

社會網路之建置、分析與視覺化： 以台灣學術社群網路為例

Construction, Analysis and Visualization Social Networks: Exemplified by the
Academia Social Network in Taiwan

李政德
Cheng-Te Li

E-mail : r96944015@csie.ntu.edu.tw

張峻銘
Chun-Min Chang

E-mail : r95137@csie.ntu.edu.tw

國立台灣大學資訊網路與多媒體研究所研究生

Graduate student, Graduate Institute of Networking and Multimedia, National Taiwan University

劉建邦
Chien-Pang Liu

E-mail : r95137@csie.ntu.edu.tw

陳尚澤
Shang-Ze Chen

E-mail : b95110@csie.ntu.edu.tw

國立台灣大學資訊工程學研究所研究生

Graduate student, Department of Computer Science and Information Engineer, National Taiwan
University

林守德
Shou-De Lin

國立台灣大學資訊工程學研究所教授

Professor, Department of Computer Science and Information Engineer, National Taiwan University

E-mail : sdlin@csie.ntu.edu.tw

【摘要 Abstract】

本論文主要針對學術社群網路進行自動建置、分析與具體視覺化的工作。首先，我們採用資料探勘的技術在網際網路上自動識別台灣學術社群之不同種類的個體以及他們之間的關係。並用研究人員、學術機構、關鍵字、以及其他相關的學術資訊間的關係建構全台灣學術社群之網路。接著，我們引進社會網路分析的方法，對已建構好的學術網路進行分析，計算這些學術實體間的重要特性，如群聚係數、中心度，進而得出該網路的一些特性。最後，我們實作出視覺化的系統具體呈現學術社群網路。

This paper discusses the process of constructing, analyzing, and visualizing an academia social network of research communities in Taiwan. Our work can be divided into three portions. First, we use web mining

technologies to identify on the Internet various kinds of research communities in Taiwan and then construct an academia social network with information of researchers, academic organizations, keywords, and other relevant academic materials that characterize these research communities. After the construction of the academia social networks, we apply the methodology of social network analysis to calculate and evaluate some important factors reflect some properties of the network, such as the clustering coefficient and centrality values of the various research communities. Finally, we design a visualization system to concretely display this academia social network to users.

關鍵詞 Keyword

社會網路分析 學術社群 資訊視覺化 資料探勘

Social network analysis ; Academic research society ; Information visualization ; Data mining

壹、簡介

近來，隨著數位資訊的發展、電腦計算能力與容量的提升、Web2.0 的竄紅，人們對於網路的仰賴與日俱增，許多以往需由人工完成的任務均已數位化，人與人的交流更打破時空的限制，建構在網際網路上的應用不計其數，如即時通訊軟體、虛擬社群、線上聊天室、電子佈告欄、網路資料庫等。於是，互動交流的資料累積相當迅速，尤其在所謂社群網路服務(Social network service)或稱社交媒體(Social media)的網站，藉由提供資訊交流的平台，建立起以人、事、物為實體的虛擬網路，通稱為社群網路(Social network)。

從學術研究的角度來看社群網路，通常我們會以圖(Graph)的結構來表示，將社群網路中的實體(Entity)與關係(Relation)分別表示成圖的點(Node)和邊(Edge)，若以一般常見的社交網站服務為例來看，實體為人，關係為兩人彼此認識，或是有特定的互動（如共同參與某事件）。一般常見的社群網路為單一關係網路(Single relation network)，意即網路中的實體類型僅為一種，而網路中的關係類型也只有一種，如同上述的人際網路，實體之間只有認識的關係。然而，這種單一關係網路並不能涵括真實世界的資訊，它所呈現出來的只是複雜網路(Complex network)中的其中一個角度，換句話說，想要窺探真實世界網路，僅僅透過單一關係網路是無法達到的。為此，我們提出多重關係網路(Multi-Relational Network, MRN)的概念，可視為一個網路中包含眾多單一網路的資訊，多重關係網路的實體與關係均可以有各種類型，換另一種角度來解釋，多重關係網路為許多單一關係網路之疊合。同樣以人際關係網路為例，實體的類型可以是人、機構、事件、地點等，而關係則為這幾類型的實體交互組成，例如說人參某些事件。

儘管多重關係網路所涵蓋的資訊較接近真實

世界，但目前國內外關於此種網路的相關研究仍不多。而近兩年專注於學術網路之社會網路分析的研究，並與多重關係網路相關的研究，以DBconnect (Zaiane, Chen & Goebel, 2007) 與 ArnetMiner (Tang, et al., 2008) 為代表。DBconnect 是從一個蒐集資訊領域相關學術著作的網路資料庫(Digital Bibliography & Library Project, DBLP, <http://www.informatik.uni-trier.de/~ley/db>)中，定義一個擷取學術網路的模型，並採取類似網頁排名(PageRank)之隨機作業(Random work)機制來偵測不同領域的研究社群，並推薦出指定領域的重要研究人員。ArnetMiner 更進一步設計出從網路上自動擷取研究人員相關資訊的演算法，將擷取的資訊與現有的線上著作資料庫整合，以建構出完整的多重關係學術社群網路。其中，它使用 Hidden Markov random field 之機率模型對同人不同名的問題做修正，最後全部整合於 Author-conference-topic 模型，能夠將各領域的專家找出來，並提供發現依據關係類型來對學術網路做搜尋與排名的服務。

本研究主要提出如何建立並分析多重關係網路研究的方法與架構。我們以台灣學術社群網路為例，以程式自動化的方式從網際網路中自動擷取學術實體，如學生、教授、關鍵字、系所學院，並建立起實體間的彼此關係，形成多重關係網路。本研究另一個重心在於對於台灣學術社群網路進行網路特性之計算與探勘，透過分析網路的特性，我們將可得知不同實體扮演的角色與重要性、實體間的距離與群聚性、不同類型實體的關連性都能因而獲得解答。本研究的最後一個貢獻在於將台灣學術社群網路具體視覺化呈現出來，我們開發出「關係探勘器」(Relation explorer)，讓使用者可以在與系統的互動中，探得學術網路不同實體類型各種組合彼此關係的真實面貌，此外，該系統也提供簡單全域與個人化的網路量測與排名。

貳、學術網路建置

本研究的資料收集網站，提供選單列出所有論文資料，因此可達成 100% 之資料收集率。在無法預知節點總數的情況時，Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee (2007) 及 Ahn, Han, Kwak, Moon, & Jeong (2007) 等人描述在不確定總下載節點個數的情況下如何抓取資料。國內徐慧成 (2003) 整合國家圖書館的碩博士論文資料和國科會的研究人才資料，分析台灣學術界的社會網路結構。張秀儀(2004)的資料來源與徐慧成(2003)相同，而研究方向在於自動建構學術界的知識來源映射圖，協助使用者了解專家的專長及聯繫專家的管道。

本文說明如何在網際網路上面擷取並建置台灣學術網路系統。我們的資料來源是網際網路上的中文學位論文資料庫，藉由自動抓取相關的學位論文資料後，經由後續處理成 GraphML 格式檔案，再進行分析，並以視覺化的方式來呈現。

中文電子學位論文服務(Chinese Electronic Theses & Dissertations Service, CETD, <http://www.cetd.com.tw>)資料庫，是華藝數位股份有限公司於西元 2005 年所推出的博碩士論文資料庫

服務網站，提供台灣、中國大陸、香港及澳門等大專院校的博碩士論文。目前共收錄 19 所台灣的大專院校，約 5 萬筆的學位論文。

由於中文學位論文資料庫的資料相當龐大，我們透過電腦程式進行查詢與下載來取得。為此，我們開發了一個自動網頁擷取程式，經由校院瀏覽方式取得每一篇論文的相關欄位資料，每一篇學位論文中都包含了學生名稱、指導教授、論文題目、關鍵字、語言、頁數等資訊。

關於我們從 CETD 所取得台灣學術社群網路，其中學生與教授的數量資訊描述如表 1。經由驗證發現在 5 萬多篇碩博士學位論文中，有 7,500 篇左右的作者名字有重複，而如何利用知識發現的方法找出有多少作者為同名不同人，將會是我們未來研究重點之一。

為了便於社群網路資料格式的傳輸，我們採用 GraphML 標準來儲存學術網路的資料。GraphML 是一種專門用來儲存圖的一種 XML 標準格式，透過結構化的描述，利於跨系統或軟體的傳輸與應用。我們定義了六種實體、九種有向邊(Directed edge)，分別列於表 1 與表 2。

表 1
節點的名稱與個數

節 點 名 稱	個 數
<學生>	51,302
<指導教授>	8,210
<學院>	117
<系所>	615
<中文關鍵字 >	4,290
<英文關鍵字 >	4,082

表 2
關係的名稱與個數

關係名稱(節點, 節點)	個 數
指導關係(<指導教授>,<學生>)	50,699
共同指導 (<指導教授 A>,<指導教授 B>)	9,756
學生之研究領域(<學生>,<關鍵字>)	112,581
指導教授之研究領域(<指導教授>,<關鍵字>)	101,076
指導教授所屬學院(<指導教授>, <學院>)	8,100
指導教授所屬學系(<指導教授>, <系所>)	10,051
學生所屬學院(<學生>, <學院>)	50,815
學生所屬學系(<學生>, <系所>)	51,045
學系所屬學院(<學院>,<系所>)	726

對於每一種節點，我們在其上定義的屬性說明如下：

- <學生> 節點的屬性有：type (std, 此為常數)、name (學生中文名字)、eng_name (學生英文名字)、chn_thesis (論文中英文標題)、eng_thesis (論文英文標題)、thesis_type (學位別)、pub_year (出版年)、page (總頁數)、language_type (論文书寫的語言別)
- <指導教授> 節點的屬性有：type (prof, 此為常數)、name (指導教授中文名字)
- <學院> 節點的屬性有：type (clg, 此為常數)、name (學院名稱)
- <學系> 節點的屬性有：type (dpt, 此為常數)、name (學系名稱)
- <中文關鍵字> 節點的屬性有：type (chn_keyword, 此為常數)、name (關鍵字內容)、chn_keyword_freq (關鍵字出現頻率)
- <英文關鍵字> 節點的屬性有：type (eng_keyword, 此為常數)、name (關鍵字內容)、eng_keyword_freq (關鍵字出現頻率)

參、台灣學術社群網路分析

在這小節中，我們描述如何用社群網路分析的方法來解釋台灣學術網路。

一、Average Path Length (APL)

一個網路的平均距離即是其中所有節點兩兩間最短路徑的平均長度。一般來說，大部分的社群網路其直徑和網路的大小無關，並且大約是 Small-world phenomenon (Watts & Strogatz, 1998) 的長度，也就是大約是 6。Small-world effect 代表發生在此網路上的活動的強度，例如直徑小時，則網路裡資訊傳播得很快。網路平均距離的計算方式如下：

$$APL = \frac{\sum_{u,v} l_{u,v}}{|E|}$$

其中 $u, v \in V$ ， $|E|$ 為邊之個數， $l_{u,v}$ 為節點 u 與 v 之最短距離。

我們對這個學術社群網路做了人與人間的最短距離的分析，其中最長的距離可達 9，表示在這

個網站上的人們之間最遠的關係可能只間隔 8 個學校、學生或是關鍵字，然而現實中他們可能會有更多關係，距離會更短，而最短的距離是 1（指導教授和其學生的關係），平均距離則是 4.4320335，由於學生或教授可能會和學院或是關鍵字有連結，可以看成是這個網站上搜集論文的學生之間，平均只相隔了 2 個領域（學生或教授→關鍵字→學生或教授→關鍵字→學生或教授）。

二、Network Cluster

一個網路中節點的群聚係數 (Clustering coefficient) 量化了它相鄰的節點之間的靠近程度。Watts 和 Strogatz (1998) 引進了這種測量來決定一個網路是否為一個小世界網路。

一個節點 v_i 的群聚係數 C_i 即為它直接相鄰節點之間真正存在的連結數，除以它們之間可能存在的最大連結數。假設一個節點 v_i 有 k_i 個相鄰節點，則這些相鄰節點間最多可以有 $k_i(k_i - 1)$ 個可能存在的有向連結，因此一個有向圖中一個節點的群聚係數即是：

$$C_i = \frac{|neighbor(v_i)|}{k_i(k_i - 1)}$$



圖 1 教授、學生與學校之間的連結關係

三、Degree Correlation

哪些節點會有連結，除了可能和節點的類型有關，和節點的度值 (Degree) 也許也會有關係。度值相關性 (Degree correlation) 就是計算在這個學術網路中，一個度值高的節點是否和相同度值的節點有連結。

Newman (2002) 提出一種度值相關性的計算公

其中 $|neighbor(v_i)|$ 代表節點 v_i 現有的連結個數，也就是 v_i 的 In-degree 與 Out-degree 之和， $v_i \in V$ 。更進一步而言，整個系統的群聚係數即為對每個頂點的係數平均：

$$\bar{C} = \frac{1}{|V|} \sum_{i=1}^{|V|} C_i$$

其中 $|V|$ 為整個網路之節點總數。

對整個學術社群網路應用了以上的方法可得到整體的群聚係數為 0.24244578，這算是低的數值，表示在這個社群網路中，節點成群聚集的現象不明顯。低的群聚係數表示這個網路的重複性比較少。從某種角度來看，相鄰節點之間的關係，可以經由 v_i 建立，所以其間並不需要特別有連結，比如說以下的連結 (圖 1) 代表某個教授指導一個學生，而這個教授來自某個學校。其實這樣的連結已經隱含了這個學生也是屬於這個學校的一員，所以無須特別將此學生與學校做一個連結。所以低的群聚係數某種程度上可以代表這個網路中並沒有太多重複的訊息。

式如下：

$$r = \frac{m^{-1} \sum_i j_i k_i - (m^{-1} \sum_i \frac{j_i + k_i}{2})^2}{m^{-1} \sum_i \frac{j_i^2 + k_i^2}{2} - (m^{-1} \sum_i \frac{j_i + k_i}{2})^2}$$

其中 j_i 和 k_i 是第 i 個邊的兩個節點的度值，而 $i = 1, 2, \dots, m$ 。

計算這個學術網路的度值相關性，得到 -0.10255072，因為其數值的範圍可能為-1 到 1 之間，接近 0 的相關性，表示這個網路中節點的度值並沒有特別相關。

四、Mixing Pattern (Newman, 2003)

在研究社會網路的領域，社群中人們彼此連結的方式是否與它們自身的屬性或種類有關呢？又個體間的連結是否會受語言、種族和年齡或其他事物的影響？這可以透過混合樣式(Mixing pattern)來探討。

如果人們傾向於和同類的人之間有關係，我們說這個網路表現出同配連接傾向 (Assortative mixing or Assortative matching)。如果反之他們和不同類型的人們比較有關聯，我們把這叫做異配連接傾向 (Disassortative mixing)。友誼通常可以發現它的特質是同類相近 (Assortative by most characteristics)。在這個實驗中，我們想知道同校教授會不會有比較相同的研究方向。我們首先把所有研究領域相同的教授配對(Pairs)找出來 (圖 2)，然後再統計這些教授屬於某些學校的機率。

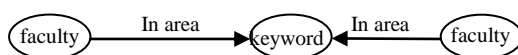


圖 2 教授與關鍵字間的連結關係。

表 3

台灣學術社群網路中，教授間研究領域相關程度之混合樣式表

	臺灣師範大學	臺灣大學	淡江大學	清華大學	元智大學	中興大學
臺灣師範大學	0.052	0.0646	0.0226	0.0135	0.0173	0.0162
臺灣大學	0.0646	0.1759	0.045	0.0374	0.0382	0.0429
淡江大學	0.0226	0.045	0.0172	0.0093	0.0134	0.0118
清華大學	0.0135	0.0374	0.0093	0.0101	0.0089	0.0094
元智大學	0.0173	0.0382	0.0134	0.0089	0.0118	0.0104
中興大學	0.0162	0.0429	0.0118	0.0094	0.0104	0.0115

表 3 是選擇了資料裡論文數最多的六所學校，對其中教授之間的研究相關程度計算做比較，而所謂研究的相關程度定義為：在該網路中，我們任取兩個連到相同關鍵字的節點 v_i 和 v_j ，該二節點分屬兩個不同的學校之邊際機率 (Marginal probability)，即 $P(v_j|v_i)$ 與 $P(v_i|v_j)$ 。其基本之假設為教授之研究方向通常可由指導學生之論文看

出，而關鍵字正是該論文研究方向之代表，因此我們透過關鍵字來決定兩位教授之研究是否相同，進而以此計算研究領域之混合樣式。我們將這六所學校兩兩之邊際機率算出後，填入一矩陣 M ，如表 3。例如，淡江與清華之間的比例是 0.093，表示如果我們任意取兩個連到一樣關鍵字的教授節點，第一個節點屬於淡江大學，而第二個節點屬於清華大



學的機率是 0.093。同配連接傾向的係數定義如下：

$$r = \frac{\sum_{j=1}^6 e_{ij} - \sum_{j=1}^6 e_j^2}{1 - \sum_{j=1}^6 e_j^2}$$

其中 $\sum e_{ij}$ 代表矩陣 M 中，對角線之值總和， e_j 表示每所學校對於所有學校各別之邊際機率。

當 $r = 0$ 表示沒有同配連接傾向，也就是教授所屬的學校和他們主題之間的研究沒有關係。 $r = 1$ 表示是完美的同配連接傾向，如果是同樣學校的教授，他們一定研究相關的主題。而不同學校一定研

究不同的主題。 $r = -1$ 表示同樣學校的教授一定研究不相關的主題。這個學術網路中我們計算出 $r = 0.05$ ，代表同校教授間同配連接傾向程度不明顯。

五、Centrality

Degree centrality：一個網路的中心節點為網路中與其他節點之互動最為頻繁之節點，其中互動的頻繁程度定義為該節點的度值，當節點的度值愈高，則表示該節點愈可能是網路的中心。

表 4

台灣學術社群網路中 Degree Centrality 前五名之列表

	學生	教授	中文關鍵字	英文關鍵字	學院
1	賴澤君	李弘暉	類神經網路	APOPTOSIS	臺灣大學工學院
2	江慧賢	張百棧	細胞凋亡	TAIWAN	臺灣大學電機資訊學院
3	溫演福	柯承恩	服務品質	DATA MINING	臺灣師範大學教育學院
4	王正德	鄭春生	基因演算法	GENETIC ALGORITHM	元智大學工程學院
5	洪嘉祥	湯玲郎	公司治理	PERFORMANCE	臺灣大學生命科學院

表 4 說明的是每種不同類的節點 Degree centrality 最大的前五名。比如說教授通常是有指導很多學生並且研究領域廣，而關鍵字的前幾名代表有很多人的研究與此相關。而學院前幾名代表其轄下的教職員學生很多。

Closeness centrality：第二個計算節點中心度 (Centrality) 的方式是以其與網路中所有節點最短距離的平均值為準。也就是中心的節點應可以在最短距離內達到網路中所有節點。這種量度基於一個節點對於其他節點的距離，所以被稱為 Closeness

centrality。在學術網路之中，一個節點和其他節點很接近時，這表示他可能是一位有多個研究領域的教授，或是一個很常被使用的關鍵字。

我們計算出來的結果如表 5 所示，就我們的觀察，這些教授們的研究領域通常為跨領域研究；而關鍵字中度值高的通常 Closeness centrality 也不會太低，但是其中也有度值不特別高，但是比較容易連到不同領域的研究學者的關鍵字（如台灣）。學院方面也顯示跨領域的學院有比較大的機會成為 Closeness centrality 高的節點。



表 5

台灣學術社群網路中 Closeness centrality 前五名之列表

	教授	中文關鍵字	英文關鍵字	學院
1	李弘暉	類神經網路	TAIWAN	臺灣大學工學院
2	張百棧	台灣	DATA MINING	臺灣師範大學教育學院
3	陳文華	服務品質	PERFORMANCE	臺灣師範大學理學院
4	邱昭彰	基因演算法	SIMULATION	臺灣大學理學院
5	莊立民	資料探勘	INTERACTION	臺灣大學生物資源暨農學院

六、Hypertext Induced Topic Selection (HITS) (Ding, Zha, He, Husbands, & Simon, 2001)

在一個網路裡探勘或搜尋時，我們常會需要評估一個節點的重要性，一個被許多頁面或節點連到的節點很可能是重要的，這種 In-degree 很大的節點叫做 Authority；一個連到許多節點的節點，提供了有用的資源，也很可能是重要的，這種 Out-degree 很大的節點叫做 Hub。

一開始各個節點都有初始的 Hub 分數 y_i 和 Authority 分數 x_i ，而一個好的 Hub 應該連到許多好的 Authority，一個好的 Authority 也會被許多好

的 Hub 連結。因此，我們可以藉由遞迴的方式，將 Hub 和 Authority 定義如下：

$$\text{Authority score : } x_i = \sum_{j:e_{ji} \in E} y_j,$$

$$\text{Hub score : } y_i = \sum_{j:e_{ij} \in E} x_j$$

HITS 分數是取決於節點彼此的連結狀況，以學術網路中的學生節點來說，學生節點可能連到所屬的學院，因此如果所屬的學院是一個許多人連到的 Authority，分數就會比較高。表 6 列出 HITS 中 Hub 和 Authority 都很高的節點。

表 6

台灣學術社群網路中 HITS 前五名之列表

	學生	教授	中文關鍵字	英文關鍵字	學院
1	廖思翰	黃漢邦	有限元素法	FINITE ELEMENT METHOD	臺灣大學工學院
2	李臻誠	曾惠斌	類神經網路	MEMS	臺灣大學電機資訊學院
3	陳雅媚	廖運炫	有限元素分析	GENETIC ALGORITHM	元智大學工程學院
4	彭建國	郭斯傑	二氧化鈦	FINITE ELEMENT ANALYSIS	臺灣大學管理學院
5	闕延洲	陳希立	奈米碳管	DATA MINING	臺灣師範大學教育學院

肆、網路視覺化

資訊視覺化(Information visualization)的目的在於有效率地讓使用者從各種不同的面向觀察資料的特性。近來，資訊視覺化尤其備受社會科學以及電腦科學領域重視，著重於由各種不同實體與關連(Entity-relation)所組成的架構之視覺化，以利相關研究人員對特定資料的了解與分析，這個資訊視覺化的子領域被稱作網路視覺化(Network visualization)。社會網路的視覺化最早可追溯至2000年 Freeman 對社群互動的視覺化原則，他主要利用實體的顏色、位置、形狀與大小來表示網路中的不同資訊意義，他同時提出，若能伴隨電腦產生的不同 Layout，如 Force-directed 與 Radical layouts，對資料的呈現將提供更多不同的觀點。

自從 Freeman 提出網路視覺化的基本原則後，相繼有人基於他的理論做了一些應用。首先，Nardi 等人於2002年提出一個名為 ContactMap 的系統，該系統主要是利用電子郵件的往來資料實作出個人社會網路的視覺化，使用者可以手動替自己經常通信的親朋好友進行分類，系統會在視覺化上以不同的顏色展示與自己有接觸的不同社群，該系統主要用於個人化信件管理。然而，ContactMap 實際上並沒有包含將人與人之間的互動，僅是以人為實體的呈現。2004年，Fisher 和 Dourish 二人提出另一個類似 ContactMap 的系統，同為以電子信件為資料來進行視覺化，可視為 ContactMap 的改進，比較不同的地方是，Fisher 等人加入了社會學的觀點，即為社會網路所具備的兩種面向：Whole network 與 Ego-centric network。前者表示網路的全貌，後者表示以個人為中心的網路視覺化。

上述兩者的視覺化呈現均著重於靜態網路，另一種能夠與使用者有豐富互動的視覺化方式是動態呈現。在網路視覺化中加入動態效果的一個典型

是 TouchGraph (<http://www.touchgraph.com>)它是一個以圖形化呈現特定領域實體之間關連的架構，其「領域」如於人、組織、知識、網站。TouchGraph 同時實作了 Freeman 所提出的 Force-directed layout，提供與使用者的互動功能。後來，Viegas, Boyd, Nguyen, Potter, and Donath (2004)將視覺化加入時間元素，開發了 Social network fragments 的系統，視覺化時間序列上信件往來的變化，讓使用者容易回顧與特定群體通信的網路變化。

關於將網路視覺化應用於學術網路的相關研究，Fujimura, Fujiyoshi, Hope, and Nishimura (2006)提出一個針對真實世界學術交流的視覺化系統，尤其用於學術會議的圓桌會談，藉由特製感應器紀錄學術人員參與討論的紀錄，將每次的參與交流的事件作為連結，建立起學術交流網路。Nakazono, Misue, and Tanaka (2006)則提出階層式結構之網路視覺化，主要用於在加入時間因素後，透過分層的方式呈現網路的變化，並且可將各階層作適當之整合。Ortega 和 Augillo (2007)則藉由現成軟體 NetDraw (Borgatti, 2002)來做學術網路的視覺化與分析，該學術網路之實體均為學術機構網站，關係則為透過網際網路之連結。

儘管網路視覺化這個領域已經有不少人提出相關的系統與架構，然而，先前的研究均著重於單一關係網路(Single relation network)，對於我們所提出的多重關係網路(Multi-relational network)的資訊呈現，幫助是有限的。因此，我們提出並實作了關係探勘器來視覺化我們提出的多重關係網路。

關係探勘器主要目的在於具體化呈現社會網路中的實體與實體之間的關係，相關研究人員與一般使用者可透過本系統觀察特定多重關係網路，它同時也作為往後社會網路分析相關研究的平台。關係探勘器與傳統視覺化系統最大的改進在於提供不同關係類型(Relation types)的觀察面向，並且如同一般搜尋引擎的搜尋方式，正確且有效率地呈現

出指定的子圖(Subgraph)。它在 Ego-centric 與 Whole networks 的互動性讓使用者能簡易地識別實體在網路中扮演的角色。此外，我們實作了一些常見的量測社會網路的計算準則於其中，如：PageRank, HITS, Centrality 等，並提供對不同類型實體的排名。

如同所謂的網路或圖形，我們視覺化是透過點與邊的結構來呈現的。在我們所取得的台灣學術網路資料中，我們識別了點的類型有學生、教授、系所學院、關鍵字四種實體類型(Node type)，而這些

構成了指導、屬於、研究領域關鍵字三種關係類型(Relation type)。例如：X 教授指導 Y 學生，Y 學生屬於 Z 科系，Y 學生論文研究領域的關鍵字為「關係探勘器」。在網路的視覺化中，對實體會以其文字名稱呈現，透過邊來連結不同實體。當使用者查詢特定的實體，系統會對應呈現與該實體相關的其他實體。以上例來說，當使用者查詢 X，則會出現相互連結的其他類型實體有 X, Z, A, B, C。圖 3 為關係探勘器的截圖。

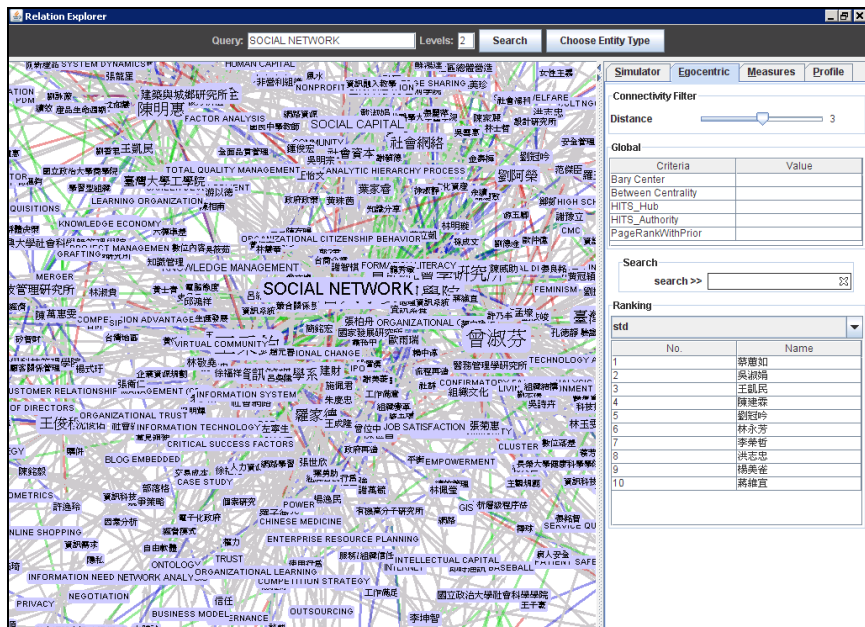


圖 3 關係探勘器的截圖

以「SOCIAL NETWORK」關鍵字為查詢對象。圖的最上方為提供使用者下查詢的地方，與提供的參數，以及選擇欲呈現的實體類型。下方主視窗被切割為左右兩部份，左半部為網路視覺化之動態呈現，右半部包含四個頁籤，分別提供網路量測、個人屬性、個人化分析、動態網路校調等。

關係探勘器的視覺化系統設計是遵從

Shneiderman (1996)所提出來的視覺化法則「先給全貌、在互動中觀察與過濾、依據使用者需求呈現細節」。當查詢被輸入後，首先被呈現出來的是根據查詢的子圖全貌，使用者在這個時候可以透過平移與縮放來觀察該網路，並利用我們提供的一些互動元素來過濾觀察不同的資訊面向，若想看得更詳細，可以點擊實體，會對應出相關屬性與度量網路



的一些計算值。以下我們針對系統的各個部份做更詳細的介紹。

一、參數化查詢與實體類型

關係探勘器為以查詢為主的關係探勘系統，使用者可以在最上方輸入已知的對象。在我們所取得的台灣學術網路中，可輸入的實體對象為教授、學

生、關鍵字與系所學院。在下查詢的同時，使用者可指定欲從該查詢向外擴張幾層(Level)關係，從圖論(Graph theory)的角度來講，即利用深度做為優先搜尋之深度為多少。另一方面，使用者可以點擊選擇實體類型，不同的類型將展示網路不同面向的資訊。參數化查詢與實體類型選擇如圖 4 所示。

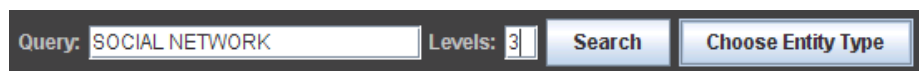


圖 4 參數化查詢與實體類型選擇圖

二、視覺化介面與互動

根據 Freeman (2000)的網路視覺化設計原則，動態的介面可讓使用者有豐富的互動感受，進而做到資訊瀏覽的目的。基於此原則，關係探勘器採用 Force-directed layout，當大量實體與關係所構成的網路產生時，系統會自動做各實體彼此間力與力必須達到平衡的物理模擬，逐漸地將畫面上的實體調整到邊相交最少、點重疊最少的程度，使用者亦可於右方的模擬器(Simulator)頁籤內手動調整各種模擬參數。而點擊拖曳各個實體時，也會改變整個網路的平衡狀態。此外，使用者可以透過滑鼠左鍵平移整個視覺化畫布，或透過滑鼠右鍵上下拖放達到畫面縮放的效果，藉此一覽全貌或深入細部探索。

三、個人化瀏覽

個人化瀏覽的方式在於藉由與使用者之間的互動性來達到，我們的系統提供了以下幾項功能性互動元素：

(一)以度為基礎的實體顯示

在視覺化網路中，可以見到不同實體的大小是不一樣的，我們的系統會計算各個實體的度，根據度值來顯示實體的大小。使用者同樣可以在模擬器中的 Enhanced revealing filtering 手動調整實體大小，可讓網路中比較重要的實體呈現愈明顯。

(二)強調關係與實體(Entity-Relation Highlighting)

由於有多種關係類型，若同時呈現出來，使用者將難以辨別，因此系統會自動依據不同類型的關係顯示不同的顏色。當整個網路的實體與關係的數目很大，將造成使用者瀏覽上的困難，因此關係探勘器提供強調的功能。使用者可於 Ego-centric 頁籤中的搜尋列中輸入查詢，系統會自動將符合查詢的實體透過顏色強調顯示。當滑鼠移動到任一實體上時，除了會透過顏色強調外，與該實體直接連結的實體亦會被強調。如圖 5 所示。

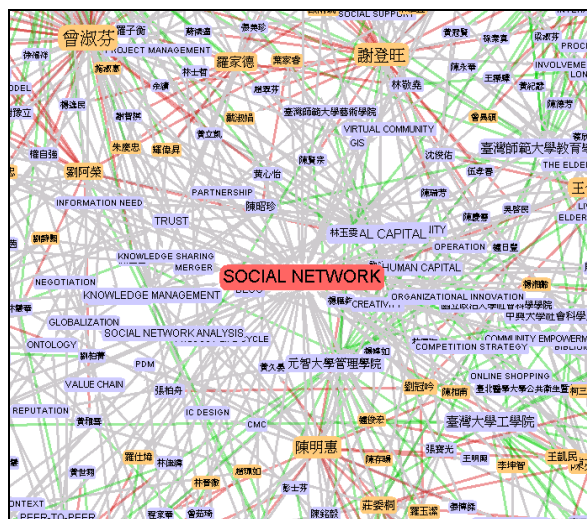


圖 5 關係與實體強調圖

(三)依查詢為中心的深度過濾(Ego-centric Connectivity Filtering)

使用者可於 Ego-centric 頁籤中指定依特定實體瀏覽時的深度 d ，然後在網路視覺化畫布中點擊不同實體(如 X)，即可見到整個網路會轉變為以被點擊的實體 X 為中心，向外擴展 d 層。透過這種過濾式瀏覽，可以見到在不同層分別是不同的實體，愈裡面與指定實體的關係愈強烈，愈外面與指定實體的關係則愈微弱，因此通常建議 connectivity 值為 2 或 3 最為恰當。

(四)個人屬性與排名(Profiling and Ranking)

針對不同類型的實體，所擁有的屬性之數目與名稱不盡相同。當使用者在我們的系統中點擊實體時，將於 Profile 頁籤內對應產生該實體的屬性列表，如此一來，使用者在瀏覽網路的同時可隨時查看實體的屬性。另一方面，在使用者於最初於上方輸入查詢後，系統會自動利用 PageRank 計算不同類型實體於對應子圖的排名，使用者可於 Ego-centric 頁籤中選擇不同類型實體來觀看。如圖 6 所示。



Ranking	
chn_keyword	
No.	Name
1	社會網絡
2	社會資本
3	社會網路
4	組織文化
5	點對點網路
6	社會網絡分析
7	信任
8	競爭策略
9	層級分析法
10	演化

圖 6 以「SOCIAL NETWORK」關鍵字為查詢兩層的子圖下，相關的中文關鍵字排名

(五)網路量測(Network measures)

針對社會網路分析的部份，關係探勘器提供 Global 與 Local 兩種角度的量測。Global 的量測的對象是我們取得的整個台灣學術網路資料，主要是計算網路平均長度與度的統計分佈量測。Local 的部份則以常見的實體重要性來度量，有 Bary Center(量測實體接近網路邊緣的程度，值愈大表示愈接近邊緣，愈小表示愈為網路中的核心)、PageRank(透過與自己連結的實體的重要性與數目來衡量自己的重要性)、HITS(由 Hub 與 Authority 所組成，Hub 值為指向重要的 Authority 的程度，Authority 為被重要的 Hub 所指向的程度)、Betweenness Centrality(透過圖論中的最短路徑來衡量實體的重要性，被最短路徑經過越多次者，其值愈大)。

伍、結論

本文提出了關於社會網路應用於學術社群之完整流程，該學術社群網路為多重關係網路。先是自動從中文學位論文資料庫網站擷取學術實體資訊，表示為 GraphML 標準格式以建構學術網路。

接著，我們對整個學術網路進行社會網路分析之相關量測，如群聚係數、實體重要性、混合樣式、度的分佈等。最後，我們開發關係探勘器以視覺化該學術社群網路，除了彈性的互動式經驗，使用者可透過查詢輕易地瀏覽網路的全貌，並且局部進行分析與觀測。這三個部份串連成社會網路分析的典型流程，有別於傳統的單一關係網路，我們提出來的多重關係網路更接近真實世界的實體互動。未來，我們將制定更多屬於多重關係網路的衡量準則，並在關係探勘器上加入專屬多重關係網路的演算法與元件。

針對台灣學術之多重關係網路所進行的分析，我們可以稍微了解目前台灣學術研究之社會網路之現況：透過網路平均距離之計算，我們得知台灣學術界之研究社群的關係相當緊密，兩個研究領域不同的教授或學生僅僅相隔兩個領域；低群聚係數告訴我們，這種多重關係網路的建構方式是精簡而有效率的，網路中少有重複之訊息；低度值相關性顯示台灣學術網路中，教授與學生之間、教授與教授之間，做研究是較不受彼此的影響力所影響的；低混合樣式顯示台灣的研究型大學的研究領域相當多元，即便同校的教授也未必做相關的研究；透過 Centrality 的計算，我們可得知近幾年熱門的

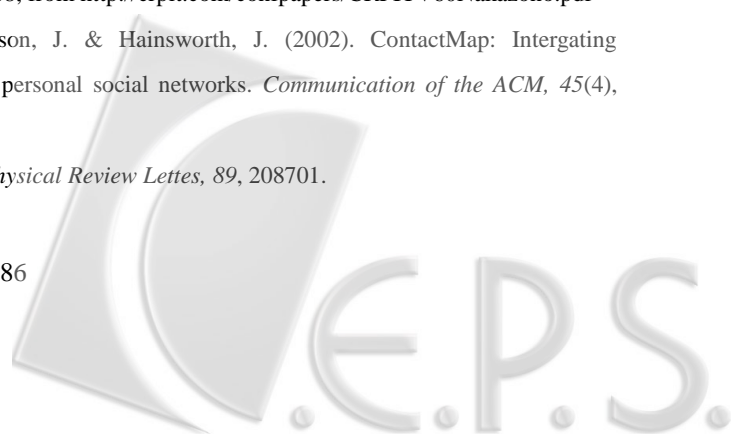
研究議題為何，同時顯示跨領域研究為目前台灣學術研究的一大潮流；HITS 值之計算則呈現學術網路中較具影響力的各類型節點；而透過關係探勘器，相關研究人員將更容易觀察學術網路中不同類

型節點彼此的互動，如學術社群之形成以及上述各種分析指標對於以個人為中心之子圖的結果。

(收稿日期：97 年 6 月 9 日)

參考書目

- 徐慧成 (2003)。利用網頁資訊建構多階層指導教授與研究生之網絡關係。未出版之碩士論文，國立中山大學資訊管理研究所，高雄市。
- 張秀儀 (2004)。利用全國博碩士論文資料庫自動化建構知識來源映射圖。未出版之碩士論文，國立中山大學資訊管理研究所，高雄市。
- Ahn, Y.-Y., Han, S., Kwak, H., Moon, S. & Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th International World Wide Web Conference*. Retrieved October 24, 2008, from <http://www2007.org/papers/paper676.pdf>
- Borgatti, S. P. (2002). *NetDraw: Graph visualization software*. Harvard, MA: Analytic Technologies.
- Ding, C., Zha, H., He, X., Husbands, P. & Simon, H. (2001). Link analysis: Hubs and authorities on the world wide web. *SIAM Review*. Retrieved October 24, 2008, from <http://www.nersc.gov/~simon/Papers/hits5.hits5.pdf>
- Fisher, D. & Dourish, P. (2004). Social and temporal structure in everyday collaboration. In Dykstra-Erickson, E., & Tscheligi, M. (Eds.), *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp.551-558). New York: ACM.
- Freeman, L. (2000). Visualizing social networks. *Journal of Social Structure*, 1(1). Retrieved October 24, 2008, from <http://www.cmu.edu/joss/content/articles/volume1/Freeman.html>
- Fujimura, N., Fujiyoshi, S., Hope, T. and Nishimura, T. (2006). Tabletop community: Visualization of real world oriented social network. In *Proceedings of the 14th Annual ACM international conference on Multimedia* (1035-1036). Retrieved October 24, 2008, from <http://portal.acm.org/citation.cfm?id=1180868>
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P. & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. Retrieved October 24, 2008, from <http://www.imconf.net/imc-2007/papers/imc170.pdf>
- Nakazono, N., Misue, K. and Tanaka, J. (2006). Nel2: Network drawing tool for handling layered structured network diagram. In K. Misue, K. Sugiyama, and J. Tanaka, (Eds.), *Proceeding of Asia-Pacific Symposium on Information Visualization* (pp. 109-115). Retrieved October 24, 2008, from <http://crpit.com/confpapers/CRPITV60Nakazono.pdf>
- Nardi, B., Whittaker, S., Isaacs, E., Creech, M., Johnson, J. & Hainsworth, J. (2002). ContactMap: Intergating communication and information through visualizing personal social networks. *Communication of the ACM*, 45(4), 89-95.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Lettes*, 89, 208701.



- Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, 67(2), 026126.
- Ortega, J. L. & Augillo, I. F. (2008). Visualization of the Nordic academic web: Link analysis using social network tools. *Information Processing and Management*, 44(4), 1624-1633.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualization. In IEEE Symposium on Visual Languages (Ed.), *Proceedings of the 1996 IEEE Symposium on Visual Languages* (p.336-343). Wash., D.C.: IEEE Computer Society.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp.990-998). New York: ACM.
- Viegas, F., Boyd, D., Nguyen, D., Potter, J. & Donath, J. (2004). Digital artifacts for remembering and storytelling: PostHistory and social network fragments. In R. H. Sprague (Ed.), *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. New York: IEEE.
- Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440-442.
- Zaiane, O. R., Chen, J. & Goebel, R. (2007). DBconnect: Mining research community on DBLP data. In *Proceedings of the Joint 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis*. Retrieved October 24, 2008, from http://workshops.socialnetworkanalysis.info/websnakdd2007/papers/submission_1.pdf

