

行政院國家科學委員會專題研究計畫 成果報告

資料庫裡尋人脈：考掘新聞人物的社會網絡(第2年) 研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC 98-2410-H-004-118-MY2
執行期間：99年08月01日至101年01月31日
執行單位：國立政治大學新聞學系

計畫主持人：陳百齡
共同主持人：李蔡彥
計畫參與人員：碩士班研究生-兼任助理人員：蕭伊貽
碩士班研究生-兼任助理人員：劉凱欣

公開資訊：本計畫可公開查詢

中華民國 101年08月20日

中文摘要：本計畫旨在發展一套社會網絡分析工具，利用新聞資料庫為素材，藉以分析新聞人物的關係脈絡。第一年計畫建立一套社會網絡分析方法，透過個案分析，以理解新聞網絡資料蒐集流程的各種相關因素；第二年則承續第一年研究成果，比較不同資料源的社會網絡分析結果。

中文關鍵詞：人脈 社會網絡分析 電腦輔助報導 資料新聞學 資料可視化

英文摘要：This project aims to develop a set of tool to analyze social network of tarhgeted public figures. The researchers develop the procedures to extract data from news archive and trans form format data into relational type, and analyze/visulaize the data set. Two cases were employed to illustrate the possiblity that reporters use this tool in future disclose the network ties of their targeted sources of news.

英文關鍵詞：social network analysis, database reporting, data journalism, visualization

資料庫裡尋人脈：考掘新聞人物的社會網絡

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 98-2410-H-004-118-MY2

執行期間：2009 年 8 月 1 日至 2012 年 1 月 31 日

執行機構及系所：政治大學傳播學院

計畫主持人：陳百齡

共同主持人：李蔡彥

計畫參與人員：蕭伊貽、劉凱欣（研究助理）

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：

- 赴國外出差或研習心得報告
- 赴大陸地區出差或研習心得報告
- 出席國際學術會議心得報告
- 國際合作研究計畫國外研究報告

處理方式： 得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

中 華 民 國 101 年 8 月 1 日

資料庫裡尋人脈：考掘新聞人物的社會網絡

本計畫旨在發展一套社會網絡分析工具，利用新聞資料庫為素材，藉以分析新聞人物的關係脈絡。第一年計畫建立一套社會網絡分析方法，透過個案分析，以理解新聞網絡資料蒐集流程的各種相關因素；第二年則承續第一年研究成果，比較不同資料源的社會網絡分析結果。

數位人文取徑

「數位人文」(digital humanities) 研究取徑，是晚近結合資訊技術和人文研究的一個取徑，企圖尋找在前數位時代難以觀察的現象、無法想像的議題與無法進行的研究(項潔/涂豐恩, 2011)。數位人文學者認為，許多研究線索已經蘊藏在故紙堆中，先前研究人員用肉眼直接觀察文本，或許難以察覺線索，但資訊技術提供一個機緣，不僅用於的檢索和計量，更可以進一步利用各種資訊技術如文本探勘(text mining)、地理資訊(GIS)、社會網絡分析(social network analysis)、資料可視化(data visualization)等技術耙梳大規模歷史文本，藉以發掘一些學者意料之外的問題。運用這些技術雖然未必可以輕易回答這些問題、但過程中浮現的研究線索，卻很可能能夠刺激研究者，而開拓出研究新視野(項潔/翁稷安, 2011a)。

西方學界近年來蓬勃發展的社會網絡領域，是探索人類社會關係的知識取徑。這個研究取徑特別關注行動者和社會結構之間的關係(Lin, 2003)。社會網絡分析聚焦於行動者之間的連結關係，使用計量工具，把文數字資料轉化為網絡節點(nodes)和連結(links)，並由此出發，計算節點之間的計量關係，以計算節點和網絡成員之間的群聚程度、接近程度、或居間關係。並由此而得出派系(cliques)和社會圈(circles)。因此，社會網絡分析可用於考掘和呈現社群成員之間的關連。由於社會學研究傳統上一向關注社會行動的結構，而結構又建立在關係的基礎上，因此研究者可以通過社會網絡分析，解構社會關係的內涵(Scott, 2007)。

使用網絡分析考掘人脈

近來文獻顯示，網絡分析技術可用來幫助研究者發現歷史人物之間複雜的網絡關係。例如美國哈佛大學歷史學者 Robert M. Hartwell 和北京大學合作的研究計畫「中國歷代人物傳記資料庫」(CBDB) 便曾以資料庫進行群體傳記學研究，此後隨著技術進展，CBDB 從群體傳記學研究轉入人際網絡分析。

歷史的人際網絡研究所看中的，並非人物群體共享特性，而是許多一對一關係組構而成的複雜網絡。正如項潔和涂豐恩（2011：21）認為，雖然學者向來關注歷史人物之間的關係。但傳統使用人工分析方法，往往只能侷限在比較明顯的、單純的互動關係，一些比較隱晦的、複雜的互動關係，往往就難以兼顧。但經過電腦程式呈現之後，歷史學者便有可能深入觀察社會網路中不同節點的互動，並提出新的論述。例如，CBDB 計畫研究人員便曾經以網絡分析呈現宋代思想學派間的互動，從而指出思想史上一些不見經傳的學者，在當時扮演了不同學派間的溝通橋梁。

除了歷史學門，來自各個領域的學者也都使用透過資料可視化，發現隱藏的人脈關係。其中最著名案例之一，便是英國經濟學者 Richards (2005: 25-32) 便曾經以David Kelly 自殺案的幕後黑手。Kelly 是英國政府的武器專家，曾因涉入伊拉克戰爭大規模武器調查，在 2003年離奇自殺。Richards以英國政府網站提供的公開文件為素材，將所有提及 Kelly 姓名的政府公文、電郵、會議紀錄和備忘錄內容轉換成為關連資料。結果顯示，案發前後布萊爾政府高階官員之間交換訊息，且所有資訊均流向內閣長官辦公室，這個結果和當時政府官員「毫不知情」宣示大相逕庭。使用政府檔案資料，並透過資料流向，整體地反證出官員宣稱不可信。Richards 應用網絡分析解讀政府文件，發現政府高層掩飾真相，其研究動機和作法近似新聞工作者的調查報導，顯示社會網絡分析技術，同樣也可用在公共事務報導和調查報導工作中。

新聞工作者早已關注，有些個案已經實際應用於公共事務新聞報導。華盛頓郵報記者Cohen（2007）認為社會網絡分析是電腦輔助採訪報導的下一個世代的主流。他指出，藉由社會網絡分析新聞事件中的關係脈絡，可以讓記者產出更具深度的報導作品。社會網絡分析應用於新聞報導工作的實例近來較為著名的案例包括：加州《橘郡記事報》（Orange County Resister）2006 年揭發慈善勸募機構和電話詐騙犯相勾結，以及《華盛頓郵報》2007 年報導布希總統和德州油商之間的錯綜複雜關係；記者使用電話通聯或往來信件紀錄等關連資料進行網絡分析，得以揭露這些不為人知的關係脈絡。上述報導的所進行的網絡分析，素材來自蒐集政府或企業機構的資料庫，因此是電腦輔助採訪報導（Computer Aided Repeating, CAR）的一環。

社會網絡分析需要關連資料做為分析基礎。「關連資料」（relational data）是社會個體在行動中相互接觸、聯絡、依附或聚合的資料。研究者必須從文本中萃取出關連資料，以做為網絡分析的對象。基本上，關連資料來自於文本中的獨立事件（events），每個事件都必須有若干節點，研究人員以事件做

為單位，萃取出節點（可以是個人或組織團體）據以進行網絡分析。

人物網絡分析通常必須取自文本。歷史文本多存在紙本之中，包括個人日記、信件等文本內容，經過萃取，固然可以找到關連資料所需要的素材（沈崇麟、夏傳玲，2007）；大眾傳播媒體如報紙或廣播電視的文字或影像內容，經過處理，也可提供關連資料的來源。近年來，隨著數位典藏計畫的推展，可資利用的關連資料範疇正不斷擴展，除了原本文字、影像、聲音等傳統素材之外，透過程式自動擷取網路媒體如部落格、網路相簿或社交媒體（如 Twitter, Plurk, Facebook）而取得的資料，也可做為分析對象。

本研究旨在利用資料庫內容做為社會網絡分析的素材，企圖建立一套系統方法，展示新聞工作者利用資料庫再現人脈網絡的可行性。第一年計畫使用日治時期報紙《臺灣日日新報》資料庫，考掘一群新竹地方仕紳的人脈關係。第二年計畫使用李國鼎網路相簿上的資料，發掘李國鼎的人脈關係；並以其口述歷史所提及的人名做為比對，以發現兩種資料所提供的人脈再現的差異。

個案一：新竹仕紳的社會網絡

我們也透過個案研究探討網絡分析方法運用於輔助新聞報導的可行性。2006 年研究人員在棄置在新竹市郊寺院後方的一方石碑後方發現卅五個人名，這些人名經比對地方志書，發現均為日治初期新竹地方仕紳，但多數未在人物志留名。研究人員先從戶政檔案、土地申告書、縣志等地方史料來源，確認人物身份，再以人名檢索《台灣日日新報》資料庫，萃取出 1897 年至1926 年相關人物共計 2602 則報導資料，去除重複資料之後，獲得卅五人的 313 則新聞報導。每一則報導均視為獨立事件，每則新聞報導包括的人名，則被視為一個網絡中的節點。研究人員從每一則報導萃取出事件、時間和人名，便取得一套共計313個事件的關連資料。

研究人員使用軟體進行運算，產出網絡資料。再將軟體運算結果匯入製圖軟體，產出網絡示意圖，圖中節點（nodes）表徵卅年間所有人物的總網絡；連接節點的線條則表徵網絡成員之間的連結。由於這個示意圖所呈現的網絡相當龐大，各個節點和鍊結互相交錯，人類感官難以直接辨識；因此研究人員操作軟體，只留下表徵碑上人物的網絡，隱去其它人物資料，如此一來，便可以得到一張網絡示意圖，這張來取自新聞報導的網絡圖表徵碑上仕紳的關係脈絡。

由網絡分析數據所形成的派系，再比對這些仕紳的身份，可知這些人物雖為昔日仕紳，卻包括地主、街坊商人和文人三類身份。這樣一個異質的網絡成員，浮現出一個進一步的問題：到底為何會形成這樣一個網絡。由此，我們進一步分析當時新竹地方社會的情境，因而發現碑上人物在政權交替後企圖轉變家族事業經營模式的線索。本個案從報紙資料庫萃取資料紀行網絡分析，從而發現地方仕紳的異質化結合，進一步發現這群人正是臺灣地方近代化企業最早的推手。這種「從網絡分析發展為研究線索」的過程，正和前述CBDB 計畫發現佚名學者扮演宋代思想學派間橋梁角色的歷程，極為類似。

個案二：李國鼎的人際網絡

本計畫的第二年聚焦在「個人為中心」的網絡（ego-centered network），並以影像為資料分析對象。相對於第一年計畫旨在建立一套萃取資料的程序，第二年偏向以新聞人物為對象的網絡分析。這類分析特徵在以一個節點（個人）為社會網絡中心。這個個案案主為李國鼎。

李國鼎生於南京，早年留學英國，中日戰爭爆發後返國服務，1949年隨政府遷臺，於1953年加入台灣的工業委員會起，1988年退休，歷經美援會、經合會、經濟部、財政部及行政院政務委員，素有「台灣經濟推手」、「台灣科技教父」、「資訊工業舵手」之譽，是戰後台灣經濟發展的主要領航員之一（康綠島，2001）。我們以李國鼎為個案，主要考量資料的豐富度。李國鼎生平史料相當完整，並已進行數位典藏，並有口述歷史材料可資佐證。

網絡分析的主要素材為數位相簿。李國鼎生前相簿捐贈中研院近史所，並進行口述訪談，所有相片影像均附有圖說，公開於近史所網站。研究人員擷取相片圖說，我們使用圖說文字做為分析對象，每則圖說均有事件、時間、相片人物姓名等。我們將每一張相片當作一個獨立事件（分析單位），形成一個社會網絡；每一位相片上的人物姓名，便是社會網絡上的節點。由此研究人員整理出相同格式之關連資料，開始進行網絡分析，除總網絡圖之外，再以李國鼎任職政府的四個時期分別紀行分析，以發現四個時期不同的節點。

過去網絡分析多以文字為主，本計畫所進行的分析則是以影像做為分析對象。為了瞭解影像為主的網絡再現，和文字為主的網絡再現有何不同，因此研究人員擷取李國鼎口述歷史的文字敘述，以作者陳述的獨立事件做為分析單位，進行分析。結果發現，口述歷史展現多為政府體系中和傳主平行或高階的人物，部屬較少出現在口述歷史的網絡再現當中。但影像為主的再現網絡則較無此偏倚。

研究啟示

誠如多位學者指出（費孝通，1947；金耀基，1992；喬健），「人脈」是華人社會日常生活運作的重要概念，新聞工作者在發想蒐尋訊息或查證線索都必須理解和運用人脈。以人為諮詢對象的傳統方法，有其侷限性，因此有必要尋求器物工具的輔助。社會網絡分析方法本質上是一種是當代社會理論所衍生的知識工具，旨在將複雜的人際網絡訊息轉換成可視資料，用以強化人類辨識和運用社會網絡的能力。正如過去統計應用於精確新聞報導（precision journalism; Meyer, 1991）。社會網絡分析做為一種知識工具，不僅適可以應用於史家考據；同時也可用於輔助新聞報導，幫助新聞工作者辨識和理解新聞事件中的的人際網絡。使用網絡分析並非仰賴電腦替決定人物和關係，而是透過器物輔助，將人類感官無法直接解譯的抽象關係，透過多媒材的再現，而得以辨識和運用，因此或可避免人類認知的偏執或誤導。

我們認為，新聞工作者必須認識人脈，但人脈隱藏在情境之中，難以直觀。但透過社會網絡分析我們得以透過資料庫採集關連資料；並再現新聞事件中的社會網絡、由此而探索特定新聞人物的社會關係，以從事新聞報導。換言之，由於當代資訊科技進步，讓社會網絡從理論知識走向物質化、具體化。本文聚焦於萃取新聞資料庫內容從事網絡分析，研究人員在過程中除了展現這項應用的可行性，同時也體認到跨領域知識整合的重要性，未來應用網絡分析於新聞工作時，必須結合資料科學學門的相關技術。

參考書目

1. 金耀基（1992）。〈關係和網絡的建構〉，《中國社會與文化》，頁 64-85。香港：牛津大學出版社。
2. 沈崇麟、夏傳玲（2007）。〈總序：社會研究方法的現狀及其發展趨勢〉，劉軍（譯）《社會網絡分析法》，頁 VII。四川：重慶大學出版社。
3. 康綠島（2001）。《李國鼎口述歷史：話說台灣經驗》。臺北市：卓越世界文化。
4. 項潔、涂豐恩（2011），〈什麼是數位人文〉，《從保存到創造：數位人文研究的發端》，臺北：臺灣大學。
5. 項潔、翁稷安（2011），〈導論—關於數位人文的思考：理論與方法〉，《從保存到創造：數位人文研究的發端》，臺北：臺灣大學。
6. 喬健（1982）。〈關係芻議〉，楊國樞/文崇一（編）《社會學及行為科學研究的中國化》，台北：中研院民族所。
7. Hockey, S. (2004). History of Humanities Computing. In S. Schreibman, R. Siemens & J. Unsworth (Eds.), *A Companion to Digital Humanities*. Oxford: Blackwell.
8. Lin, Nan (2006). Social capital. In Jens Beckert and Milan Zagiroski (eds.). *The Interenational Encyclopedia of Economic Sociology*, New York: Rutlege.
9. Meyer, P. (1991). *The new precision journalism*. Bloomington: Indiana University Press.
10. Richards, S. (2005). A social network analysis into the David Kelly Tragedy, *Connections*, 26(2), 25-32.
11. Scott, J. (2007). *Social network analysis: A handbook*. 2nd edition, New York: Sage.
12. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York, NY: Cambridge University Press.

國科會補助專題研究計畫成果報告自評表

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

 達成目標 未達成目標 (請說明, 以 100 字為限) 實驗失敗 因故實驗中斷 其他原因

說明：第二年計畫中，原預計以新聞資料庫之李國鼎新聞文本做為資料比對之用，但在過程中和某知識庫廠商洽談使用付費，發現資料價格過高（依照資料量計費，五十年資料費近三十萬元），不合乎比例原則，因此改萃取李國鼎口述歷史內容之人物，做為比對。

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

其他：第一年成果已發表，共二篇；第二年成果仍在撰寫中。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本計畫意義、價值、創新和未來發展，依序陳述如下：(1) 跨領域知識整合：本計畫是一個跨領域研究，針對新聞資料庫文本資料進行分析，以發現公眾人物的社會網絡。本計畫以兩年為期，研究人員自資料庫萃取新聞文本，藉由網絡分析處理人物社會關係脈絡，並應用電腦圖學呈現資料。(2) 萃取傳媒內容做為分析素材：網絡分析需要關連資料，本研究使用報紙資料庫文本、網路相簿影像，做為分析素材。結果顯示，當代傳播媒體產生的內容，的確可以做為人物網絡分析的素材來源。(3) 邁向資料新聞學 (data journalism)：新聞學界先前提出「電腦輔助報導」，近年來則提及「資料新聞學」，其核心概念便在使用有系統的科學方法，以分析電腦資料庫或網路上產出的大量資料，做為新聞線索或素材。本研究過程證實概念可行，並具體展現資料處理的解題策略，本研究結果呼應資料新聞學發展趨勢，結果將可應用於新聞媒體的深度報導和調查報導之用。

(4) 未來當可繼續和資訊學門合作，發展鉅量資料分析 (Big Data Analysis)：鉅量資料已成為當代傳媒現象，人物關係資料也有此問題，本研究下一步發展應和資料探勘領域學者繼續合作，發展鉅量資料技術。

國科會補助計畫衍生研發成果推廣資料表

日期:2012/08/20

國科會補助計畫	計畫名稱: 資料庫裡尋人脈: 考掘新聞人物的社會網絡
	計畫主持人: 陳百齡
	計畫編號: 98-2410-H-004-118-MY2 學門領域: 電訊傳播/傳播科技
無研發成果推廣資料	

98 年度專題研究計畫研究成果彙整表

計畫主持人：陳百齡		計畫編號：98-2410-H-004-118-MY2					
計畫名稱：資料庫裡尋人脈：考掘新聞人物的社會網絡							
成果項目		量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）	
		實際已達成數（被接受或已發表）	預期總達成數（含實際已達成數）	本計畫實際貢獻百分比			
國內	論文著作	期刊論文	1	0	100%	篇	本計畫第一年成果部分研究結果已被引用（引註於《黃旺成先生日記》）
		研究報告/技術報告	0	0	100%		
		研討會論文	2	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	3	0	100%	人次	
		博士生	0	0	100%		
博士後研究員		0	0	100%			
專任助理		0	0	100%			
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
博士後研究員		0	0	100%			
專任助理		0	0	100%			

<p>其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p>	2011年11月南京大學新聞傳播學院講座
--	----------------------

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本計畫意義、價值、創新和未來發展，依序陳述如下：(1) 跨領域知識整合：本計畫是一個跨領域研究，針對新聞資料庫文本資料進行分析，以發現公眾人物的社會網絡。本計畫以兩年為期，研究人員自資料庫萃取新聞文本，藉由網絡分析處理人物社會關係脈絡、並應用電腦圖學呈現資料。(2) 萃取傳媒內容做為分析素材：網絡分析需要關連資料，本研究使用報紙資料庫文本、網路相簿影像，做為分析素材。結果顯示，當代傳播媒體產生的內容，的確可以做為人物網絡分析的素材來源。(3) 邁向資料新聞學(data journalism)：新聞學界先前提出「電腦輔助報導」，近年來則提及「資料新聞學」，其核心概念便在使用有系統的科學方法，以分析電腦資料庫或網路上產出的大量資料，做為新聞線索或素材。本研究過程證實概念可行，並具體展現資料處理的解題策略，本研究結果呼應資料新聞學發展趨勢，結果將可應用於新聞媒體的深度報導和調查報導之用。(4) 未來當可繼續和資訊學門合作，發展鉅量資料分析(Big Data Analysis)：鉅量資料已成為當代傳媒現象，人物關係資料也有此問題，本研究下一步發展應和資料探勘領域學者繼續合作，發展鉅量資料技術。