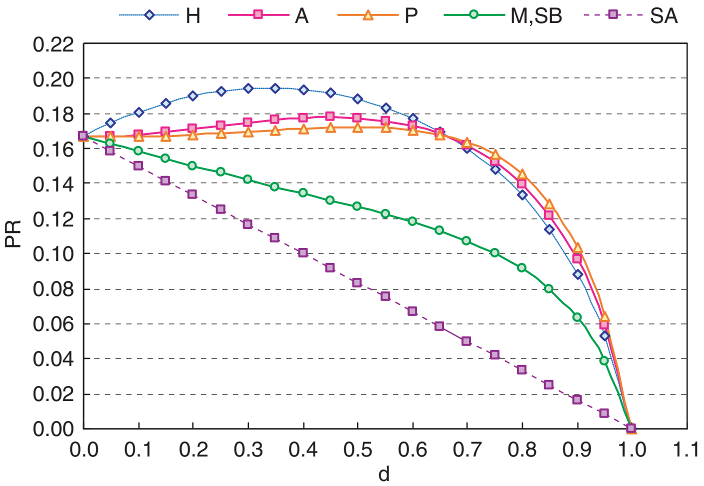

Figure 4. Inconsistent ranking under Second Page Rank algorithm.

Simultaneously, Equations (7)–(12) can be solved to obtain each page's PageRank value. Figure 4 displays the relationship between the PageRank and the damping factor *d*. The diagram reveals that lines *H* and *A* cross line *P*, indicating that this hyperlink diagram also lacks consistent ranking under different values of *d*.

This study has so far demonstrated that *d* is a significant factor in deriving the page ranking in this hyperlink diagram by determining which category of Figure 1 is presented. Values of $d=0.85$ and 0.15 can be adopted to calculate the PageRank of all pages. If both $d=0.85$ and 0.15 have the same webpage rank, then adopting $d=0.85$ is a robust and reliable way to determine the importance of web pages. Otherwise, hyperlinks diagram have inconsistent rankings under different *d* values. The damping factor *d* is a manipulative element to change the page ranking in hyperlink diagram. To solve this problem, this study proposes improved PageRank algorithms, called the New1 and New2 PageRank algorithms.

## 4. MODIFIED GOOGLE PAGERANK ALGORITHM

This section introduces the improved PageRank algorithms. Section 4.1 illustrates the PageRank algorithms. Section 4.2 contrasts the Original, Second and New1 algorithms. Section 4.3 provides illustrative examples.

### *4.1. Formulas of modified Google's PageRank algorithm*

Google's PageRank algorithm determines the importance of each page that casts a vote, since votes from some pages are considered to have a greater value than others. A link from a page with a high ranking also increases the ranking of the linked page. An important page has a higher PageRank and is placed at the apex of the search results.

DOI: 10.1002/asmb

The damping factor can, therefore, be considered as a weight. An important web page should obtain a high weight, and a less important web page should obtain a lower weight. The damping factor is defined as $n/\sum C(T_i)$. Restated, the damping factor refers to the total number of all pages that link to page $A$ divided by the sum of $C(T_i)$. The damping factor is equivalent to the input–output ratio.

This study modifies the Original PageRank algorithm as Equation (13), which is called the New1 PageRank algorithm. The Second PageRank algorithm is also transformed to Equation (14), which is called the New2 PageRank algorithm

$$\mathrm{PR}_{\mathrm{New1}}(A) = \left(1 - \frac{n}{\sum C(T_i)}\right) + \frac{n}{\sum C(T_i)} \times \left(\frac{\mathrm{PR}_{\mathrm{New1}}(T_1)}{C(T_1)} + \frac{\mathrm{PR}_{\mathrm{New1}}(T_2)}{C(T_2)} + \cdots + \frac{\mathrm{PR}_{\mathrm{New1}}(T_n)}{C(T_n)}\right) \tag{13}$$

$$\mathrm{PR}_{\mathrm{New2}}(A) = \left(1 - \frac{n}{\sum C(T_i)}\right) \times \frac{1}{N} + \frac{n}{\sum C(T_i)} \times \left(\frac{\mathrm{PR}_{\mathrm{New2}}(T_1)}{C(T_1)} + \frac{\mathrm{PR}_{\mathrm{New2}}(T_2)}{C(T_2)} + \cdots + \frac{\mathrm{PR}_{\mathrm{New2}}(T_n)}{C(T_n)}\right) \tag{14}$$

Differences between Original, Second, New1 and New2 algorithms are important and noteworthy questions and are discussed in Section 4.2.

### 4.2. Comparisons

This section discusses how the Original, Second, New1 and New2 algorithms differ from each other, again by using the hyperlink of six pages in Figure 2 as an example. For algorithms mentioned above, the PageRank of each page can be calculated with $d=0.85$ under the Original and Second formulae. The PageRank of each page can also be calculated under the New1 and New2 formulae. Table I and Figure 5 show the analytical results. In the diagram in Figure 5, the top line denotes the results of New1; the second from top denotes the results of the Original algorithm; the third from top denotes the results of New2, and the bottom line denotes the results of the Second algorithm. The observation results indicate that the New1 and Original algorithm lines have similar formulae, and the New2 algorithm and Second algorithm lines also have similar formulae. Therefore, the New1 and Original algorithms are grouped into one

Table I. PageRanks value of Original, Second and New algorithms for Figure 2.

| Rank | Web | Original PageRank | Second PageRank | New1 PageRank | New2 PageRank |
|---|---|---|---|---|---|
| 1 | Product | 0.77 | 0.129 | 2.00 | 0.333 |
| 2 | About | 0.73 | 0.129 | 2.00 | 0.333 |
| 3 | Home | 0.68 | 0.114 | 2.00 | 0.333 |
| 4 | More | 0.48 | 0.080 | 1.00 | 0.167 |
| 4 | Site *B* | 0.48 | 0.080 | 1.00 | 0.167 |
| 5 | Site *A* | 0.15 | 0.025 | 1.00 | 0.167 |

DOI: 10.1002/asmb

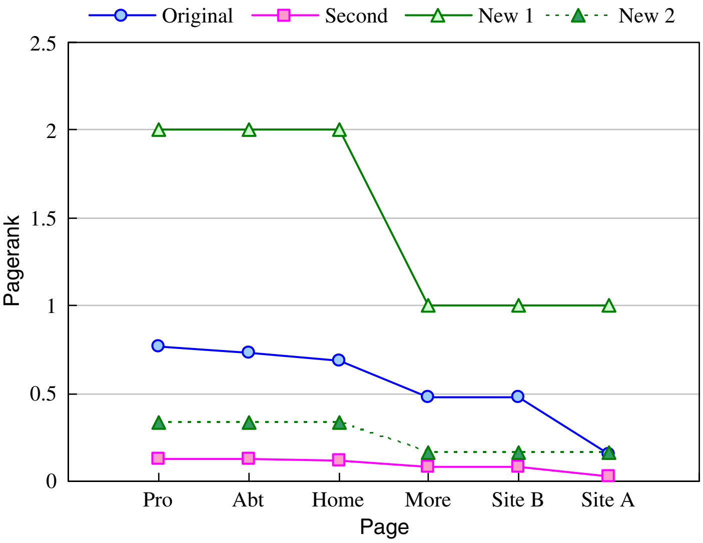

Figure 5. Comparison between Original, Second, New1 and New2 algorithms.

Table II. The comparison between Second and New2 algorithms.

| Algorithm | *d* | *H* | *A* | *P* | *M* | SB | SA |
|---|---|---|---|---|---|---|---|
| New2 | | 0.333 | 0.333 | 0.333 | 0.167 | 0.167 | 0.167 |
| Second | 0.00 | 0.167 | 0.167 | 0.167 | 0.167 | 0.167 | 0.167 |
| | 0.05 | 0.174 | 0.167 | 0.167 | 0.163 | 0.163 | 0.158 |
| | 0.10 | 0.181 | 0.168 | 0.167 | 0.158 | 0.158 | 0.150 |
| | 0.15 | 0.186 | 0.170 | 0.167 | 0.154 | 0.154 | 0.142 |
| | 0.20 | 0.190 | 0.171 | 0.168 | 0.150 | 0.150 | 0.133 |
| | 0.25 | 0.193 | 0.173 | 0.168 | 0.146 | 0.146 | 0.125 |
| | 0.30 | 0.194 | 0.175 | 0.169 | 0.142 | 0.142 | 0.117 |
| | 0.35 | 0.195 | 0.176 | 0.170 | 0.138 | 0.138 | 0.108 |
| | 0.40 | 0.194 | 0.177 | 0.171 | 0.134 | 0.134 | 0.100 |
| | 0.45 | 0.192 | 0.178 | 0.172 | 0.130 | 0.130 | 0.092 |
| | 0.50 | 0.188 | 0.177 | 0.172 | 0.126 | 0.126 | 0.083 |
| | 0.55 | 0.183 | 0.176 | 0.172 | 0.122 | 0.122 | 0.075 |
| | 0.60 | 0.177 | 0.173 | 0.171 | 0.118 | 0.118 | 0.067 |
| | 0.65 | 0.170 | 0.169 | 0.168 | 0.113 | 0.113 | 0.058 |
| | 0.70 | 0.160 | 0.162 | 0.163 | 0.107 | 0.107 | 0.050 |
| | 0.75 | 0.148 | 0.153 | 0.156 | 0.100 | 0.100 | 0.042 |
| | 0.80 | 0.133 | 0.140 | 0.145 | 0.091 | 0.091 | 0.033 |
| | 0.85 | 0.114 | 0.122 | 0.129 | 0.080 | 0.080 | 0.025 |
| | 0.90 | 0.089 | 0.096 | 0.103 | 0.063 | 0.063 | 0.017 |
| | 0.95 | 0.053 | 0.059 | 0.064 | 0.039 | 0.039 | 0.008 |
| | 1.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

cluster, and the New2 and Second algorithms are grouped into another cluster. The rest of this study focuses only on the New2 algorithm.

An additional point must be clarified. The New2 formula is not a special case of the Second algorithm. The damping factor of the Second algorithm is a numerical constant (generally $d = 0.85$). By contrast, the damping factor in the New2 algorithm changes according to a page's importance level. Table II helps to illustrate this phenomenon. Clearly, the New2 algorithm markedly differs from Google's PageRank algorithm.

### 4.3. *Illustrative examples*

This section includes four interesting illustrative examples of difference in PageRank of the Second and New2 algorithms. The list is in order of hyperlink size. Figure 6 shows a hyperlink with only four pages; Figure 7 shows a hyperlink with eight pages; Figure 8 shows a hyperlink with different connections, and Figure 9 shows a hyperlink with 12 pages.

The question to consider next is whether New2 is an improved page-ranking algorithm, based on whether it has the same page rank as the Second algorithm. The PageRanks of the Second and New2 algorithms were calculated by a series of four diagrams. Tables III–VI list the

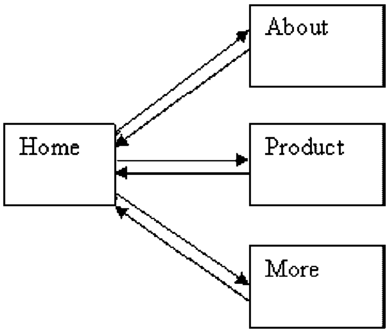

Figure 6. Hyperlink diagram with four pages.

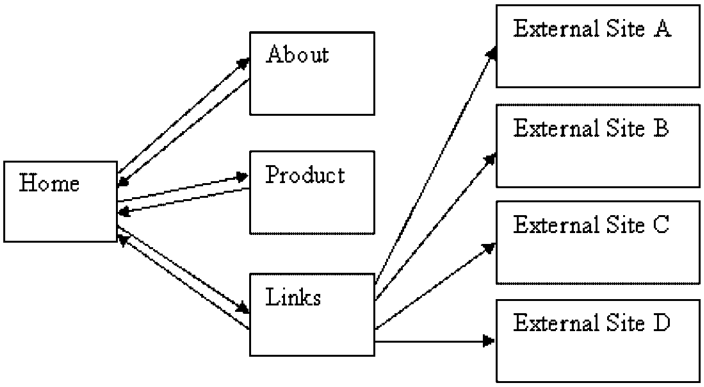

Figure 7. Hyperlink diagram with eight pages.

DOI: 10.1002/asmb

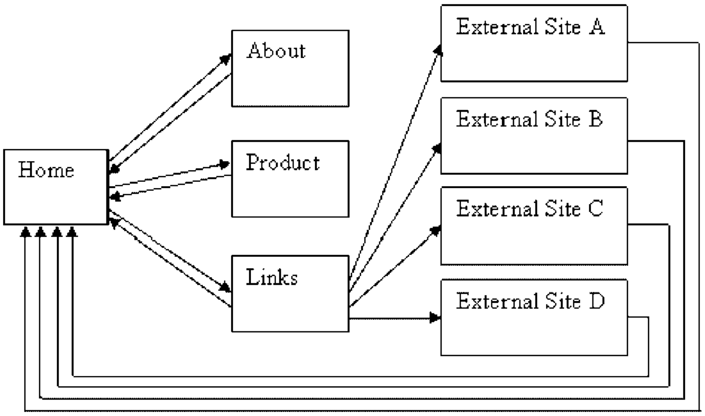

Figure 8. Complex hyperlink diagram with eight pages.

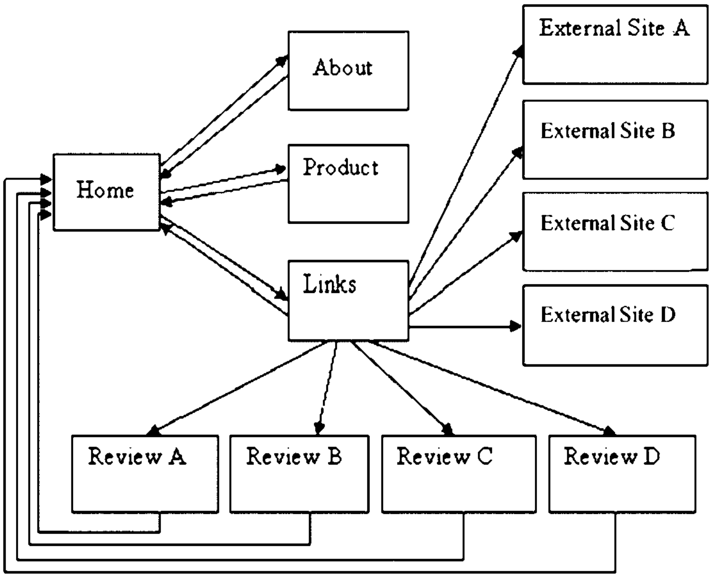

Figure 9. Hyperlink diagram with 12 pages.

Table III. Difference in PageRank of the Second and New2 algorithms for Figure 6.

| Second algorithm | | | New2 algorithm | | |
|---|---|---|---|---|---|
| Rank | Web | PageRank | Rank | Web | PageRank |
| 1 | Home | 0.48 | 1 | Home | 0.75 |
| 2 | About | 0.17 | 2 | About | 0.25 |
| 2 | Product | 0.17 | 2 | Product | 0.25 |
| 2 | More | 0.17 | 2 | More | 0.25 |

Table IV. Difference in PageRank of the Second and New2 algorithms for Figure 7.

| Second algorithm | | | New2 algorithm | | |
|---|---|---|---|---|---|
| Rank | Web | PageRank | Rank | Web | PageRank |
| 1 | Home | 0.114 | 1 | Home | 0.178 |
| 2 | About | 0.051 | 2 | Site *A* | 0.104 |
| 2 | Product | 0.051 | 2 | Site *B* | 0.104 |
| 2 | Links | 0.051 | 2 | Site *C* | 0.104 |
| 3 | Site *A* | 0.027 | 2 | Site *D* | 0.104 |
| 3 | Site *B* | 0.027 | 3 | About | 0.103 |
| 3 | Site *C* | 0.027 | 3 | Product | 0.103 |
| 3 | Site *D* | 0.027 | 3 | Links | 0.103 |

Table V. Difference in PageRank of the Second and New2 algorithms for Figure 8.

| Second algorithm | | | New2 algorithm | | |
|---|---|---|---|---|---|
| Rank | Web | PageRank | Rank | Web | PageRank |
| 1 | Home | 0.42 | 1 | Home | 0.510 |
| 2 | About | 0.138 | 2 | About | 0.140 |
| 2 | Product | 0.138 | 2 | Product | 0.140 |
| 2 | Links | 0.138 | 2 | Links | 0.140 |
| 3 | Site *A* | 0.042 | 2 | Site *A* | 0.106 |
| 3 | Site *B* | 0.042 | 3 | Site *B* | 0.106 |
| 3 | Site *C* | 0.042 | 3 | Site *C* | 0.106 |
| 3 | Site *D* | 0.042 | 3 | Site *D* | 0.106 |

Table VI. Difference in PageRank of the Second and New2 algorithms for Figure 9.

| Second algorithm | | | New2 algorithm | | |
|---|---|---|---|---|---|
| Rank | Web | PageRank | Rank | Web | PageRank |
| 1 | Home | 0.203 | 1 | Home | 0.267 |
| 2 | About | 0.070 | 2 | About | 0.085 |
| 2 | Product | 0.070 | 2 | Product | 0.085 |
| 2 | Links | 0.070 | 2 | Links | 0.085 |
| 3 | Site *A* | 0.019 | 3 | Site *A* | 0.075 |
| 3 | Site *B* | 0.019 | 3 | Site *B* | 0.075 |
| 3 | Site *C* | 0.019 | 3 | Site *C* | 0.075 |
| 3 | Site *D* | 0.019 | 3 | Site *D* | 0.075 |
| 3 | Review *A* | 0.019 | 3 | Review *A* | 0.075 |
| 3 | Review *B* | 0.019 | 3 | Review *B* | 0.075 |
| 3 | Review *C* | 0.019 | 3 | Review *C* | 0.075 |
| 3 | Review *D* | 0.019 | 3 | Review *D* | 0.075 |

analytical results. Tables III, V and VI indicate that the New2 algorithm has exactly the same ranking as the Second algorithm. The results in Tables III, V and VI are schematized in Figures 10, 12, and 13, respectively. The shape of the New2 algorithm is similar to that of the

DOI: 10.1002/asmb

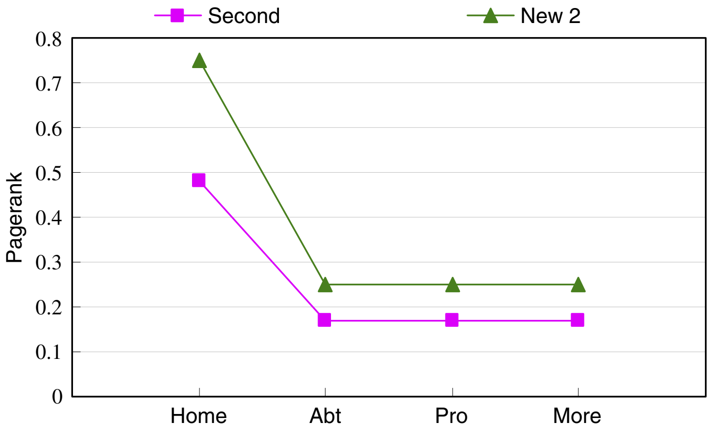

Figure 10. Difference in PageRank between Second and New2 algorithms for Figure 6.

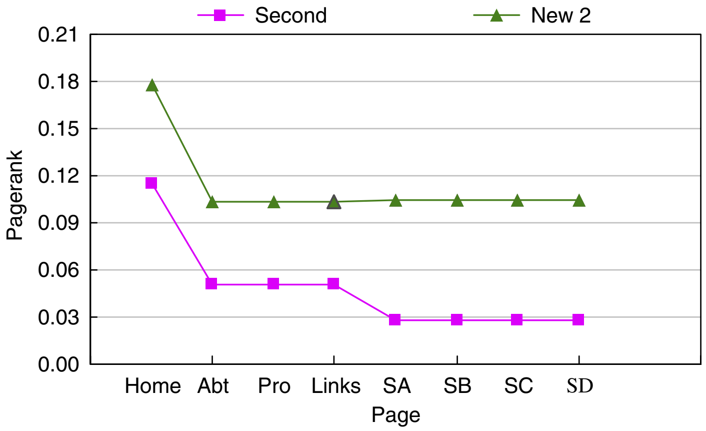

Figure 11. Difference in PageRank between Second and New2 algorithms for Figure 7.

Second algorithm in Figures 10, 12 and 13. In other words, both algorithms have similar page rankings in Figures 6, 8 and 9.

The question that must be considered next is why different ranking exists between Second algorithm and New2 algorithm in Figure 7. Table IV shows two rankings. The result in Table IV is schematized in Figure 11. Experimental results indicate that a terminal website having back links but no forward links, or having forward links without back links, has higher weight under New2 algorithm, leading to inconsistent ranking between the Second and New2 algorithms. In reality, few websites can have back links without forward links or forward links without back links. Therefore, the rankings of the New2 and Second algorithms are consistent in real hyperlinks structures. For the above reasons, New2 algorithm has two advantages: (1) it does not have to seek the optimal damping factor value and (2) it has the same effect on computation as the Second algorithm.

DOI: 10.1002/asmb

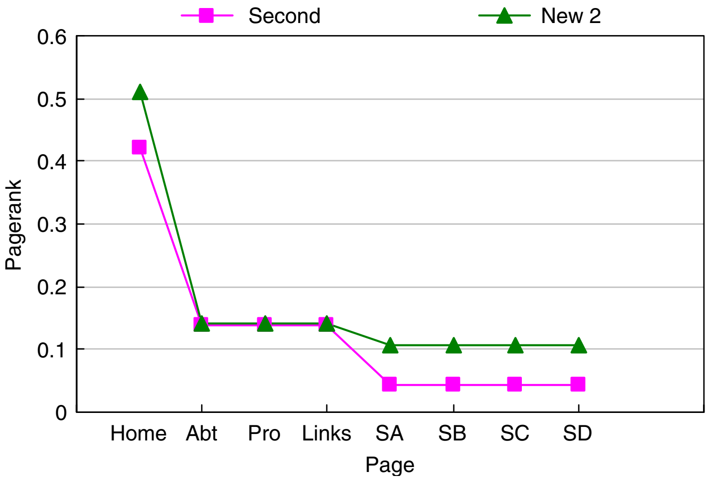

Figure 12. Difference in PageRank between Second and New2 algorithms for Figure 8.

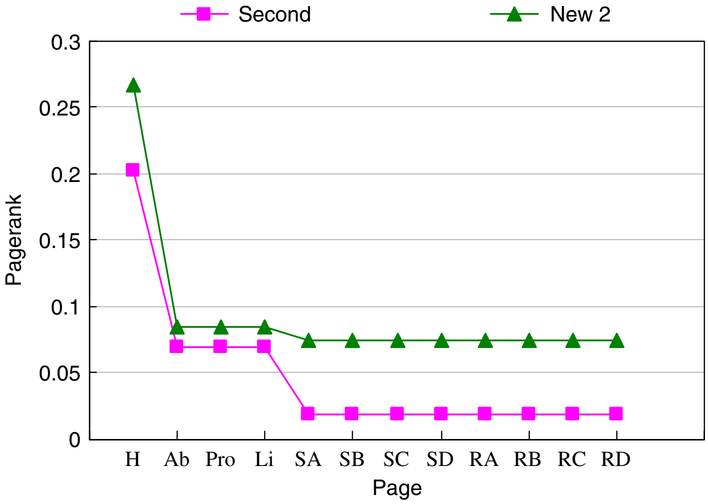

Figure 13. Difference in PageRank between Second and New2 algorithms for Figure 9.

## 5. CONCLUSIONS

This study describes how damping factor is a critical factor in changing a website's ranking in traditional Google PageRank technology. A modified algorithm based on input–output ratio concept is proposed to replace the damping factor. In the traditional Google PageRank algorithm, the damping factor is the major element to change the page ranking in hyperlink diagrams and can be set to any value in the range [0, 1]. Analysis results indicate four categories

DOI: 10.1002/asmb

of PageRank based on the damping factor *d*: (1) all websites have equal PageRank value regardless of how *d* changes; (2) the difference in PageRank increases as *d* increases; (3) the largest difference in PageRank occurs in mid-range values of *d*; and (4) websites with hyperlinks are not ranked consistently according to *d*.

In the proposed modified Google PageRank algorithm, each website hyperlink diagram has a different damping factor value depending on its importance. Besides, there is no need to choose an optimal damping factor value, the modified algorithm has an equivalent effect on computation as the traditional Google's PageRank algorithm. The proposed algorithm offers damping factor an innovative algorithm.

## REFERENCES

1. Fu HH, Lin KJ, Bai F, Tsai HT, Wei D. Sensitivity analysis on damping factor in Google PageRank. *Journal of Chinese Statistical Association* 2005; **43(2)**:15–31.
2. Mangalindan M. Seeking growth, search engine Google acts like ad agency. *Wall Street Journal* 2003; October 16.
3. Mcuhan R. Search for a top ranking. *Marketing* 2000; October 19.
4. Weidlich T. Search engine marketing revving up. *Catalog Age* 2002; **19**:3–6.
5. Notess GR. Rising relevance in search engines. *Online* 1999; **23**:84–86.
6. Ward E. Market through 'link analysis' to improve popularity quality. *Advertising Age's Business Marketing* 2000; **85**:32–33.
7. Meghabghab G. Google's web page ranking applied to different topological web graph structures. *Journal of the American Society for Information Science and Technology* 2001; **52**:736–747.
8. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 1998; **30**:107–117.
9. Lin KJ, Lin CF. Search engine and optimal list ranking: review, managerial implication and future research. *Sun Yat-Sen Management Review* 2007, forthcoming.
10. Chakrabarti S, Dom B, Gibson D, Kleinberg J, Kumar SR, Raghavan P, Rajagopalan S, Tomkins A. Hypersearching the web. *Scientific American* 1999. http://www.iprcom.com/papers/pagerank/ (5 May 2005).
11. Seneta E. *Non-Negative Matrices and Markov Chains*. Springer Series in Statistics. Springer: Berlin, 1981.
12. Haveliwala TH, Kamvar SD. The second eigenvalue of the Google matrix. http://dbpubs.stanford.edu:8090/pub/2003-20 (11 March 2003).
13. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking. Bringing order to the web. Stanford Digital Library Technologies Project 1999; November 11.
14. Rogers I. The Google PageRank algorithm and how it works. http://www.iprcom.com/papers/pagerank/ (5 May, 2006).

DOI: 10.1002/asmb