

阻尼因子對網頁排名之敏感度分析

傅懷慧¹ 林共進² 白峰杉³ 蔡憲唐⁴ 韋伯韜⁵

¹樹德科技大學企業管理系

²國立政治大學管理學院暨美國賓州州立大學供應鏈與資訊系統學系

³中國北京清華大學數學科學系

⁴國立中山大學企業管理學系

⁵淡江大學財務金融學系

摘 要

全球第一大搜尋引擎 Google 以 PageRank 技術來決定網頁排名。因此，深入瞭解 PageRank 的基本原理就更顯得有其必要性。本文利用所有基本典型範例來探討公式中阻尼因子 (*damping factor*) 變化對 PageRank 值的影響趨勢。本研究結果顯示：在不同網頁超連結方式，阻尼因子變動確實會對超連結內個別網頁 PageRank 值產生不同程度的影響，甚至於影響網頁排名，此技術即所謂索引分析 (*Citation analysis*)。進一步研究發現：馬可夫鏈 (*Markov chain*) 理論恰好能解釋網頁搜尋特性，並推導得知阻尼因子靠近 1 較好，但是又不能過分靠近 1，此與 Google 取阻尼因子值為 0.85 狀況不謀而合。

關鍵詞：搜尋引擎，索引分析，馬可夫鏈。

美國數學會分類索引：主要 15A18。

1. 前言

眾所周知, 搜尋引擎是上網查詢資訊的首要工具。網站能夠在搜尋中獲得較高的排名意味著有更多的潛在人潮, 當然也蘊藏著更多的商機 (Mangalindan, 2003; Weidlich, 2002)。由於資訊科技的日新月異以及網路的蓬勃發展, 搜尋引擎檢索及排序網站或網頁也不斷地更新演進, 以便提供給搜尋網友最新最合適的搜尋結果。

根據美國網路顧問行銷公司 WebGuerrilla 總裁 Greg Boser 說:『全球資料量最龐大, 查詢次數最多, 全球知名度最高的搜尋站是 Google。如果在 Google 上查不到, 就好比這個人或這家公司不存在一樣』。目前 Google 資料庫儲存了 80 億個網頁內容, 每天需要提供查詢服務 2 億多次, 提供 35 種搜尋結果語言, 是具有壟斷優勢的全球第一大搜尋引擎 (截止到 2005 年 3 月的資料)。倘若您從事網路營銷業務, 高速而有效地搜索資訊十分重要, 良好的 Google 排名可以迅速拓展業務至全球市場。換言之, 誰能在 Google 排名第一頁誰就是贏家。

Notess (1999) 與 Ward (2000) 兩位學者均提到 Google 爲了提供使用者想要的資訊, 使搜尋引擎功能達到完美的境界, 採用「網頁排名值 (PageRank)」與「頁面分析」兩種技術來確保檢索品質與精確率, 維持網頁搜尋的領導地位。所謂「PageRank」技術係網頁鏈結廣泛度 (Link Popularity) 之排序, 用以評估網頁的重要性。而「頁面分析」乃依頁面標題是否出現關鍵字、網頁內關鍵字出現頻率 (Keyword Density) 及關鍵字出現之位置來決定特定查詢的相關性。這部分請參照 McLuhan (2000) 與 Notess (1999) 文章。結合了所有的網頁重要性和相關性, Google 才得以將最相關和最可靠的結果放在搜尋結果頂端。一般而言, PageRank 對於排名的重要性比頁面分析高。因此, 欲針對 Google 來優化網頁, 必須瞭解 Google 的 PageRank 技術。

本文將除了介紹 Google PageRank 技術外, 將進一步藉由 Ian Rogers (2002) 中的典型範例探討 Google PageRank 公式中阻尼因子 (damping factor) 對 PageRank 值的影響, 並提出決定阻尼因子值之理論基礎。對搜尋引擎 (search engine) 歷史背景與研究有興趣者, 可參照林共進與林勤豐 (2004) 論文。

2. Google 網頁排名值 (PageRank) 之介紹

Google 所使用的 PageRank 排序技術是由 Sergey Brin 與 Lawrence Page 於 1998 年提出, 用來衡量某個網站是否被其他網站管理者所肯定 (Page, 1998)。PageRank 值的大小用來標識網頁的等級及其重要性, 級別從 1 到 10 級, 10 級為滿分。PR 值越高說明該網頁越受歡迎也越重要。一個 PR 值為 1 的網站表明這個網站不太具有流行, 而 PR 值為 7 到 10 則表明這個網站非常受歡迎。網頁 A 連至網頁 B 可解釋為網頁 B 得到網頁 A 的一張選票。網站管理者通常不會隨便去連結別人的網站, 一旦連結出去就等於投對方網站一票。

Sergey Brin 與 Lawrence Page (1998) 對 PageRank 排序觀念為: 被連結較多的網頁, 代表該網頁具有較高的重要性, 網頁排名值亦較高; 若是被重要性較高的網頁所連結, 則被連結網頁的重要性與排名值也相對較高。Page 等 (1998) 解釋其超連結 (hyperlink) 的特性結構, 可分為連結 (forward links) 與被連結 (back links) 兩種, 連結與被連結可視為引用 (citing) 與被引用 (cited) 的關係。藉由此種連結 (引用) 關係, 可透過公式計算網頁排名值 (PageRank), 來決定網頁的排列順序。依據上述 PageRank 的計算觀念, 則網頁 A 的 PageRank 值可由以下公式算出。

若 網頁 A 被 T_1, T_2, \dots, T_n 等 n 個網頁連結 (back links)

$$\text{則 } PR(A) = (1 - d) + d \times \left[\frac{PR(T_1)}{C(T_1)} + \frac{PR(T_2)}{C(T_2)} + \dots + \frac{PR(T_n)}{C(T_n)} \right]$$

其中 d 稱為阻尼因子 (damping factor), 取值為 0 到 1, Google 取值為 0.85。

$PR(T_i)$: 表示網頁 A 被連結 (back links) 網站 T_i 本身的 PR 分值, $i = 1, 2, \dots, n$ 。

$C(T_i)$: 表示網頁 A 外部鏈接網站 T_i 本身所擁有連結 (forward links) 數量, $i = 1, 2, \dots, n$ 。

因而所求得的 $PR(A)$ 值, 即為網頁 A 在所有網頁連結中代表的重要性。

PR 公式隱含一些特性如下:

1. 若網頁 A 沒有被任何網頁連結 (no back links), 即 T_i 不存在, 則得到 $PR(A) = 1 - d$ 。當 d 取為 0.85 時, 則 $PR(A) = 0.15$ 。
2. 若網頁 A 的外部鏈接數越高, 即 T_i 的個數越多 (n 越大), 則網頁 A 被越多網頁所連結, 得的投票越多, 網頁 A 越重要, $PR(A)$ 值也就越高。
3. 若網頁 A 的外部鏈接站點的級別越高, 即 $PR(T_i)$ 值越高, 則網頁 A 越重要,

$PR(A)$ 值越高。

4. 若網頁A外部鏈接網站 T_i 本身所擁有連結 (forward links) 數量越多, 即 $C(T_i)$ 值越高, 網頁 A 所能夠得到的 PR 分值反而越低, 呈反向關係。
5. 綜合特性 3 與特性 4 得知: 因 T_i 鏈接 $C(T_i)$ 個網頁, 所以 $PR(T_i)$ 值的 $\frac{1}{C(T_i)}$ 比例才是 T_i 網頁給網頁 A 的分值。即 T_i 網頁投給網頁 A 的分值為 $\frac{PR(T_i)}{C(T_i)}$ 。
6. T_i 網頁投給網頁 A 的分值為 $\frac{PR(T_i)}{C(T_i)}$, 但網頁 A 僅得到 $d * \left[\frac{PR(T_i)}{C(T_i)} \right]$ 。當 d 設 0.85 則網頁 A 僅得 T_i 網頁投給 A 網頁分值之 85%。

簡單地來說, 要提高網頁的 Google PageRank 排序基本重點: 1. 連結進來的數量愈多愈好。2. 外部連結網站本身的 PageRank 值愈高愈好。3. 外部連結網站連結出去的數量越低越好。然而, 卻少有文章去探討 d 為什麼設 0.85? d 從 0 到 1 變動, 會對 PageRank 值產生何種變化? 本文將在下一段落分析 d 變動對 PageRank 值產生何種變化? 對連結網頁排名是否造成影響?

3. 阻尼因子之敏感度分析

依據上述 PageRank 的計算公式可確定阻尼因子 d 值變動必定會影響 PageRank 值。Rogers (2002) 一文中列出 14 種基本網頁連結。通過所有 14 種基本網頁連結, 本文設定不同 d 值來觀察 PageRank 值變化情形。此 14 道網頁連結範例共可歸納出三大類型, 三大類型之 d 值變化對 PageRank 值影響趨勢圖分別如下列三範例所示。

範例一僅有兩網頁 A 與 B 相互連結, 如圖 1 所示。網頁 A 與網頁 B 之 PageRank 值可由 (1) (2) 兩式聯立求解。此時 d 值由 0 至 1 的調整, 網頁 A 與網頁 B 之 PageRank 值如圖 2 所示。由圖 2 發現網頁 A 與 B 相互連結, 互給相同 PageRank 分值, 所以 d 值由 0 至 1 的變化對網頁 A 與網頁 B 之 PageRank 值均無影響, 此時網頁 A 與 B 的 PageRank 值均固定為 1, 重疊成一直線。與此類型相同之其他範例因大同小異而被列在附錄 A。

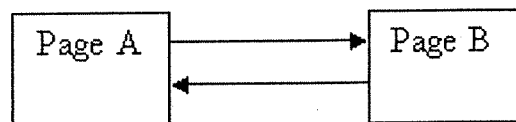


圖 1 範例一網頁連結圖

$$PR(A) = (1 - d) + d \left[\frac{PR(B)}{1} \right] \quad (1)$$

$$PR(B) = (1 - d) + d \left[\frac{PR(A)}{1} \right] \quad (2)$$

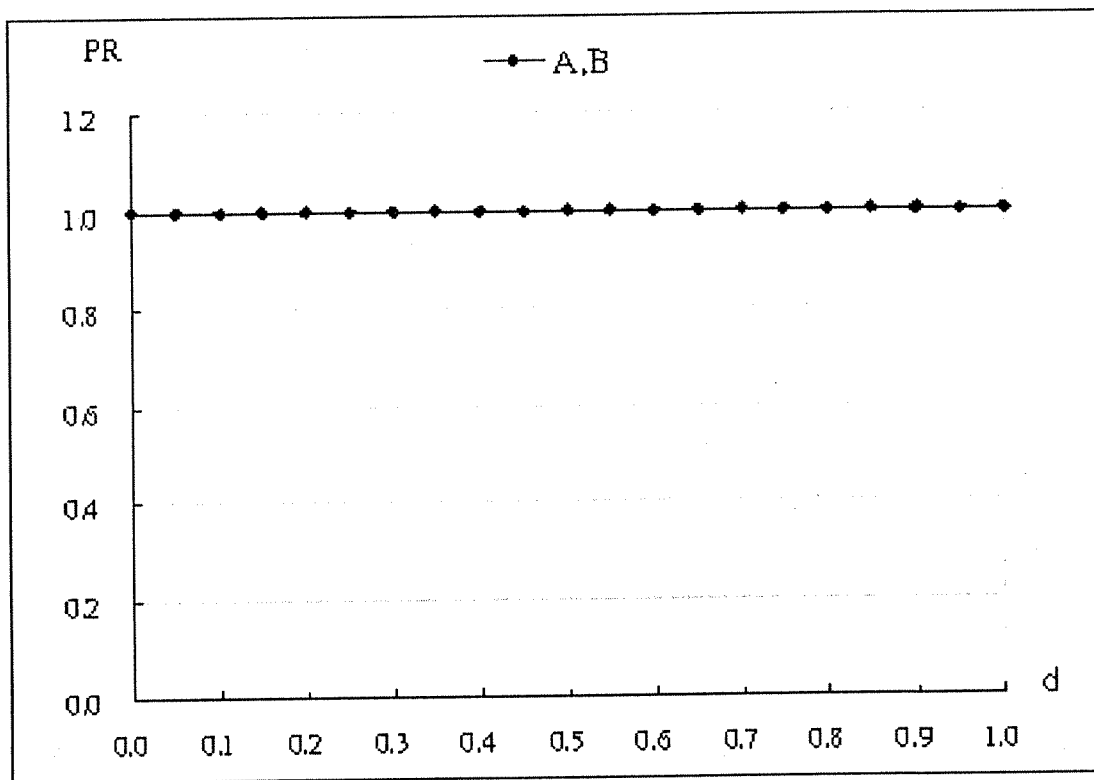


圖 2 範例一阻尼因子 d 變化對各網頁 PageRank 值影響趨勢圖

範例二有四個網頁 Page A、Page B、Page C、Page D 連結方式如圖 3 所示。四個網頁之 PageRank 值可由 (3) (4) (5) (6) 式聯立求解，此時 d 值由 0 至 1 變化，相應的 PageRank 值的變化如圖 4 所示。由圖 4 發現：不論 d 值如何由 0 至 1 變化，對網頁 C 的 PageRank 值均最高，其次是網頁 A，再其次是網頁 B，最後為網頁 D。 d 值越大，網頁 A、B、D 三者間 PageRank 值差距越大；但網頁 C 與 A 兩者間 PageRank 值差距卻越小，即 d 值取接近 1 時網頁 C 與 A 越分不出優劣。依 Google 採 d 值 = 0.85，得 $PR(A) = 1.49$ 、 $PR(B) = 0.78$ 、 $PR(C) = 1.58$ 、 $PR(D) = 0.15$ 。與此類型相同之其他兩範例因大同小異將被列在附錄 B。

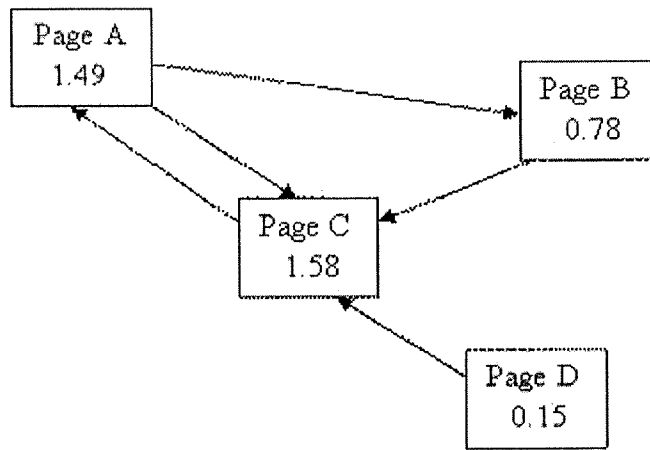


圖 3 範例二網頁連結圖

$$PR(A) = (1 - d) + d \left[\frac{PR(C)}{1} \right] \tag{3}$$

$$PR(B) = (1 - d) + d \left[\frac{PR(A)}{2} \right] \tag{4}$$

$$PR(C) = (1 - d) + d \left[\frac{PR(A)}{2} + \frac{PR(B)}{1} + \frac{PR(D)}{1} \right] \tag{5}$$

$$PR(D) = (1 - d) \tag{6}$$

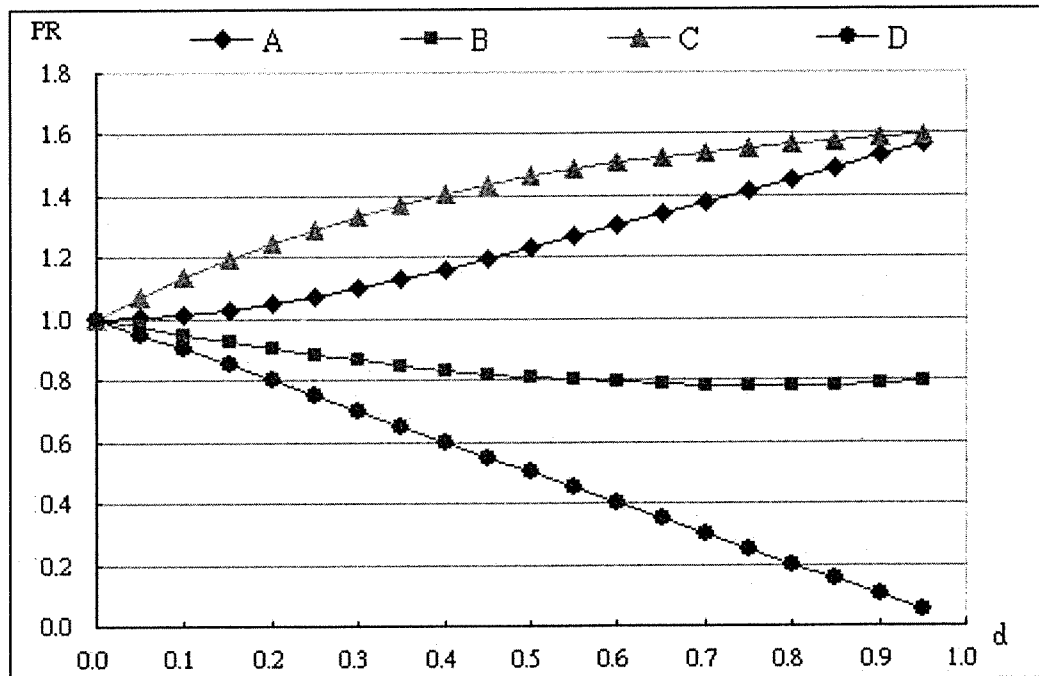


圖 4 範例二阻尼因子 d 變化對各網頁 PageRank 值影響趨勢圖

範例三有八個網頁 Home、About、Product、Links、External Site A、External Site B、External Site C、External Site D 連結方式如圖 5 所示。八個網頁之 PageRank 值可由 (7) (8) (9) (10) (11) (12) (13) (14) 式聯立求解。此時 d 值由 0 至 1 的調整, 相應的 PageRank 值如圖 6 所示。

由圖 6 發現: 網頁 PageRank 值分三群, 網頁 Home 自成一群, 網頁 About、Product、Links 三者 PageRank 值相同, External Site A、External Site B、External Site C、External Site D 四者 PageRank 值相同。 $d \in [0, 1]$ 網頁間 PageRank 值高低次序分別為網頁 Home 的 PageRank 值最高, 其次是網頁 About、Product、Links, 最後為 External Site A、External Site B、External Site C、External Site D。 d 值越接近 0 或 1, 三群網頁 PageRank 值差距越小, 即較不易分出優劣; 相反的, d 值近 0.5 時, 三群網頁 PageRank 值差距越大。依 Google 採 d 值 = 0.85, 則 $PR(H) = 0.9$ 、 $PR(A) = PR(P) = PR(L) = 0.41$ 、 $PR(EA) = PR(EB) = PR(EC) = PR(ED) = 0.22$ 。與此類型相同之其他四個範例將被列在附錄 C。

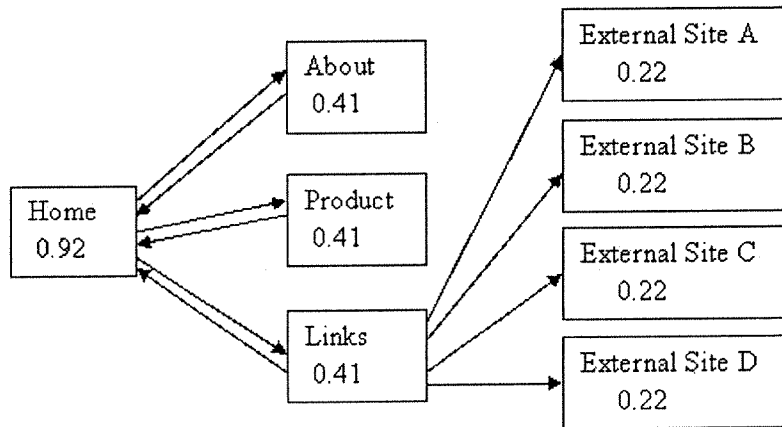


圖 5 範例三網頁連結圖

$$PR(H) = (1 - d) + d \left[\frac{PR(A)}{1} + \frac{PR(P)}{1} + \frac{PR(L)}{5} \right] \quad (7)$$

$$PR(A) = (1 - d) + d \left[\frac{PR(H)}{3} \right] \quad (8)$$

$$PR(P) = (1 - d) + d \left[\frac{PR(H)}{3} \right] \quad (9)$$

$$PR(L) = (1 - d) + d \left[\frac{PR(H)}{3} \right] \quad (10)$$

$$PR(EA) = (1 - d) + d \left[\frac{PR(L)}{5} \right] \quad (11)$$

$$PR(EB) = (1 - d) + d \left[\frac{PR(L)}{5} \right] \quad (12)$$

$$PR(EC) = (1 - d) + d \left[\frac{PR(L)}{5} \right] \quad (13)$$

$$PR(ED) = (1 - d) + d \left[\frac{PR(L)}{5} \right] \quad (14)$$

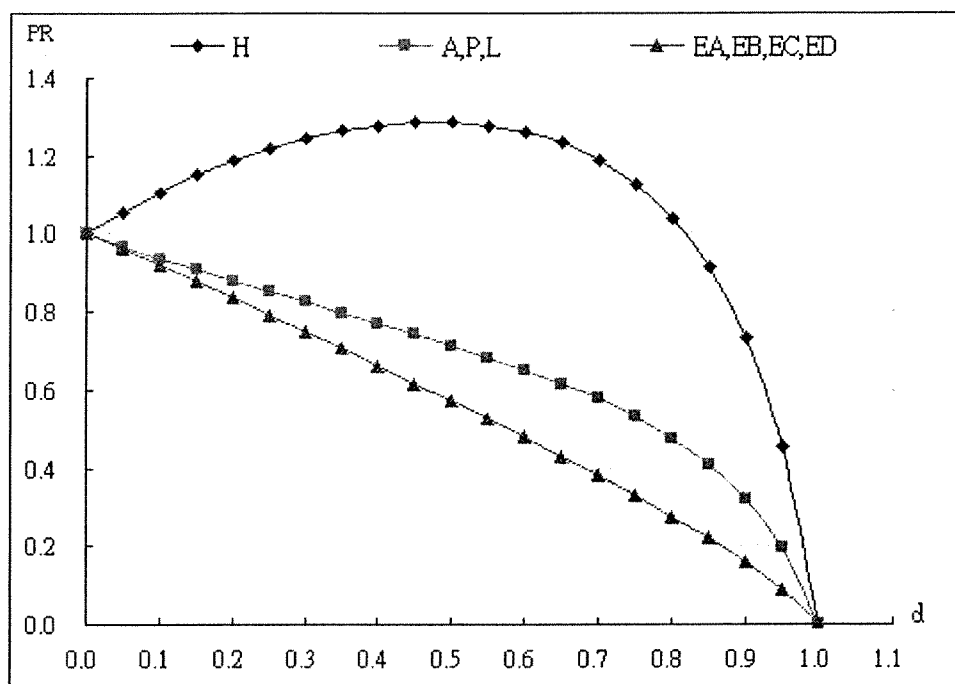


圖 6 範例三阻尼因子 d 變化對各網頁 PageRank 值影響趨勢圖

由上述三個範例可歸納出阻尼因子變動對超連結內個別網頁 PageRank 值影響的三種類型：第一種類型是不論阻尼因子如何變動，各個網頁 PageRank 值均固定為 1。第二種類型是阻尼因子值越接近 1，各個網頁 PageRank 值差距越大，越能分辨各網頁之優劣。第三種類型是阻尼因子值越接近兩端 0 或 1，各個網頁 PageRank 值差距越小，較不易分出優劣；相反的，阻尼因子值近區間中段，各個網頁 PageRank 值差距越大，越能顯現優劣。綜言之，在不同網頁超連結方式， $d \in [0, 1]$ 變化對超連結內個別網頁 PageRank 值變動敏感度不具相同規則，甚至於影響網頁排名，此技術即所謂索引分析 (citation analysis)。因此， d 值選取應慎選合宜理論來套用，方可達到區分網頁優劣之差異。下一段落將介紹如何引進 Markov 鏈，來解釋 d 在 PageRank 中扮演的角色。

4. Markov 鏈模式下之阻尼因子

由上節陳述可得知：不同網頁超連結方式， $d \in [0, 1]$ 變化對超連結內個別網頁 PageRank 值影響趨勢不相同。因此，討論阻尼因子應從網頁搜尋特性著手。假設我們在上網流覽頁面，選擇下一個頁面的過程，與過去流覽過哪些頁面並無關係，而僅依賴於當前所在的頁面。這是一個簡單的有限狀態、離散時間的隨機過程，可以用 Markov 鏈描述。

假設 Google 資料庫共有 N 個網頁，定義一個 $N \times N$ 的方陣 G ：如果從網頁 j 到網頁 i 有超鏈結，則 $g_{ij} = 1$ 否則為 0。顯然 G 是巨大的但非常稀疏的矩陣，其中非零元素的總數即是網頁超鏈結的總數。

令

$$c_j = \sum_i g_{ij}$$

$$r_i = \sum_j g_{ij}$$

分別是矩陣 G 的列和及行和，它們分別給出了頁面 j 的鏈出數目和頁面 i 的鏈入數目。再定義矩陣 A

$$a_{ij} = \frac{(1-d)}{N} + d \frac{g_{ij}}{c_j}$$

其中 d 是模型參數，通常取 $d = 0.85$ ，則 A 是 Markov 鏈的轉移概率矩陣。可以證明 A 的最大特徵值為 1，相應的特徵向量之各元素一定非負並滿足

$$Ax = x$$

如果假設

$$\sum_i x_i = 1$$

則 x 是 Markov 鏈的平穩分佈。假設 x 已經得到，則它按分量滿足

$$x_k = \sum_{j=1}^N a_{kj} x_j = \frac{(1-d)}{N} + d \sum_{j=1}^N \frac{x_j}{c_j}$$

網頁 j 的 PageRank 值是 x_j ，它鏈出的頁面有 c_j 個，於是頁面 j 將它的 PageRank 值分成 c_j 份，分別“投票”給它鏈出的網頁。 x_k 為網頁 k 的 PageRank 值，即網路上所有頁面“投票”給網頁 k 的最終值。倘若將公式中 $\frac{(1-d)}{N}$ 替換成 $(1-d)$ ，概念上仍類似於 Markov 鏈特性，因而不再做細部討論。

PageRank 是 Markov 鏈轉移概率矩陣 A 特徵值 (eigenvalue) 所對應的特徵向量 (eigenvector), 它對矩陣擾動的敏感性, 依賴於其他特徵值與 1 的分離程度。Haveliwala 與 Kamvar (2003) 提出阻尼因子 d 值靠近 1, 矩陣 A 的次特徵值會隨之靠近 1, 從而會導致 PageRank 對 d 值的選擇敏感依賴, 此時計算特徵向量方法的收斂速度降低。

如範例二類型之網頁超連結其阻尼因子 d 值越靠近 1, PageRank 差距拉大, 越能顯示出網頁優劣, 此與 Markov 鏈模式下 d 值選取吻合。倘若網頁超連結如範例三類型時, 其阻尼因子 $d = 1$ 時卻無法判別各網頁優劣, 顯而易見, 此類網頁超連結方式僅能讓 d 值靠近 1 但不能等於 1。由此看來, Google 選取 $d=0.85$ 確實具有理論基礎與考量網頁超連結方式之雙重功效。

5. 結論

PageRank 是 Google 用於評定一個網頁重要性的一種方法, 越具重要性的網頁排名越前面, 進而提高 Google 搜索的品質。網站管理者若要提升在 Google 網頁排名不外乎: 連結進來的網頁數量愈多愈好; 外部連結網站本身的 PageRank 值愈高愈好; 外部連結網站連結出去的數量越低越好。然而, PageRank 計算公式中阻尼因子亦也是影響網頁排名的重要參數, 經由本研究深入探討阻尼因子, 得到下列結論:

阻尼因子變動確實會對超連結內個別網頁 PageRank 值產生不同程度的影響, 依影響程度可分為三種不同類型: 第一種類型是網網相互連結, 此時不論阻尼因子如何變動, 均無法影響各個網頁 PageRank 值。第二種類型是阻尼因子值越接近 1, 各個網頁 PageRank 值差距越大, 越能分辨各網頁之優劣。第三種類型是阻尼因子值越接區間中段, 各個網頁 PageRank 值差距越大, 越能顯現網頁排名優劣。

經由理論驗證發現: 馬可夫鏈理論的基本假設與網頁搜尋特性相同, 經推導得知阻尼因子越靠近 1 越好。但第三種類型阻尼因子值太靠近 1, 則無法區分網頁優劣。Google 公司在不同時點可能碰到不同超連結方式, 取阻尼因子值為 0.85 必有其實務面考量。

PageRank 將本來無序的、浩如煙海的網頁賦予合理的次序, 從而幫助我們尋找最有價值的頁面。資訊時代高速膨脹的不僅僅是網路頁面而已, 所以 Rank-

ing 無疑是資料挖掘中基本而重要課題,它幫助我們從海量的資訊中尋找甚麼是最重要的。本文主要是對此重要課題起了個開頭,希望達成拋磚引玉之功效,讓有興趣讀者能共襄盛舉,對 PageRank 技術提出精闢見解,使之更趨於完善。再則,當今科學技術的發展日新月異,每年發表的研究論文數量呈指數增長,如果將研究論文的參考文獻視為搜尋引擎的連結,即類似於 PageRank 的思維,或許也可以考慮由 PaperRank 來評價科學論文的重要性,從而對學術研究排名給一個公平合理的次序。

附錄 A: 與範例一特性相同之網頁連結圖與阻尼因子之敏感趨勢圖

附錄 A-1

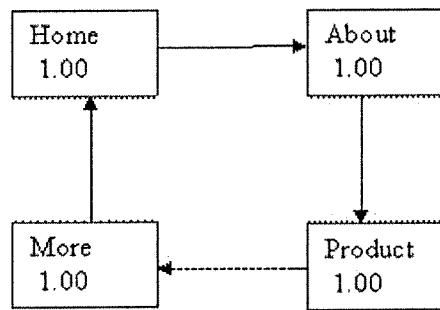


圖 7 網頁連結圖

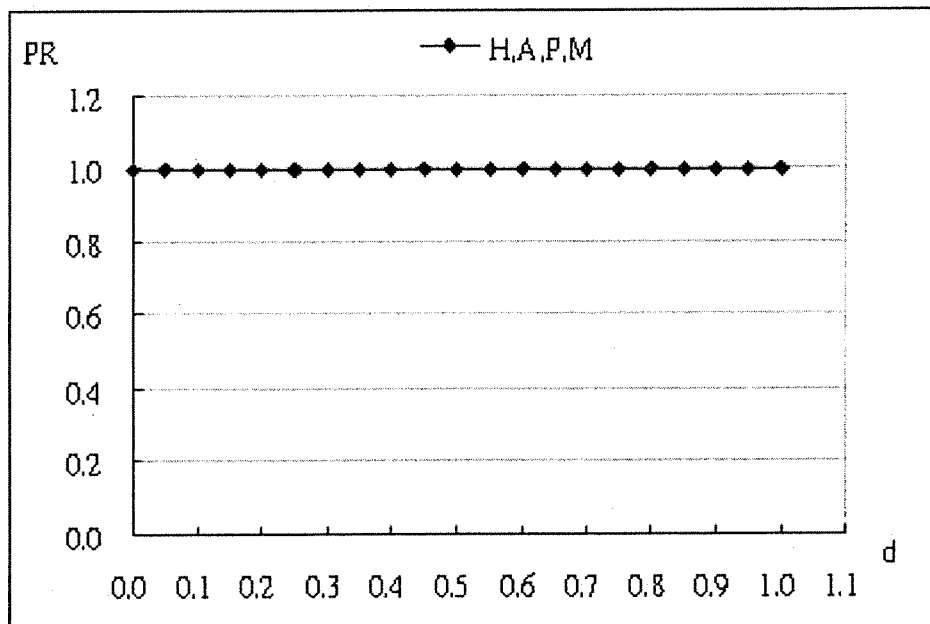


圖 8 d 值變化對各網頁 PageRank 值之影響

附錄 A-2

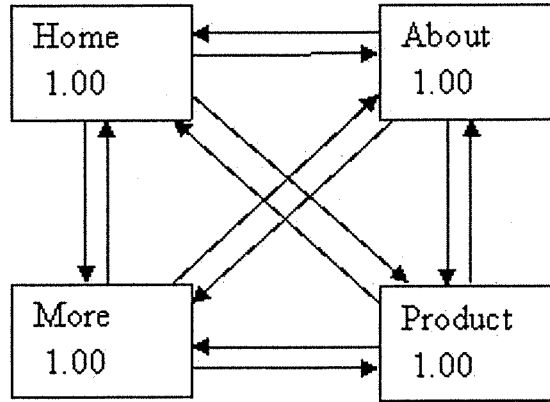


圖 9 網頁連結圖

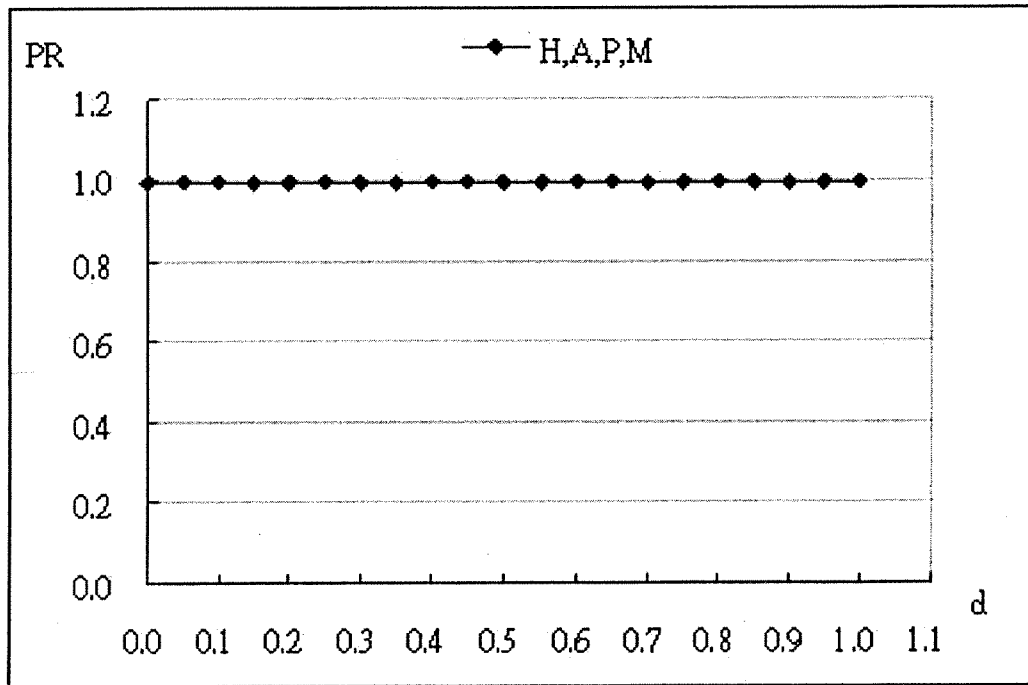


圖 10 d 值變化對各網頁 PageRank 值之影響

附錄B: 與範例二特性相同之網頁連結圖與阻尼因子之敏感趨勢圖

附錄B-1

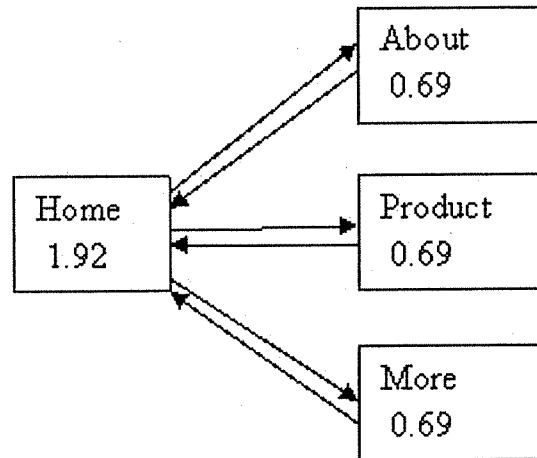


圖 11 網頁連結圖

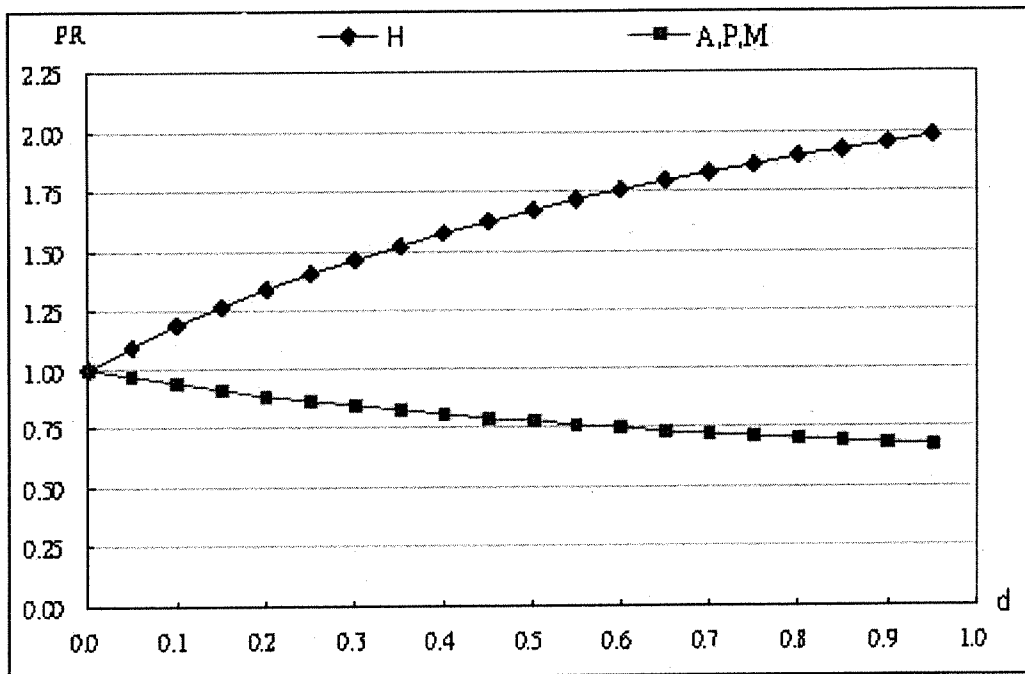


圖 12 d 值變化對各網頁 PageRank 值之影響

附錄B-2

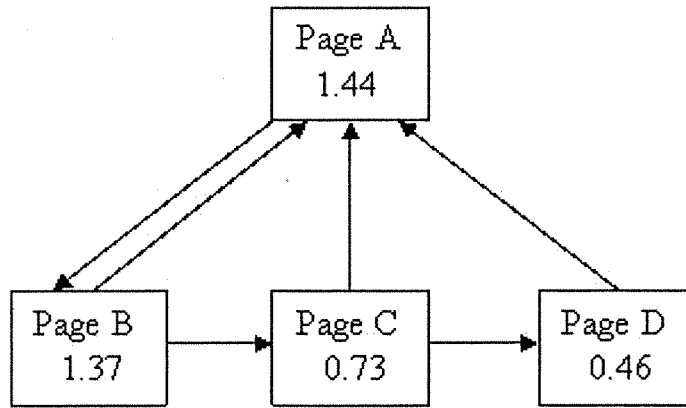


圖 13 網頁連結圖

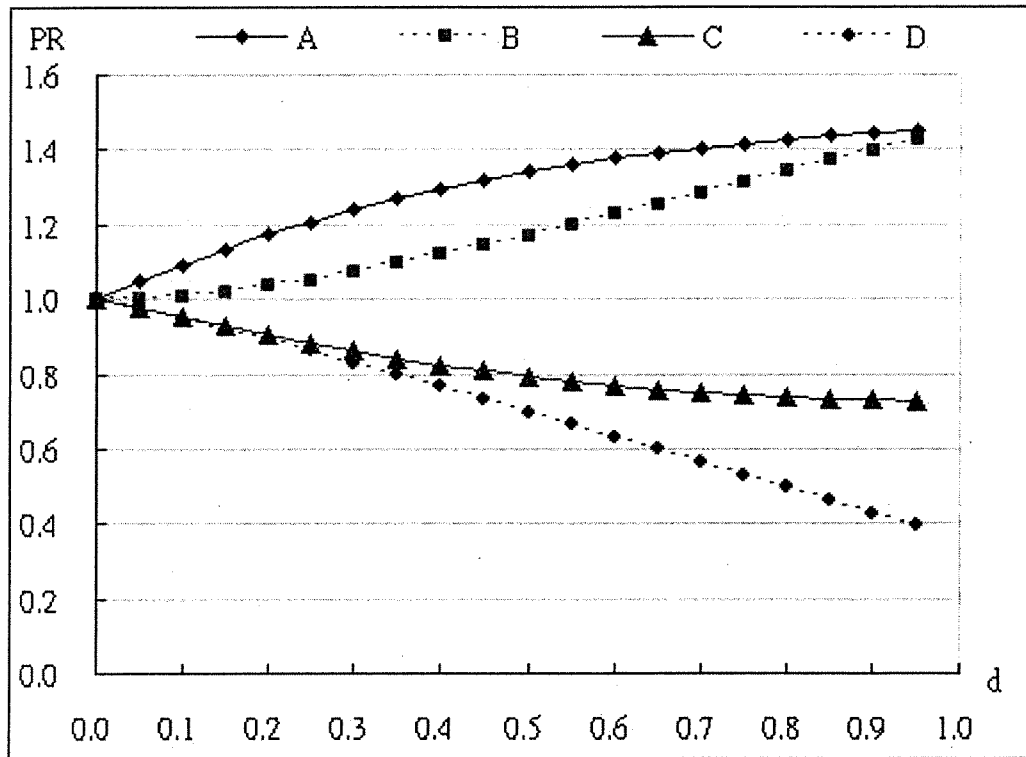


圖 14 d 值變化對各網頁 PageRank 值之影響

附錄 C: 與範例三特性相同之網頁連結圖與阻尼因子之敏感趨勢圖

附錄 C-1

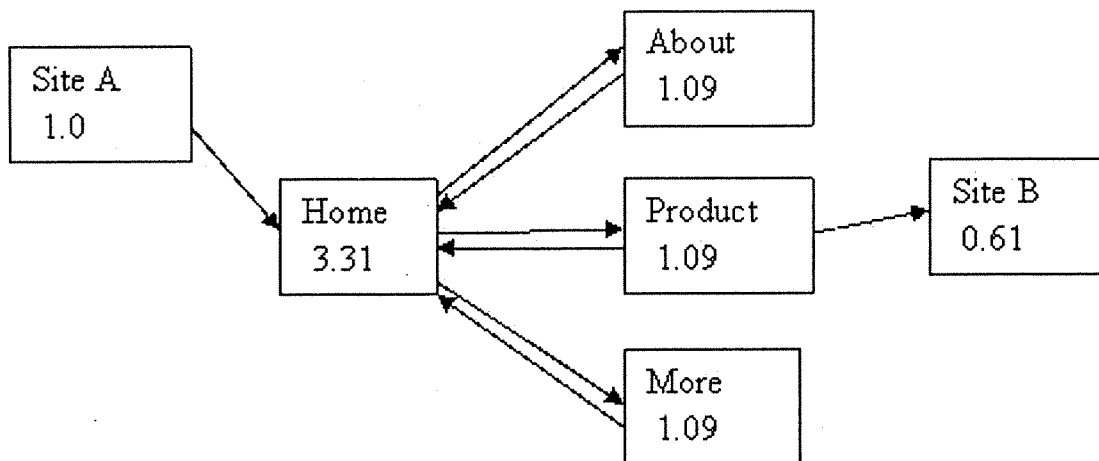


圖 15 網頁連結圖

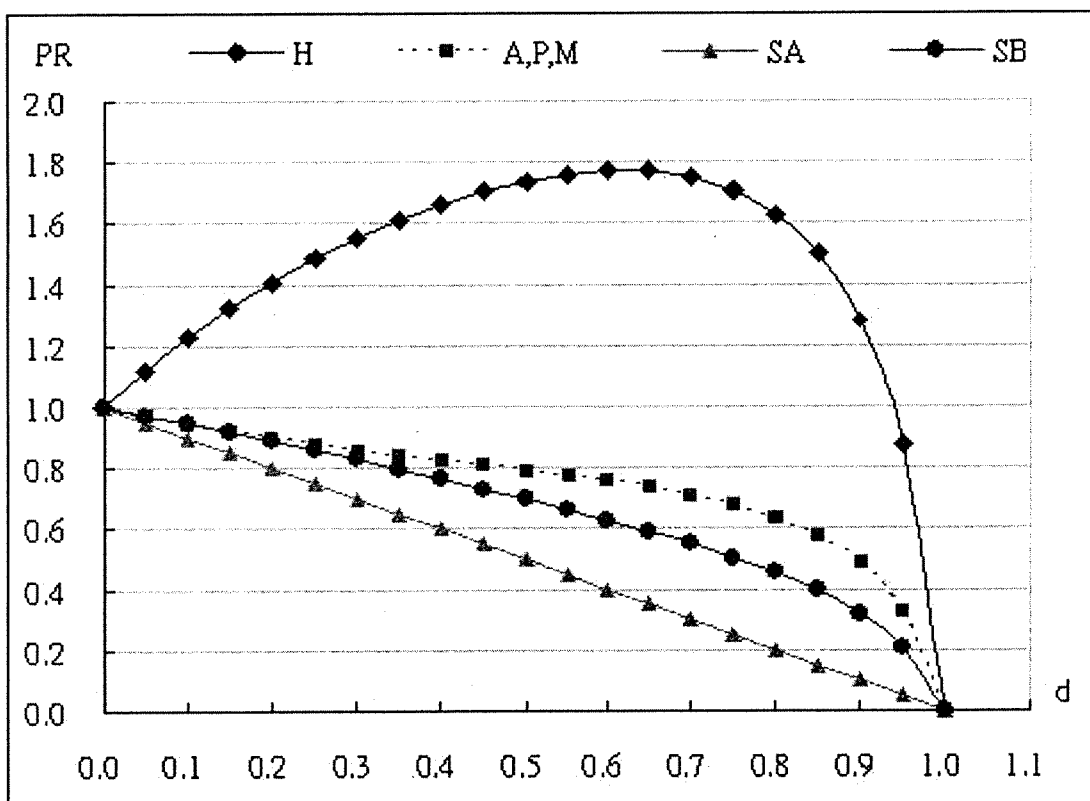


圖 16 d 值變化對各網頁 PageRank 值之影響

附錄 C-2

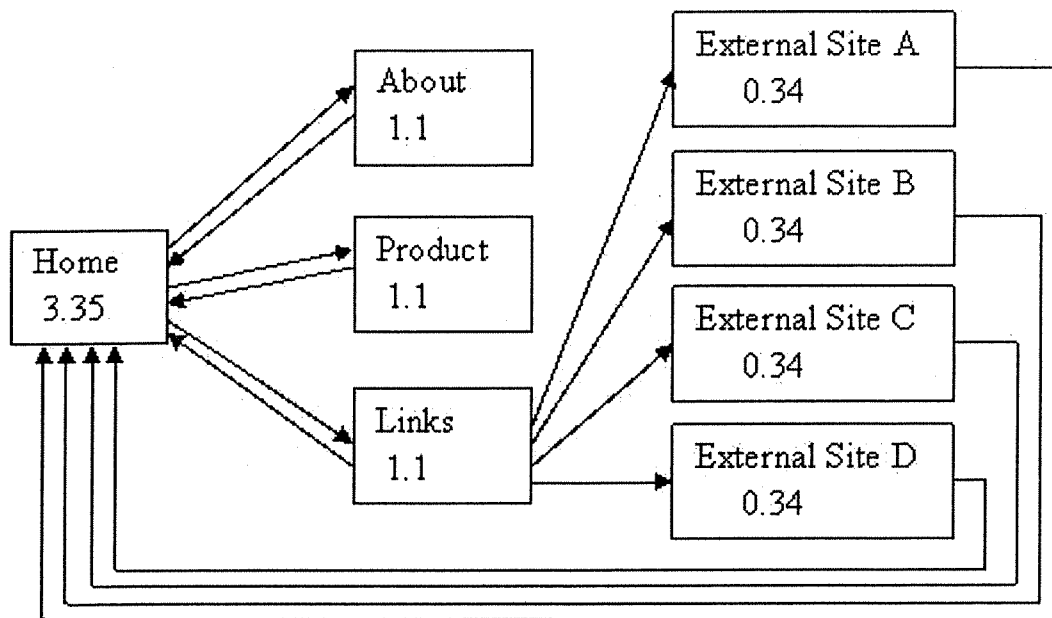


圖 17 網頁連結圖

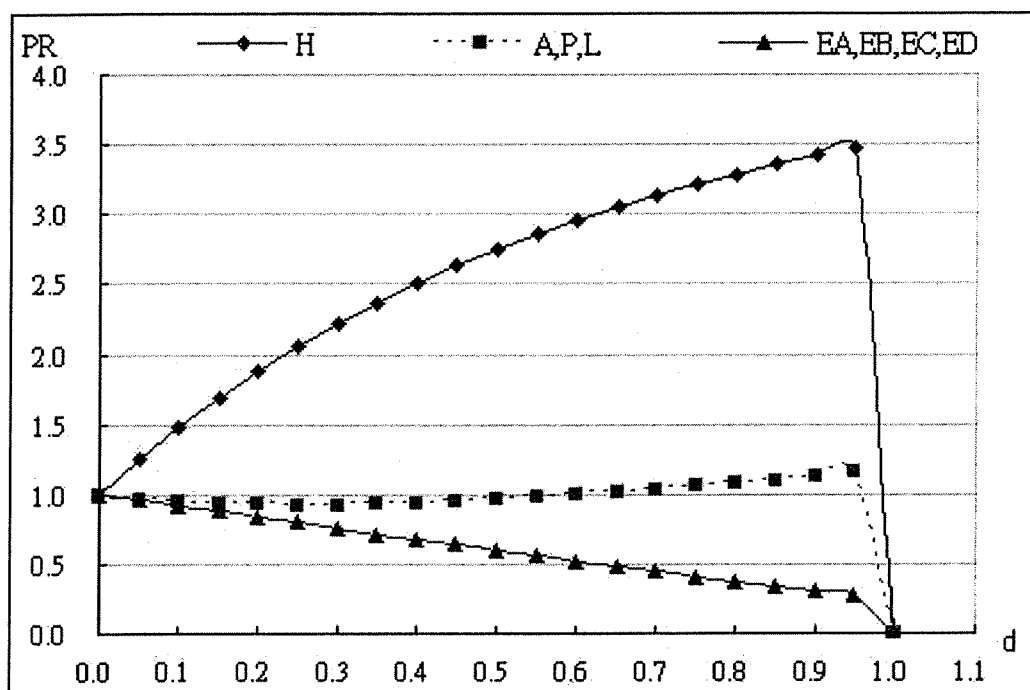


圖 18 d 值變化對各網頁 PageRank 值之影響

附錄 C-3

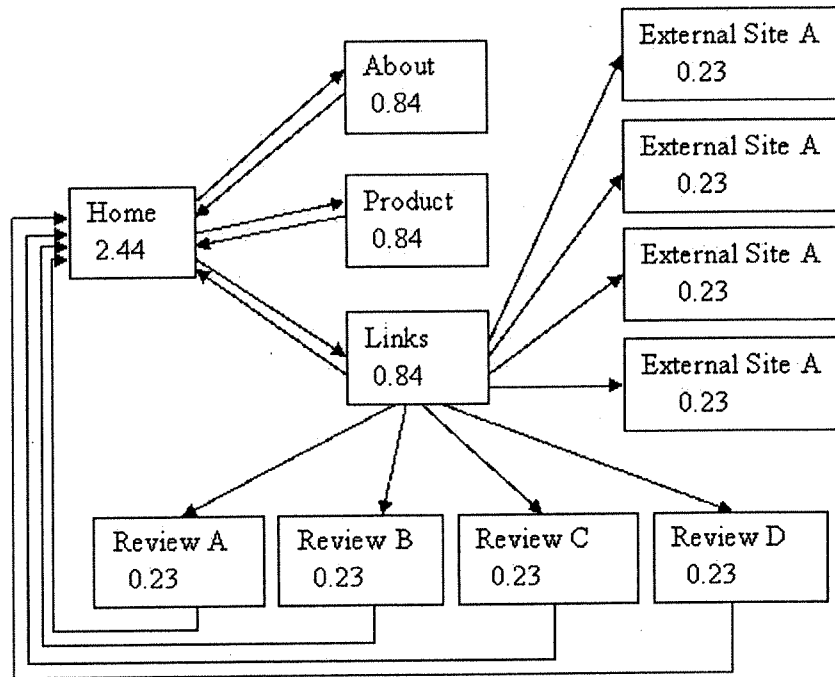


圖 19 網頁連結圖

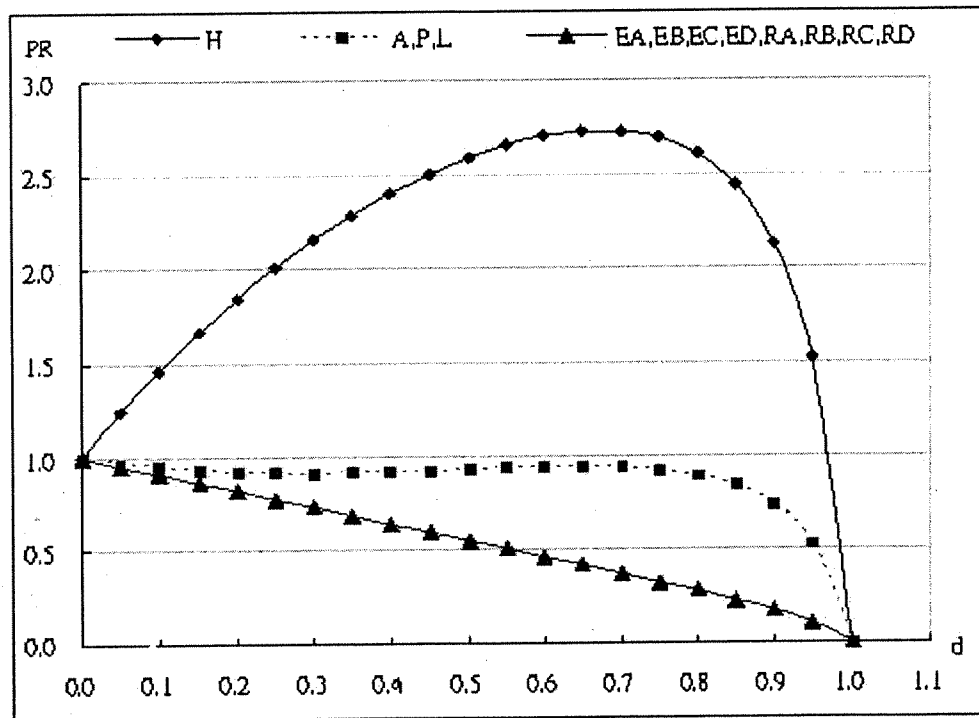


圖 20 d 值變化對各網頁 PageRank 值之影響

附錄 C-4

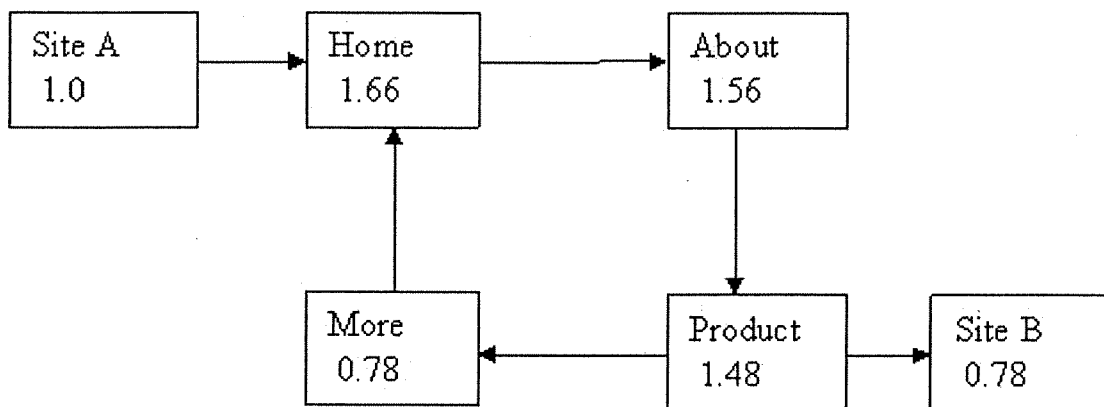


圖 21 網頁連結圖

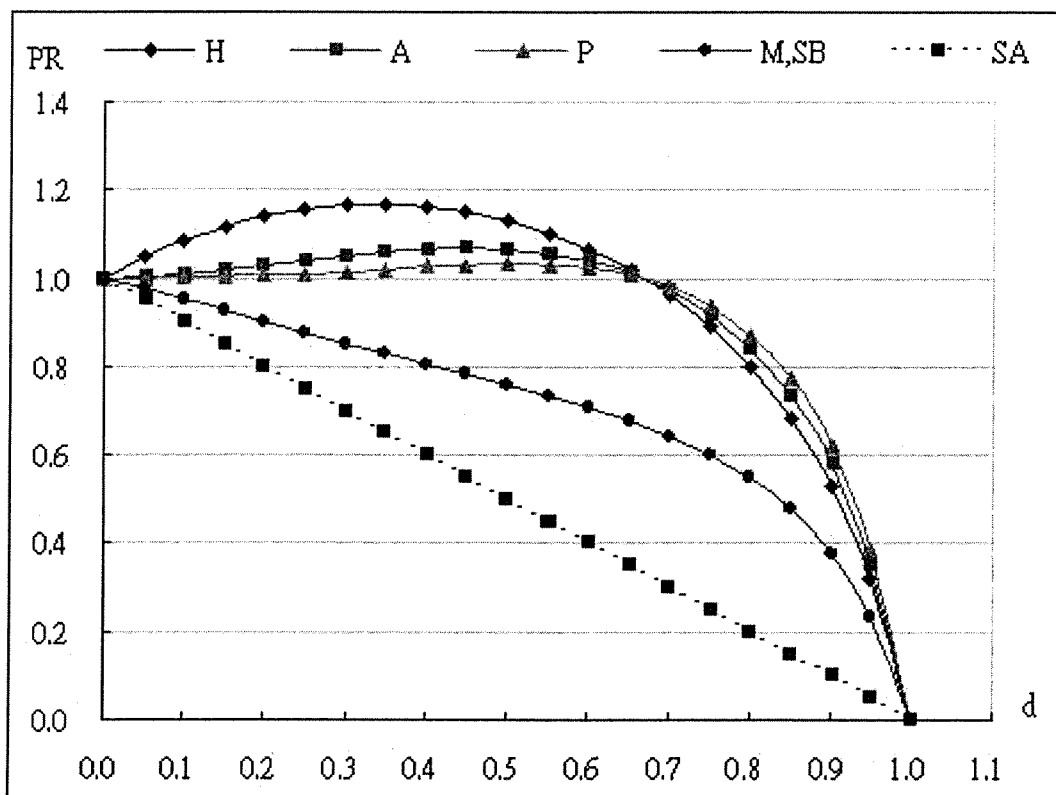


圖 22 d 值變化對各網頁 PageRank 值之影響

參考文獻

- 林共進、林勤豐 (2004)。搜尋引擎之最優排序: 回顧綜覽、管理意涵與研究展望。 <http://bm.nsysu.edu.tw/tutorial/htt/dkl/search01.pdf>
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* **30**, 107-117.
- Haveliwala, T. H. and Kamvar, S. D. (2003). The Second Eigenvalue of the Google Matrix. <http://www.stanford.edu/~taherh/papers/secondeigenvalue.pdf>
- Mangalindan, M. (2003). Seeking growth, search engine google acts like ad agency. *Wall Street Journal* **Oct. 16**.
- Mcuhan, R. (2000). Search for a top ranking. *Marketing* **Oct. 19**, 47.
- Meghabghab, G. (2001). Google's web page ranking applied to different topological web graph structures. *Journal of the American Society for Information Science and Technology* **52**, 736-747.
- Notess, G. R. (1999). Rising relevance in search engines. *Online* **23**, 84-86.
- Page, L., Brin, S., Motwani, R. and Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. *Stanford Digital Library Technologies Project*.
- Rogers, I. (2002). The Google Pagerank Algorithm and How It Works. <http://www.iprcom.com/papers/pagerank/>
- Ward, E. (2000). Market through 'Link Analysis' to improve popularity quality. *Advertising Age's Business Marketing* **85**, 32-33.
- Weidlich, T. (2002). Search engine marketing revving up. *Catalog Age* **19**, 3-6.

[民國 94 年 2 月收稿, 民國 94 年 4 月接受。]

SENSITIVITY ANALYSIS ON DAMPING FACTOR IN GOOGLE PAGERANK

Hwai-Hui Fu¹, Dennis K. J. Lin², Fengshan Bai³,

Hsien-Tang Tsai⁴ and Duan Wei⁵

¹Department of Business Administration, Shu-Te University,

²Department of Statistics, National Chengchi University and
Department of Supply Chain and Information Systems,

Pennsylvania State University,

³Department of Mathematical Sciences, Tsinghua University,

⁴Department of Business Management, National
Sun Yat-Sen University and

⁵Department of Banking and Finance, Tamkang University

ABSTRACT

Google--the largest search engine in the world--uses PageRank technology to determine the rank of website listings. It is thus important to have a decent understanding on PageRank technology for a fair ranking system. In this paper, we study the sensitivity on damping factor in PageRank, by investigating all typical linkage examples in the article "The Google Page Rank Algorithm and How It Works" (Roger, 2002). It is shown that the choice of the value for damping factor is critical to the PageRank. The degree of impact varies from one case to another. This ranking technology is called Citation analysis. Furthermore, Markov chain theory can be used here for modeling webpage's characteristics. The applications of Markov chain indicates that damping factor should be closer, but not equal to 1. These results support the choice of damping factor $d=0.85$ currently used by Google.

Key words and phrases: Citation analysis, markov chain, search engine.

AMS 2000 subject classifications: Primary 15A18.