

The Impacts and Values of Book Reference-Citation in Citation Networks

引文網絡中專書的影響與價值[#]

楊志堅*

「建立專書引文索引資料庫」的重要性，陳東升（2014）從其學術視野及臺灣特有的學術環境背景論述這件工作，本文則以計量（metrics）觀點，討論專書論著的文獻引用（references）及專書本身被引用（citations）的狀況，以及其在整體學術研究中的量化意義，並試圖考驗陳東升（2014）假說的可能性。本文先從學術現況實例顯現專書文獻引用及專書本身被引用的獨特意義；第二部分以統計學的學理基礎，尤其是社會網絡分析（social network analysis, SNA）的計量觀點試圖解讀專書引文的網絡意義；第三部分以模擬的數值資料作為說明，再佐以語言學門的專書引文試辦資料庫作為實徵範例。

一、專書文獻及專書被引的獨特意義

陳東升（2014）有兩個重要的假設支持其論述，第一：「專書被引用和引用的資料是可以描繪出專書作品學術影響的範圍，這樣的社群組群和以期刊的引用和被引用文獻所構成的社群範圍會有差異」；第二：「因為探究問題的性質和引文類型不同。以專書引文為依據可以顯現出一個完整的知識系譜，發掘專書開展出的學派或典範之理論源頭，並且爬梳這個學派在不同階段所

[#] 這篇文章若沒有臺灣大學陳東升教授的原創想法及推動，它不可能發生，科技部人文司吳淑真副研究員不厭其煩協調聯絡、國家圖書館蔣嘉寧博士與其研究團隊的盡力協助都是最重要的支持助力，敬表誠摯感謝。本文僅為作者個人想法，不代表任何官方立場，唯作者雖盡所有可能勘誤，若本文有疏失不當之處，概由作者個人承擔。

* 國立臺中教育大學教授、科技部人文司教育學門（含體育學、圖書資訊學）召集人。



產生的知識傳播效用」。在數學與統計科學的研究領域中，關鍵假設、假說的提出的重要性與這些假說的考驗及證明的重要性不分上下，有時候甚至成為某個重要論述創新改革的原動力，以下分別從幾個觀點討論陳氏假說的可行性。

其實現有國內外常見的期刊引文為主的期刊資料庫中，多少也會將期刊引註專書的狀況計算統計出來，換言之，專書在期刊引文的地位或影響力不就可以因此而被顯現出來？以專書為主角，計算專書內的文獻引用又有何特別意義？有一個例子可以說明，以臺灣知名生物統計學者梁賡義 (Kung-Yee Liang) 在 1986 年時發表於 *Biometrika* 這本期刊的一篇文章為例，當時他所提的方法 Generalized Estimating Equation (GEE) 篇幅不過 9 頁左右，隨後學界眾多學者因為這個想法所作的延續性或擴充性相關研究不知有多少！尤其在當時期刊論文所引用的數目更是如過江之鯽，這些後來發表的相關期刊論文，從某些角度來看，不是應該比原文更完備嗎？可是有趣的是，16 年後還有人願意為這個方法、這篇 1986 年的文章寫了一本長達 240 頁的書 (Hardin & Hilbe, 2002)，書名與當年 Dr. Kung-Yee Liang 及其同事在文章中 (Liang & Zeger, 1986) 所提方法的名稱一模一樣！16、17 年前的一篇期刊論文能被一本書所引用、論註，若要顯現這篇期刊論文在該領域的 Prestige 或 Authority 應該沒有更好的方式了，不是嗎？有趣的是，2002 年剛好也就是梁賡義老師當選中央研究院院士的時間，這可能不只是巧合吧？

所以，專書引文的建立不僅可以回應專書本身或專書社群的引用狀況，甚至填補了期刊引文資料庫一個長期被忽視但卻相當重要的訊息；請注意，以上的範例還是從講究時效性的數理科學領域的角度出發，若從講究長期社會影響力的人文社會科學領域而言，其相對重要性，就更不容忽視了。能被專書、專著所引用的期刊文章與僅能被同時期的同儕期刊所引用的期刊文章，有何不同影響力？這個問題，藉由以專書為主角的引文資料庫，才能合適且正確的回答。因此延伸而來的問題就是「專書所引的期刊或期刊引用專書」應該在各自引用的網絡中給予不同加權，就如同期刊引文因期刊的優劣應有權重上的差異，同理，專書引文或期刊引用專書也應有類似的考量。

此外，期刊引文常見的時限與專書引文時限的長短差異，也很容易預期，這個差異可能會造成明顯的歷史視野差異。期刊引文內專書比例與專書引文內的專書比例，自然也會有所不同，這些都是可能造成期刊引文網絡與專書引文網絡不同的重要因素。

二、專書在引文網絡分析之意義：學理基礎

期刊、專書、學術論著的引用情形，如果從社會網絡 (social networks) 的觀點來分析可以得到系統性的揭露。無論期刊或論文都需要引述一些已經具有 Privilege 或 Authority 的論述來支持或驗證自己的想法，就好像是人際社會網絡中友誼連結，「談笑有鴻儒，往來無白丁」(劉禹錫)，就是很好的說明。專書引文與期刊引文的來往對象，有其各自對應的特殊社群，應該沒有太大的問題。

(一) 小世界現象 (small world phenomenon)

不過，很有趣的，現代計量觀點的社會網絡分析卻反而說，鴻儒與白丁的網絡距離其實並沒有太遠；「遠在他鄉異地，卻能遇見同鄉、同胞」，不是常常讓我們驚嘆 What a Small World！社會網絡的數學分析可以很容易就證明一群彼此似乎無關的陌生人，只要少數個連結點 (connection)，就可以把這些個體連結成一個連繫在一起的網絡，這就是網絡分析方法中有名的「小世界」現象 (small world phenomenon; Travers & Milgram, 1969)。

因此，期刊與專書的引文連結或網絡距離似乎應該也不會太遠，不是嗎？只要是研究相同領域，專書或期刊論文的參考文獻 (references) 或被引文 (citation) 應該很快就會鏈結在一起，引文網絡的「小世界」(Small World; Burt, 2002 & 2004) 現象，只要有相關實徵資料，應該不難證實，既是如此，又何必勞費心力建立以專書為主角的引文資料庫？

(二) 結構洞現象 (Structural Holes)

換個角度從結構洞現象解讀引文網絡，視野卻完全不同。「樹倒猢猻散」就是解釋網絡世界的結構洞現象的最通俗的說法，原本緊密連結的猴群，只要「樹」一倒，也就是關鍵的連結點一旦消失，網絡內的原本連結各個獨立點，便從此無法彼此聯繫，專書所引述的文獻或作為被引述的來源自然比起篇幅較短的期刊論文的機會多出不少，因此合理的推測或假設專書可以成為引文網絡「結構洞」的機率應該也不低。這種很可能就發生於專書引文的引文結構洞現象，如何能忽視？

尤其，如果小世界現象與結構洞的情形同時發生於引文網絡中，整體網絡狀況又會如何？如果學術研究社群的大小規模有限 (例如：臺灣的某些學術領域或學門)，小世界現象是否加劇？加上傳統的「尊師重道」的習性，又有「大師」、「宗師」的階層組織結構，結構洞的問題是否更嚴重？兩因素交互作用下，專書引文的學術影響力是否會被加乘、加倍？這些足以充分反映學



術社群具體結構的關鍵問題，若沒有專屬的專書引文資料庫的實徵佐證資料，又如何能回答？

(三) 橋梁的衰退現象 (Bridge Decay)

大概沒有人會反對臺灣學界這幾年非常熱衷期刊論文的發表，相對之下，各方的數據也可以反映出臺灣學界本土的專書發表現況並不熱烈，甚至令人擔憂；網絡分析的領域中有個現象稱為 Bridge Decay (Burt, 2002 & 2004)，這個現象若用非專業的普通說法來形容臺灣的專書引用網絡的衰退，應該相去不遠。專書出版大量萎縮的結果，可以預期專書在整體學術引文網絡所占的連結（橋梁）地位的減退，尤其在長時間的作用下，Decay 的現象依現有文獻的研究結果指出應該會更加嚴重，提供專書引文的數值資料及相關資料庫，應該可以適當反映專書的重要性，或許在不久的將來可以導正一部分獨尊期刊發表及其引文的現況。

因此，僅有期刊引文的資料庫所能呈現的引文分布情形是不完整的，由此資料庫反映部分的引文網絡的結果也很可能僅是殘缺的學術研究脈絡。專書引文資料庫的建置不僅是重要的而且其時效性也同樣重要，而且可能是促成臺灣學術環境發展多元化及正常化的關鍵作為。

三、專書經典值 (Book Legendary Scores, BLS) 的意義與計算

好吧，就算以上所有的假設 (hypothesis) 或論點都超乎完美的逐一成真，專書引文與期刊引文確實有所不同，又如何？專書與期刊本來就不同，隨便找個路人甲、路人乙也能說兩者的厚薄、頁數就不同，兩者的不同又有何實質意義？

在運動競技的世界中，職業網球社群常常在討論一個永不過時的話題，「誰才是真正的世界第一？」、「誰才是真正的史上最強？」直接計算誰的勝場數或得勝率比較高，不就解決了？可是，因為沒有任何兩位選手的狀態是能長期完全公平的對比：年齡不同、職涯長短不同、對手程度不同、比賽的規模不同、甚至賽制也不同，僅僅比較勝場數的多寡，就像僅比較專書引文的條目數遠大於期刊論文的條目數，就直接說專書引文比較重要一樣可笑。所以，有沒有一種量化的分析方法足以公平的比較某本專書引文與某篇期刊論文在引文網絡中的「權威地位」？這個權威地位的數值是必須同時兼顧引用數量及引用品質的綜合指標，換句話說，引文世界中的專書或期刊的地位能不

能客觀的被比較？

Radicchi (2011) 及 Yang, Tsai and Chang (2013) 討論了網球選手威望值 (Prestige Score) 的概念，威望值的發明讓打敗了一位高品質的選手的加權分可以大於打敗了一位低品質的選手，也就是兩位選手雖然擁有一樣的勝場數，他們的威望值卻不見得相等，這個威望值的關鍵概念在於對兩兩選手間的對戰網絡的實質分析，而非將每場賽局一視同仁，Radicchi (2011) 宣稱獲得了對傑出網球選手的歷史地位的全新詮釋意義；而 Yang, Tsai, & Chang (2013) 也指出 Radicchi (2011) 的計算方式不見得適用在另一群選手的排名上，所以又提出了另一種針對選手網絡對戰品質好壞的數學加權方式。

專書或期刊引文在引文網絡世界中的地位，其實很像是運動選手之間對戰網絡的優勝劣敗？一篇論文能被嚴謹的期刊所引述，自然不能等同於被粗糙期刊引述而相提並論，專書既然不同於期刊，自然的，被專書所述或專書的參考文獻的意義或價值也應該要有別於期刊引文。因此，「專書經典值」(Book Legendary Scores, BLS) 的意義與價值就很容易理解了：理想的專書經典值 (BLS) 的內涵應至少能反映 (1) 專書引文權威值的獨特差異；(2) 專書引文在多年期間的長期 (經典) 地位。至此，專書經典值 (BLS) 的基礎與計算應該不難推導，依據前述 Radicchi (2011) 與 Yang, Tsai, & Chang (2013) 對網絡節點的加權方法，以及 Yang & Yang, C.C. (2003) 與 Muthén et al. (2002) 對於縱貫資料在一段時間內的成長或衰退情形的分類方法 (mixture of Growth)，理論上提出符合網絡架構的 BLS 加權方法應該不難。當然，任何新工具或新指標的提出都有其假設與前提，不盡然就能完美及全能，而任何濫用或誤用的指標都是另一個災難的開始。使用量化指標應搭配質性的批判及廣泛的多元觀點，才能符合客觀健康的視野。

四、數值資料作為說明範例

(一) 模擬研究

本文依據二項分配 (binomial distribution) 模擬了 4 本書 20 篇期刊論文的互相引用網絡，設計如表一：

表一

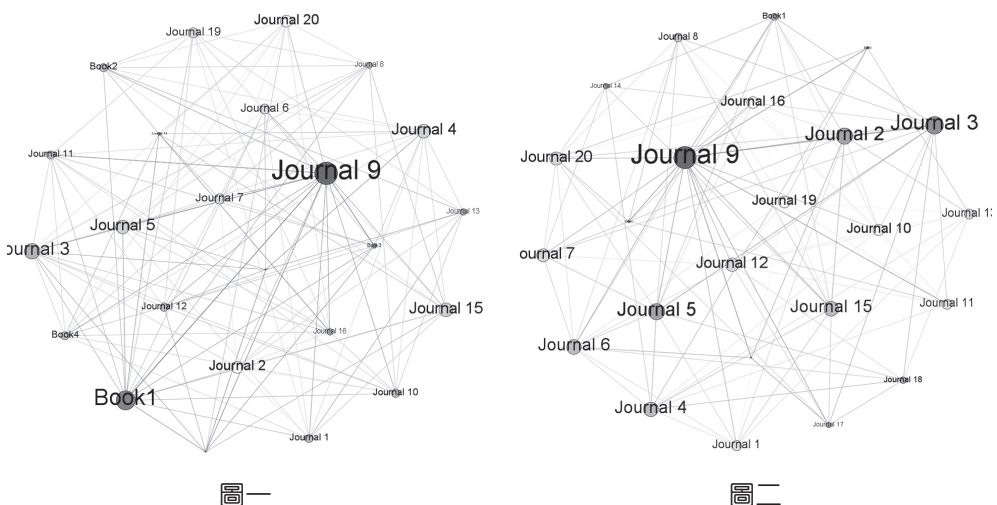
	Book 1, ..., Book 4	Journal 1, Journal 2, ..., Journal 20
Book 1, ..., Book 4	書引用書	書被書引用
Journal 1, Journal 2, ..., Journal 20	期刊引用書	期刊被期刊引用

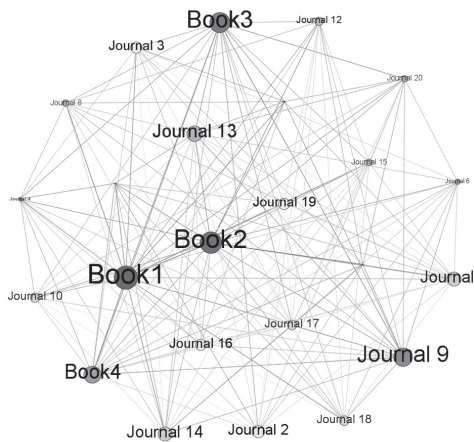


這個表所說明的模擬設計是：假設不論某篇期刊論文或專書彼此間的引用或被引用機率都是 30%，也就是「專書引用某篇期刊文章的機率是 30%、專書引用專書是 30%、期刊引用期刊 30%、期刊引用專書也是 30%」的「**無經典的純隨機引文網絡**」，亦即專書及單篇期刊論文的引用率是無差別的。執行這個數值模擬後，再依被引用的指標數值 (in degree) 比例形成節點大小及顏色，將所得引用數據繪製成以下的引文網絡圖，如圖一、圖二。其中，圖二是將專書互引及專書引用期刊的部分刪除，也就是現行國內無專書引文資料庫的情況下，以期刊的引文資料為主的分析方式，僅分析期刊互引或專書被期刊引述的現況。

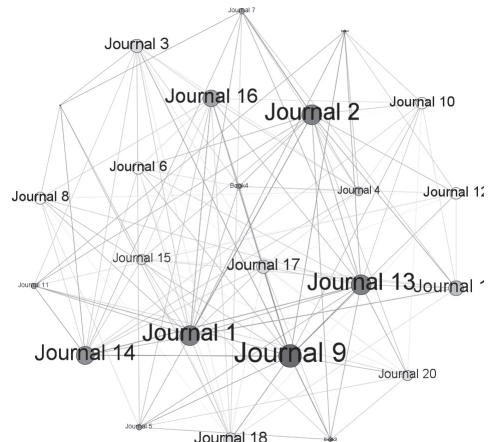
圖一則是包含了專書引文資料庫的理想完整引文網絡，其中，很明顯的，期刊 9 (Journal 9) 及專書 1 (Book 1) 是網絡中重要的引文網絡核心點。但是，臺灣的現況則是圖二的**移去專書引文** (專書互引、專書引期刊) 的殘缺網絡，此時僅維持期刊互引、期刊引專書的期刊引文網絡，如圖二。請注意：雖然**專書被期刊引述**的紀錄依然存在於新的網絡中，但是專書 1 (Book 1) 的核心地位不見了，其他專書的影響力更是大幅減弱至近乎無形，雖然期刊 9 的關鍵位置依然存在，但是一些單篇期刊論文的重要性 (例如: Journal 2, 3, 5) 則可能會被放大。

如果整體引文網絡中出現了「經典專書」的情況：「專書引用專書的機率是 80%、期刊引用專書的機率是 60%，而專書引用期刊的機率是 30%、期刊引用期刊也是 30%」，執行這個模擬設計所得的數據資料，繪製成具經典專





圖三



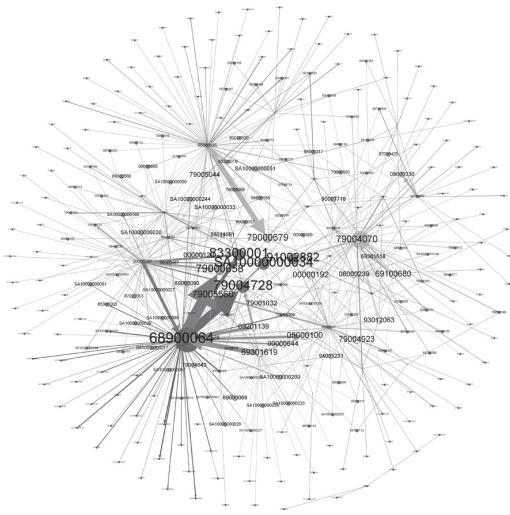
圖四

書的引文網絡的網絡圖，即如圖三、圖四。圖四是將專書互引及專書引用期刊的部分刪除，也就是現行無專書引文資料庫的情況下，以期刊的引文資料為主的分析方式，也就是只能記錄期刊互引或專書被期刊引述的現況。

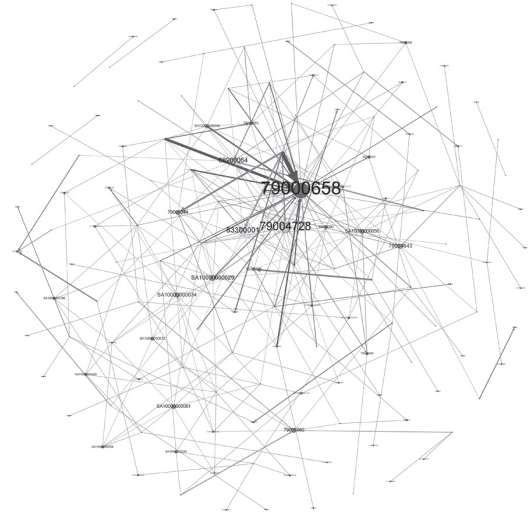
圖三是完整的引文網絡，專書 1, 2, 3 (Book 1, Book 2, Book 3) 都是預期中的經典專書，專書四 (Book 4) 及期刊論文 9 (Journal 9) 也有相當核心地位。但是如果移去專書引文 (專書互引、專書引期刊) 的網絡後，僅維持期刊互引、期刊引專書的網絡，如圖四。此時，可以明顯發現，雖然專書被期刊引述的紀錄依然存在網絡中，但是專書的核心地位都不見了，所有經典專書的影響力無法被顯現，期刊本身順序或領先地位應該不致動搖 (例如：Journal 9)，但是期刊整體的地位相對於原先的完整資料庫，其影響力或重要性又更大幅的被放大。

(二) 實徵範例：語言學的引文網絡

模擬資料畢竟是虛擬的，真實的臺灣學界現況又是如何？語言學門的專書引文資料庫經由科技部 (原國科會) 的補助及國家圖書館的同仁們的辛勤努力已經先行試辦，資料建置工作進行相當順利也獲得良好的成效，以下引文網絡分析即以目前已經建置完成的部分資料，作為前段文中模擬研究的實徵驗證。這些網絡圖的每一個節點都是一本在國內發行且具相當條件 (發行數量、期數等) 的期刊或專書；圖中，純數字的節點是期刊代碼，SA 加數字的節點則是專書代碼。由於這是試行資料庫，而且為了避免過度推論或其他不當用途，實際的期刊名稱或專書名稱已經隱匿不擬公布，讀者也不應妄加推



圖五 期刊引期刊、期刊引專書網絡圖



圖六 專書引專書、專書引期刊網絡圖

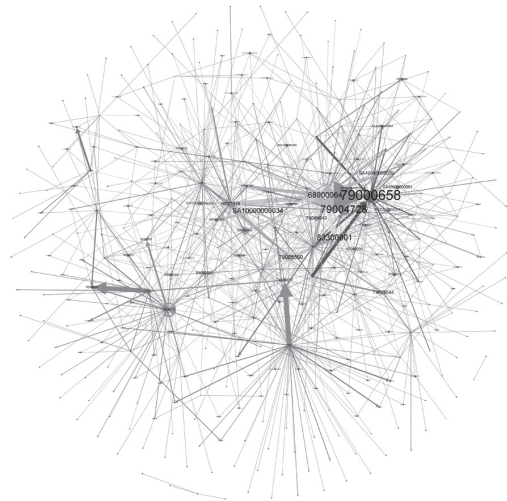
測。本次各式分析的專書約有 70 幾本左右，期刊大約有 30 本，其中實際納入分析的期刊**總篇數遠大於本數**，單就期刊單篇互相引用的次數就超過 1,400 多筆。另外，被引用的書籍或期刊出版年介於 1956-2013 年之間，引用書籍或引用期刊論文的出版期間則介於 1988-2013 年之間。

圖五是「期刊引期刊及期刊引專書」的網絡圖，這是傳統的引文資料庫所能提供分析結果。圖六則是「專書引專書、專書引期刊」網絡圖，亦即是建立專書資料庫後所能提供的資料來源，這些網絡圖的連結再依被引用的指標數值 (in degree) 比例形成節點大小及顏色深淺。節點越大或顏色越深皆顯示被引用次數越多，代表其所在引文網絡中的地位越加重要。以圖五為例，期刊 68900064 是在這個網絡中相對重要的節點，這本期刊引用他本期刊或專書的數目或是它自己的總被引用數都是相對多數；圖六則顯示 79000658 這本期刊被專書引用的情形是相當重要而值得注意。除了個別期刊或專書的地位之外，更重要而值得關注的是：這兩者的**網絡圖整體系統**所呈現出不同結構與連結的引文**生態差異** (ecological difference)，建議應以更宏觀的視野來檢視專書引文與期刊引文兩系統的不同處。

若是進一步將專書視為核心，排除引用數量最多的期刊論文間互引後，可得「專書引專書、專書引期刊、期刊引專書」的引用網絡圖，如圖七；這時候期刊 79000658 的重要性又再度顯現，而專書 SA1000000034 的重要性則



圖七 專書引專書、專書引期刊、期刊引專書



圖八 完整引文網絡

被凸顯出來。最後將完整的四種引用狀況都納入分析，則可得圖八的「完整引文網絡」圖；此時，期刊 79000658 依然重要，但在它的左方不遠處，專書 SA1000000034 似乎也有一席相當的地位。同樣的，若整體來看，兩張圖的引文生態系統也確實具有相當程度的不同。

五、結論與建議

就學理上的論述而言，專書引文與期刊引文應有其基礎上的不同與差異，將這些理論上的差異反映到數值模擬的設計上，也可以產生如預期一般的引文網絡差異，尤其當有經典專書出現時，兩者的引用網絡的差異會更加明顯。

真實學術社群的實際引用網絡涵蓋多種可能因子及未知結構，情形自然更加複雜，但就目前語言學門所收集的試辦資料中的期刊與專書引用情形而言，整體引文生態 (citation ecology) 的不同應該也無疑義。至於實際差異的形成原因應留待相關語言領域專家進行仔細的學理內涵實質辨認，並非簡單的數量檢測可以回應。

專書經典值 (Book Legendary Scores, BLS) 的提出及計算則是建置專書資料庫的重要關鍵配套措施，專書經典值 (BLS) 除了彰顯專書引文的獨特價值外，也要能反映專書的長期地位，可有別於期刊論文強調立即影響力的作



法，BLS 重要性不在話下。由於臺灣學界這幾年過度獨尊期刊論文的現象，影響所及，臺灣學界本土專書的發表現況令人擔憂，專書可以作為學術引文的長期穩定支撐的能力可能已經有所衰退。提供專書引文的數值資料（例如：BLS 指標）及相關資料庫，應該可以適當反映專書的重要性，讓專書的重要性具有實證資料的基礎，或許在不久的將來可以導正一部分獨尊期刊發表及其引文的不正常現況。不過，任何新工具或新指標的提出都有其假設與前提，任何濫用或誤用的指標都可能是災難的開始，使用者應該搭配出版物的品質專業批判及參考廣泛的多元指標，才是客觀健康學術態度。

參考文獻

- 陳東升 (2014)。〈建立專書引文索引資料庫的重要性〉，《人文與社會科學簡訊》，第 15 卷第 4 期，頁 42-45。
- 劉禹錫〈陋室銘〉。
- Burt, R.S. (2002). Bridge decay. *Social Networks*, 24, 333-363.
- Burt, R.S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, 110 (2), 349-399.
- Hardin, J., & Hilbe, J. (2002). *Generalized Estimating Equations*. London: Chapman and Hall/CRC.
- Liang, K.Y., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73 (1): 13-22.
- Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3 (4), 459-475.
- Radicchi, F. (2011). Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis. *PLoS ONE* 6 (2): e17249
- Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, 32 (4), 425-443.
- Yang, C.C., Tsai, L.T., & Chang, B.F. (2013). Reevaluating Rankings of Female Tennis Players of All Time by Win-or-Lose Network Analysis (in Chinese). *Sports & Exercise Research*, 15 (2), 219-228.
- Yang, C.C., & Yang, C.C. (2003). Longitudinal study of early Chinese linguistics growth: revealing sequential relations of linguistic antecedents and consequences. In Haruo Yanai, Akinori Okada, Kazuo Shigemasu, Yutaka Kano, and, Jacqueline J. Meulman, (eds.) *New Developments on Psychometrics*, pp. 109-116, Springer-Verlag Tokyo, Inc., Tokyo, Japan.