

教育資料與圖書館學

Journal of Educational Media & Library Sciences

<http://joemls.tku.edu.tw>

Vol. 50 , no. 4 (Summer 2013) : 565-596

利用網路結構分析的研究主題視覺化
A Study of Research Topic Visualization
Using Network Structure Analysis

林 頌 堅 Sung-Chien Lin

Assistant Professor

E-mail: scl@cc.shu.edu.tw

[English Abstract & Summary see link](#)

[at the end of this article](#)





利用網路結構分析的研究主題視覺化

林頌堅

助理教授
世新大學資訊傳播學系
Email: scl@cc.shu.edu.tw

摘要

路徑搜尋網路(PFNet)方法與社群偵測演算法經常被應用在研究主題的視覺化呈現與分析。PFNet方法能夠在保留原先網路的結構特性下，刪除大量不重要的連結線。社群偵測演算法則能夠將網路劃分成凝聚性子群。然而這兩種方法都有不足的地方：PFNet方法無法自動從輸入的網路上發現重要的子群，社群偵測演算法無法保證同一子群的節點會映射在鄰近的區域。本論文建議整合這兩種方法以減輕上述的問題：利用社群偵測演算法將PFNet方法產生的新網路劃分成子群。且本研究也建議利用子群內出現頻率較高的詞語做為研究主題的標示，讓結果分析與詮釋更加容易。本研究以臺灣資訊傳播學領域為範例，利用相關系所的碩士論文為分析資料。研究結果發現：整合PFNet方法和社群偵測演算法有利於從論文相關網路上發現代表重要研究主題的子群。子群內最高出現頻次的詞語大多和資訊傳播學以及其基礎領域的問題、方法、理論和技術非常相關，可以做為研究主題的標示。

關鍵詞：研究主題分析，資訊視覺化，路徑搜尋網路，社群偵測演算法

緒 論

過去已有相當多研究將資訊視覺化(information visualization)的概念與方法應用在資訊計量學，將相關論文本身的特性以及彼此間的關係製成容易進行視覺思考(visual thinking)的圖表，根據產生圖形上呈現的型態(patterns)發現重要研究主題，並從研究主題上分布的論文資料推測其中蘊含有關的研究實務，例如發現從事各研究主題的重要研究人員、重要研究機構以及研究的發展趨勢。進行資訊視覺化時，可針對分析的學術領域，蒐集在一段時間內發表的論文，從研究問題與蒐集的論文資料，估測論文之間的相關性。相關性較大的

2012/11/20投稿；2013/05/24修訂；2013/06/13接受

論文代表它們可能探討相似的研究問題，從相同的理論觀點來分析研究問題，或利用一樣的研究方法與工具。由於探討相同主題的論文極有可能會參考並引用相同的文獻，在內容裡也會包含許多相同的詞語，一般資訊視覺化的研究常利用論文間引用資料或詞語出現情形的相似程度來評估它們彼此間的相關性，包括以引用資料為基礎的共被引次數 (co-citation counts) 統計訊息 (如相關係數) 和以詞語出現為基礎的向量空間模式 (vector space model)。最後再使用映射演算法，根據相關性測量得到的結果，將論文的關連繪製成各種圖表，藉由圖形呈現的型態，提供分析人員探索學術領域的研究主題以及主題間的關連，並利用視覺化的結果解釋研究發展的情形 (Börner, Chen, & Boyack, 2003; Morris, & Van der Veer Martens, 2008)。

為使得產生的圖形能呈現論文間的相關性，資訊視覺化的研究需要根據分析的問題特性以及論文間的相關性，選擇能夠表現這些特質的圖形類型，並計算出所有論文資料在呈現圖形上的位置。這過程稱為「部署」(ordination) (Börner, Chen, & Boyack, 2003)。在學術領域的研究主題分析中，網路式 (network) 的呈現方式是經常應用於部署的圖形類型。運用網路圖進行研究主題分析的構想是希望藉由相關論文被映射到圖形上距離較近的範圍，透過密集的映射點與連結線在視覺上形成的群聚效果，讓資料分析者發現可能的主題，且資料映射點間的距離愈小、連結線愈多，便表示這群論文間的關係愈強，也就是這個主題愈重要 (Small, 2006)。網路圖的優點是可以表現出研究領域的整體樣貌，藉由密集叢聚的節點與連結線呈現出領域內的重要研究主題，了解整個領域的知識結構 (intellectual structure)；再加上網路分析 (network analysis) 演算法的快速進步 (Börner, Sanyal, & Vespignani, 2007)，使網路圖的呈現方式可以分析各節點的中心性 (centrality)，並觀察出哪些群體存在較密集的關係 (Chen, 2006)。但利用圖形上密集的節點與連結線判斷研究領域的重要研究主題時卻也同時產生了一個問題：愈重要的研究主題有著越多的資料、越複雜的相關性，使得網路圖上密集的区域反而不易清晰辨認出研究主題包含的所有資料 (Chen, & Morris, 2003; Small, 2006)。

更詳細地說，網路圖的資料表現方式是以圖形上節點代表將分析的每一筆論文，並使用每一對論文的相關性做為建立網路上連結線的參考訊息，當兩筆論文間的相關性超過某一個預設的程度時，其對應的節點之間便有連結線。進行資訊視覺化時，可以使用合適的映射演算法盡可能讓相關性愈大的論文在圖形上映射點的距離愈接近，例如 Kamada-Kawai 演算法 (Kamada, & Kawai, 1989)。但網路圖也有以下的問題：若連結線資料的數量相當龐大，二維的圖形上將佈滿連結線，很難清楚地呈現網路的型態，且大量的資料也很難在同一畫面同時觀察詳細的節點與連結線資料以及整體的結構。另外，在利用網路圖分析研究主題時，需要依賴分析人員判讀網路圖的型態，容易產生不一致的結

果。過去的研究曾建議運用縮減連結線數目的「路徑搜尋網路」(pathfinder networking, 以下簡稱PFNet)和將網路劃分為凝聚性子群(cohesive subgroups)的「社群偵測」(community detection)等技術來改善網路圖的資訊視覺化。

將PFNet演算法應用於資訊視覺化,不僅可容易地發現網路整體的型態特性,同時也可細膩地表現出局部的連結結構(Chen, 1998),因此已經被應用於資訊科學(White, 2003)、超文件(Chen, & Carr, 1999)、電腦圖學(Chen, & Paul, 2001)、地質學與宇宙學(Chen, 2003, Chapter 6)等多個學術領域的研究主題視覺化。PFNet演算法裡使用 r 和 q 兩個參數決定產生的新網路,這些網路都保留了原先網路的「最接近的非直接路徑」(closest indirect path)(李敦仁、余民寧, 2007)。參數 r 值的範圍是1與無限大(∞)之間,是以Minkowsky距離的計算方法計算非直接路徑上的權重值。當 r 為1時,路徑的權重值為路徑上所有連結線的權重值總和;當 r 為無限大時,則將路徑的權重值設定為路徑上各連結線的權重最大值。參數 q 則用來限制路徑上的連結線數目,範圍在1到 $n-1$ 之間, n 表示網路的節點數量。由於網路圖上最長的路徑為 $n-1$,因此當 $q=n-1$ 時,表示對路徑長度沒有限制。當兩個節點間連結線的權重大於它們之間非直接路徑的權重時,此時這條連結線將被刪除。在PFNet演算法應用上,可以調整 r 和 q 來產生各種網路,一般而言, r 和 q 的值愈大,刪除的連結線愈多。應用PFNet處理的研究主題分析研究,大多是將 r 值設定為無限大,且令 $q=n-1$,將新網路上連結線數量減少到最小。具有相同主題的論文,由於彼此間相關性很高,在原先網路上對應的節點會匯集成叢集;經過PFNet處理($r=\infty, q=n-1$)後,這些節點叢集則會形成網路上的一個分支。然而PFNet處理無法自動從輸入的網路上發現重要的子群,需要分析人員仔細觀察網路上的分支情形。為讓分析人員在發現研究主題時更加便利且產生一致的結果,Chen與Carr(1999)及Chen與Paul(2001)利用Pearson相關係數(Pearson's correlation coefficients)對共被引次數進行因素分析(factor analysis),把因素分析的叢集結果呈現在網路圖上做為研判研究主題的參考資訊。然而在上述研究中,PFNet處理與因素分析是兩個各自獨立的過程,PFNet處理所產生的網路型態並沒有被妥善地運用在因素分析的過程裡。將PFNet處理產生的網路和因素分析的叢集結果一起繪製在圖形上,可能有些節點與同一集合的其他節點在網路上屬於不同分支,在圖形上映射點間的距離相當大,使得圖形變得複雜,不利於觀察與探索。

將論文與它們之間的相關程度表示成節點與連結線後,可根據社群結構(community structure)將網路劃分成凝聚性子群,網路分析將這樣問題稱為「社群偵測」。所謂社群結構是指在同一子群內節點之間彼此有密集的連結線,但在不同子群間的連結線則相當稀疏。一般在網路分析的社群偵測以模組性(modularity)來測量某種劃分方式(partition)所產生子群的社群結構品質(Girvan, & Newman, 2002; Clauset, Newman, & Moore, 2004; Newman, & Girvan, 2004;

Newman, 2006)。模組性的詳細定義與測量方式可參見上述Newman等人的論文。近年來資訊計量學研究經常將社群偵測應用於各種書目資料的分析(Waltman, van Eck, & Noyons, 2010)。例如針對ISI主題分類的引用網路進行社群偵測，將發現的子群與期刊內容的分類結果進行比較，探討ISI主題分類的品質與合適性(Zhang, Liu, Janssens, Liang, & Glänzel, 2010)；分析期刊引用網路上具有社群結構的子群，產生能夠代表科學領域的結構(Schubert, & Soós, 2010)；對於某一學科的論文根據它們之間的引用關係建立網路，然後劃分成子群，探討這個學科的研究結構(the structure of research)，並探討各子群代表的次領域規模大小、發展時間和引用影響(citation impact)等量化特性(Chen, & Redner, 2010)，發現正在興起中的研究主題(Takeda, & Kajikawa, 2009)；藉由發現作者共被引網路的社群結構，了解學科的主要研究主題(Wallace, Gingras, & Duhon, 2009; Groh, & Fuchs, 2011)；整合社群偵測和主題確認(topic identification)，藉以探討研究社群與研究主題間的彼此交織(interwoven)且共同演化(co-evolving)的互動關係(Yan, Ding, Milojević, & Sugimoto, 2012)；甚至利用社群偵測的結果子群評估相似性測量方式(similarity measures)的選擇(Colliander, & Ahlgren, 2012)。使用社群偵測方法的優點在於：(一)能夠利用資料項目的相關性構成的網路型態；(二)利用單一最佳化函數(optimization function)自動判斷子群及其數量；(三)以模組性來判斷網路上的社群偵測結果，無須受限於網路的大小與型態；(四)以網路本身既有的結構劃分成子群，無需事先對網路本身進行改變(Wallace, Gingras, & Duhon, 2009)。社群偵測演算法都是以尋找能夠產生最大模組性的網路劃分方式為基礎(Girvan, & Newman, 2002; Newman, & Girvan, 2004; Newman, 2006)，但由於尋找最佳的網路劃分方式相當困難，因此大部分演算法都是以經驗法則來求解，在資訊計量學研究裡常用的社群偵測演算法包括Newman與Girvan(2004)提出的根據連結線中介性(edge betweenness)的分裂式演算法(Takeda, & Kajikawa, 2009)、Clauset、Newman與Moore(2004)提出的聚合式方法(Groh, & Fuchs, 2011; Yan, et al., 2012)和Blondel、Guillaume、Lambiotte與Lefebvre(2008)提出的多層次聚合方法(Multi-level Aggregation Method)(Wallace, Gingras, & Duhon, 2009; Chen, & Redner, 2010; Zhang, et al., 2010; Colliander, & Ahlgren, 2012)。雖然社群偵測方法能清楚地表示整個網路的區分情形，但在有較多連結線的情況下顯得雜亂(Zhao, & Lin, 2010)，因此Groh與Fuchs(2011)在以社群偵測方法劃分網路上的子群後，再利用PFNet刪除不重要的連結線，將保留原先網路結構的新網路進行視覺化，在新網路的映射圖上標示子群。然而社群偵測演算法無法保證同一子群的節點會映射在鄰近的區域。

雖然PFNet處理和社群偵測都曾分別應用於學術領域的研究主題分析，但先前的研究少有整合這兩種技術，如Zhao與Lin(2010)比較多維尺度法(mul-

tidimensional scaling, MDS) 搭配聚合式階層叢集法 (agglomerative hierarchical clustering, AHC)、自組織映射圖 (self-organizing map, SOM)、PFNet 處理和社群偵測等四種方式在資訊視覺化結果的成效；Groh 與 Fuchs (2011) 的研究中，PFNet 處理和社群偵測兩種技術的應用過程是分開進行的，最後再一起繪製於圖形上。本研究嘗試整合 PFNet 處理和社群偵測兩項技術，先利用 PFNet 處理過濾較不重要的連結線，接著將新產生的網路輸入社群偵測演算法，尋找網路上叢集的子群，最後在呈現的網路圖上標示出主題相關的節點。整合兩種技術的優點如下：

- (一)對於 PFNet 處理後的網路而言，利用社群偵測演算法能將網路上相關的節點劃分成子群，使得研究主題確認更加容易，並可避免主觀判斷，產生一致的結果。
- (二)Groh 與 Fuchs (2011) 的研究將社群偵測和 PFNet 處理分別進行，可能產生原先網路屬於同一子群的節點，在 PFNet 網路上距離相當遠的問題。本研究則採取以 PFNet 處理後產生的新網路做為社群偵測演算法的輸入，由於經過 PFNet 處理的網路具有保留原先網路的結構特性，研究主題相近的節點可以劃分在同一子群或鄰近子群，容易利用繪製成的網路圖進行研究主題的觀察與探索。
- (三)因 PFNet 處理可大幅減少網路的連結線數量，且容易從繪製成的圖形看出節點從屬於子群以及子群彼此之間的關係，因此整合兩種方法所繪製的網路圖應較容易觀察。

除了整合 PFNet 處理和社群偵測兩種技術外，本研究還利用各子群內的詞語出現頻次訊息做為研究主題的表徵，進行研究主題標示。選用詞語出現頻次的原因乃以本研究在建立論文相關網路與 PFNet 處理時也同樣利用詞語出現頻次的統計訊息，使用相同的訊息將能夠使資訊視覺化的各個過程更緊密的整合。另外，從論文資料內出現詞語的語意也較容易讓分析人員明了節點集合所對應的研究主題。

本論文的其餘章節，除了說明本研究的方法流程與使用的工具外，並提出實際在學術領域進行研究主題分析的結果。另外，本研究也將利用「刪減相關性較小的連結線再進行網路劃分」，以及 Groh 與 Fuchs (2011) 的「分別進行網路劃分和 PFNet 處理」等兩種方法產生的結果，與本研究所建議的整合方法進行比較。

二、方法流程

本研究進行資訊視覺化的流程，如圖 1 所示，首先從論文的題名與摘要抽取關鍵詞語，運用關鍵詞語出現頻次估算論文間的相關性，利用這些資訊建立論文相關網路，以 PFNet 演算法縮減網路上連結線數量，以社群偵測演算法將

PFNet處理產生的網路劃分成子群，將PFNet處理與網路劃分的結果繪製成網路圖，並根據子群內論文的詞語訊息辨認可能的研究主題。以下說明論文相關性估算、論文相關網路建立、PFNet處理、網路劃分、網路圖繪製、研究主題標示等各處理階段。

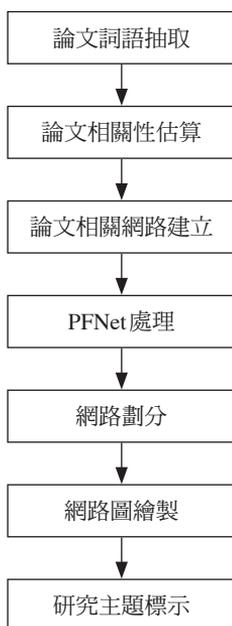


圖1 本研究進行資訊視覺化方法流程

(一)論文相關性估算

本研究根據論文的題名與摘要等文字資料中出現的關鍵詞語，估算論文間的相關性。首先統計各種關鍵詞語出現在每一筆論文文字資料內的頻次以及出現論文資料的筆數，計算詞語的出現頻率 (term frequency, tf) 及反轉文件頻率 (inverse document frequency, idf)，並將這兩種資訊整合做為可表現出論文研究主題的特徵向量。論文特徵向量上的每一個特徵值由相對應的詞語在這筆論文上的出現頻率和其反轉文件頻率的乘積產生，並經過正規化，使特徵向量間的歐幾里得距離 (Euclidean distance) 在0與1之間。將論文間的相關性定義為1減去對應的特徵向量間的歐幾里得距離。特徵向量間的距離愈小，代表這一對論文資料的詞語出現分布愈相像，也就是彼此的研究主題愈相關。

(二)論文相關網路建立

建立論文相關網路時，將每一筆論文資料指定為網路上的一個節點，並根據論文間的相關性，建立節點間的連結線。為減少接下來處理的運算量，可視情況將相關性非常小的連結線刪除。

(三) PFNet 處理

本研究以網路分析軟體 NWB Tool (NWB Team, 2006) 提供的 Fast Pathfinder Network Scaling 分析功能，對論文相關網路進行 PFNet 處理。為縮減最多的連結線，將參數 r 設為無限大。

(四) 網路劃分

許多在原先網路上彼此連結成叢集的節點，在經過 PFNet 處理後會形成新網路上的一個分支。由於同一分支上的節點，其對應的論文很有可能研究相同主題，因此本研究假定可利用適合的演算法劃分新網路上的分支，確認學術領域內的重要研究主題。

在經過 PFNet 處理後的網路上，在分支內的節點連結到其他分支的路徑都必須通過位在分支前端的連結線，如此一來在分支前端的連結線有比位於末端連結線更高的中介性。因此，本研究利用 Girvan 與 Newman (2002) 所提出根據連結線中介性的分裂式演算法來劃分網路。這個演算法每一次刪除一條當時網路上中介性最大的連結線來產生新的網路，然後重新對新網路計算連結線中介性，做為下次刪除連結線的依據，一直到網路上所有的連結線都被刪除為止。在刪除連結線時，若有某一個相連成分 (connected component) 分裂，便以當時的成分做為子群來劃分網路，並測量模組性。Girvan-Newman 演算法以分裂過程中模組性最大的網路劃分做為輸出結果。本研究利用網路分析軟體 NodeXL (Hansen, Shneiderman, & Smith, 2010) 所提供 Group by Cluster 功能中的 Girvan-Newman 方法。

(五) 網路圖繪製

將 PFNet 處理後產生的網路以及網路劃分的子群輸入網路分析軟體 Pajek (de Nooy, Mrvar, & Batagelj, 2005)，利用 Kamada-Kawai 演算法進行圖形繪製。

(六) 研究主題標示

在進行網路劃分之後，將產生的各個子群視為研究主題，並根據每個子群所屬節點對應的論文資料，標示研究主題。本研究以論文資料的詞語出現頻次做為參考，以個子群出現頻次最高的前五個詞語做為標示。

三、研究結果

本節將說明研究資料、PFNet 處理以及網路劃分和主題標示的結果。為了解研究結果的適用性，本研究將從子群包含的論文討論主題標示的結果，並以統計訊息分析子群的特性。本研究也將比較整合方法與「刪減相關性較小的連結線再進行網路劃分」，以及「分別進行網路劃分和 PFNet 處理」等三種方法產

生的視覺化呈現結果。最後，討論連結線刪減對網路劃分的影響。

(一)研究資料

本研究的研究資料為林頌堅(2012)檢索臺灣博碩士論文知識加值系統(<http://ndltd.ncl.edu.tw/>)得到的778筆資訊傳播系所相關論文資料，並以同一研究採用的詞語抽取及相關性評估方法，從論文的題名與摘要抽取關鍵詞語，運用關鍵詞語出現頻次估算論文之間的相關性。抽取出來的詞語種類共有293種，有1筆的題名與摘要中未被抽取關鍵詞語。除了沒有包含關鍵詞語的1筆論文資料以外，本研究以其餘的777筆論文資料進行研究。

建立論文相關網路時，以777個節點代表777筆論文資料，刪除兩筆論文資料間的相關性小於或等於0.1的連結線，最後共得到7168條連結線，網路密度則為 2.37×10^{-2} 。且此時網路上只有一個相連成分，也就是所有的節點之間都有至少一條路徑相連結。

(二)PFNet處理的結果

將論文相關網路經過PFNet處理刪除權重不符合三角不等式的連結線之後，保留下來的連結線並不多，產生的新網路共有786條連結線，密度為 2.61×10^{-3} 。換言之，留下的連結線約為原先網路的10.97%。將網路資料輸入Pajek，以Kamada-Kawai演算法繪製網路圖，其結果如圖2。圖形上每一個節點便對應一筆論文資料。

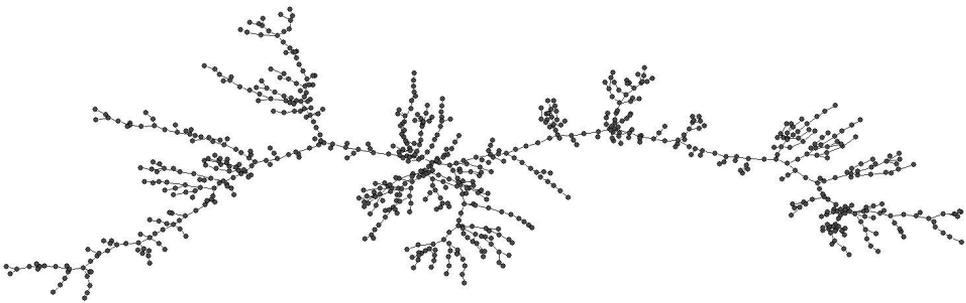


圖2 經過PFNet處理(的值設為無限大)的網路圖

(三)網路劃分與主題標示的結果

經過PFNet處理的網路輸入Girvan-Newman演算法進行網路劃分，模組性最大時共分為30個子群。子群在PFNet的分布情形，如圖3所示。為方便進行接下來的分析，為這些子群編號並附註出現頻次最高的詞語。表1則是各子群內出現頻次最高的前五個詞語。以子群C1為例，在這個子群裡，出現頻次最高的詞語是「中華電信」，出現頻次較高詞語還有「整合行銷」、「數位電視」、「雜誌」和「電信」等常見於傳播媒體研究的相關詞語。觀察表1可發現，各子

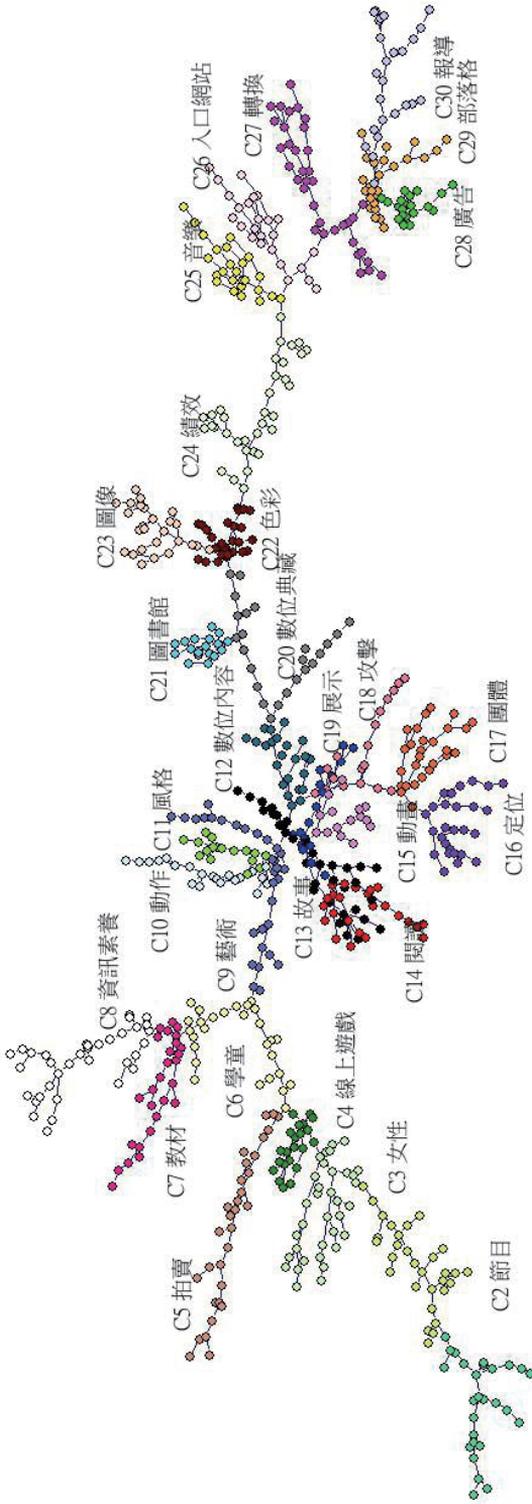


圖3 對PFNet進行網路劃分及主題標示後結果

註：圖上數值是各子群編號，文字是其出現
頻次最高詞語，例如C1為編號第一子群，
其出現頻次最高詞語為「中華電信」



群標示的詞語很多與資訊傳播學領域或其基礎領域有很高的相關性，所以表1的結果可視為資訊傳播學領域的重要研究主題。表2則是各子群的一些統計訊息。首先從各子群包含的節點數量計算各研究主題內的論文數，其次利用熵值(entropy)估計各研究主題在論文的畢業年份以及系所分散情形。從表2可發現最大的子群有36個節點，最小的則有17個節點，平均為25.90個，標準差為5.86。大致上各研究主題內的論文數分布平均。本研究所使用的論文資料，其畢業年份分布在民國88年到民國98年期間，且論文數量隨著系所以及每個系所的畢業生逐年增加而增加；大部分研究主題的論文畢業年份也是分布在民國88年到民國98年。小部分主題在畢業年份上較為分散，例如C26、C4和C19等，這些主題每年都受到資訊傳播學相關系所碩士生的重視。也有小部分主題的論文畢業年份很集中，例如C6和C11。但事實上大部分主題在論文畢業年份的分

表1 各子群內出現頻次最高前五個詞語

子群	出現頻次最高詞語				
C1	中華電信	整合行銷	數位電視	雜誌	電信
C2	節目	影音	廣播	影片	電視台
C3	女性	體驗	消費	網頁	男性
C4	線上遊戲	玩家	online game	論壇	女性
C5	拍賣	網絡	交易	信任	玩家
C6	學童	教材	國小	玩家	年級
C7	教材	e learning	數位學習	web 2	電子商務
C8	資訊素養	e learning	數位學習	技能	員工
C9	藝術	成員	研發	國小	創作
C10	動作	特徵	中文	使用性	文件
C11	風格	圖形	造形	藝術	平面
C12	數位內容	創意	學習者	評量	創作
C13	故事	兒童	實境	閱讀	體驗
C14	閱讀	出版	電影	讀者	電子書
C15	動畫	人才	遊戲產業	大專	創意
C16	定位	訊號	動態	無線區域網路	追蹤
C17	團體	家庭	電話	壓力	健康
C18	攻擊	通訊	團隊	協定	交換
C19	展示	數位媒體	專案	digital media	學系
C20	數位典藏	導覽	digital archives	博物館	典藏
C21	圖書館	館員	館藏	社區	讀者
C22	色彩	印刷	色差	記憶	檢測
C23	圖像	檢索	雜訊	複製	影像品質
C24	績效	知識管理	通路	顧客	knowledge management
C25	音樂	知覺	獨立	工業	行銷策略
C26	入口網站	手機	電子化	忠誠度	涉入
C27	轉換	標籤	視覺化	演算法	投資
C28	廣告	包裝	部落格	強度	置入
C29	部落格	搜尋	民眾	信度	搜尋引擎
C30	報導	框架	形象	中國	編碼

布都和系所及畢業生逐年增加有關。各主題的論文資料在相關系所的分散情形則可觀察到某些主題是各資訊傳播學相關系所碩士生都重視的研究方向，例如C23和C29分別有8個系所的論文資料涉及，且各系所所占比率都不高。某些主題則是某些個別系所的專長，例如C19、C22、C11、C21、C15和C6等子群內的論文資料則集中於某一個系所。

表2 各子群內的統計訊息

子群	論文數	年份分散情形	系所分散情形
C1 中華電信	27	1.95	1.78
C2 節目	30	1.80	1.52
C3 女性	35	1.91	1.67
C4 線上遊戲	22	2.13	1.86
C5 拍賣	31	1.80	1.69
C6 學童	28	1.58	1.27
C7 教材	27	1.83	1.80
C8 資訊素養	35	1.87	1.58
C9 藝術	31	1.97	1.55
C10 動作	21	2.03	1.87
C11 風格	17	1.56	1.09
C12 數位內容	36	2.03	1.69
C13 故事	17	1.89	1.33
C14 閱讀	24	1.95	1.66
C15 動畫	15	1.83	1.66
C16 定位	27	1.86	1.26
C17 團體	27	1.98	1.85
C18 攻擊	18	1.94	1.40
C19 展示	23	2.11	0.90
C20 數位典藏	24	2.08	1.64
C21 圖書館	19	1.89	1.24
C22 色彩	27	1.91	1.05
C23 圖像	26	1.74	1.90
C24 績效	34	1.94	1.86
C25 音樂	25	1.96	1.52
C26 入口網站	25	2.17	1.64
C27 轉換	36	1.94	1.79
C28 廣告	20	1.75	1.82
C29 部落格	25	1.84	1.90
C30 報導	25	1.84	1.60

註：年份與系所分散情形欄位，粗體數字表示較集中（年份分散情形 < 1.60，系所分散情形 < 1.30）；斜體數字則表示較分散（年份分散情形 > 2.00，系所分散情形 > 1.80）

以下針對每個子群的研究主題利用出現頻次較高的詞語以及論文資料進行分析：C1的論文主要探討傳統傳播產業在推動與經營新式服務業務上的議題，所涉及的傳播產業包含電信業、電視媒體、雜誌業等，探討的各種新式服務業務則很明顯的都是因資通訊科技的發達而興起，包含數位電視、MOD、互動

雜誌、整合行銷傳播等。C2的相關論文為有關各種影音媒體平台的建置以及電視與廣播節目等內容製播的研究。C3所包含論文資料的研究主題與行銷策略、消費者生活型態、品味和風格等有關，許多論文甚至特別針對女性消費者的行為特質與其應用在體驗行銷的研究。C4的論文關注的研究主題則大多是線上遊戲，包括線上遊戲的玩家特質以及線上遊戲產業的經營與行銷策略等。線上遊戲一直都是資訊傳播學相關系所研究生喜愛的研究主題，從民國92年起，每年均有2到3篇學位論文產出。

C5的論文著重的主題比較多元：有些與網路拍賣等線上交易的商業模式有關，且特別著重於線上交易過程中人際間的信任關係；有些探討Facebook等新興的社會媒體如何促進人際關係的增長以及信任；有些則是運用社會網路(social network)的概念探討現在社會中的人際關係如何影響人們的行為。這個集合的論文主題雖比較多元，但足可發現資訊傳播學領域的研究相當關注於網路上的社會關係，同時也嘗試援用社會學等其他領域的成果來豐富其本身的研究基礎與內涵。C6、C7和C8在圖3的網路圖上彼此非常接近，而事實上這三個節點集合所對應的論文大多都有關於數位學習的研究。C6包含了許多運用數位遊戲、Web 2.0和多媒體技術等資訊科技進行國小教材設計的研究；C7的論文研究主題大多與數位學習的教材設計與平台設計有關；C8的論文則包含數位學習的成效評估、職能教育與訓練和資訊素養等多種人類學習的研究主題，也有些研究將這些主題整合，探討資訊素養對工作人員在從業技能表現的關連。C6的28筆論文資料分別有4筆和13筆的畢業年份為民國96年和97年，且主要產出者為立德大學資訊傳播學系，該系所在民國96年和97年共有13筆相關的論文。

從圖3的圖形上，C10和C11看起來像是C9旁出的分支。C10相關的論文主題為網站等互動設計的偏好(preference)與使用性(usability)研究。C11所涉及的主題有許多將數學模式運用藝術設計的研究，並特別著重於造型風格的分析，相關的17筆論文資料主要集中於民國96到98年，分別有4筆、5筆和5筆論文資料，明顯可見這個主題在這幾年相當受到重視。而C9的節點分為兩部分，一部分節點較靠近C11，這部分研究與C11的論文主題很接近，都是將數學模式運用藝術創作的研究，另一部分接近C8的節點則是探討有關於資通訊技術應用於自然科、語文學習或人文藝術的教學設計。C11的相關論文大多來自於元智大學資訊傳播學系，該系所的論文佔C11的64.71%，由此可見該系所在這領域的努力與成果。

C12的研究主題與數位內容及創意管理有關，數位內容可再分為包含數位內容的創作及數位版權管理系統，而數位內容與創意管理兩方面的交集則為文化創意的加值運用。C13所包含論文為將體感互動裝置及擴增實境等資訊科技

運用於數位敘事及情境設計的相關研究。C14的論文主題大多與出版與閱讀有關，特別是由於電子報及電子書等數位文本的載具與內容技術快速進展之際。C15則以媒體產業的內容產製、核心資源與人力資源培育等主題的相關研究為主，其中資訊傳播學領域碩士研究生最關注的媒體產業是電腦動畫及遊戲軟體產業，這兩個產業是由傳播媒體和藝術設計引入資訊技術進行整合後所產生。

C16、C17、C18彼此接近而離其他子群較遠，這些子群包含較多以通訊技術開發、應用和電信產業經營為研究主題的論文。C16的論文大多有關於網路技術的研發與應用研究，特別是有關於定位系統的追蹤技術與應用服務。這些論文很多來自於銘傳大學資訊傳播工程學系，佔C16的62.96%。C17對應的研究主題較多，包括網路電話的技術發展與產業經營、數位家庭應用到日常生活網路資訊使用行為等都是這個子群對應論文探討的主題。C18的論文研究與網路安全技術的開發與應用相關。

C19的研究主題有數位媒體設計以及網站、博物館、櫥窗等各種展示場所的導覽設計，這個子群內有73.91%的論文來自於元智大學資訊傳播學系。C20關注的研究主題為數位典藏，研究範疇包含了新聞報導、博物館文物及影音資料等各種形式典藏品的典藏技術與應用，此外C20也有部分論文與博物館的網站與行動裝置等數位導覽系統有關。C21大部分論文的研究主題為圖書館的經營與服務；在經營方面的相關研究有館員的人才培育、館藏發展、委外編目、圖書館行銷、收費機制；圖書館服務則有圖書館導覽、讀者對圖書館空間利用等服務項目以及服務品質的研究。此外也有無線射頻辨識(RFID)與無線寬頻網路等資訊科技在圖書館的利用。前身為圖書資訊學系的世新大學資訊傳播學系是這個子群的主要貢獻者，共有57.89%的論文來自該系所。

C22與視覺傳播技術及印刷出版產業有關；視覺傳播技術包括使用者對色彩的知覺、認知與喜好研究、色差校正、色彩影像頻譜等研究；印刷產業的相關研究則有經營環境與策略、電子化作業、服務品質、印刷技術等主題。這個子群有70.37%的論文來自於先前名為印刷傳播系的中國文化大學資訊傳播學系。C23的研究主題主要是圖像處理與檢索等數位典藏的重要技術以及圖像對人類認知與喜好的影響。C24的相關研究主題為圖書出版業、電視台、電視購物、平版印刷業等傳播產業的經營策略與績效評估，特別是產業導入各種知識管理方法與技術的評估。

C25則與音樂產業的生產行銷相關，特別是音樂數位化後，合法與非法的音樂都能迅速方便地透過網際網路傳播，整個音樂產業的版權、生產與銷售管道都受到非常大的衝擊，也提供了獨立音樂人及數位音樂服務平台創新的發展機會，是相當富有挑戰與開創性的研究，因此吸引許多資訊傳播學的碩士生以此為論文的研究主題。C26的論文有許多關於行動服務與網站的使用態度與行

為研究；行動服務涵蓋行動上網、手機電視、行動電玩等無線通訊應用服務；網站部分則涉及各種不同類型的網站，包括入口網站、拍賣網站、品牌網站及社群網站等。

C27的論文為各種網路通訊、影像處理以及資訊視覺化等演算法。演算法是針對運用電腦解決某些問題而發展出的一套系統化而有一定進行程序的方法。這個子群的許多論文以此為研究主題，說明了資訊傳播這個領域著重於運用資通訊科技專業提供的邏輯思維與整合技術(梁朝雲, 2011)，致力於發展解決各種傳播、資料處理與視覺呈現上發生的技術問題。C28的主要研究主題為廣告，特別是由資通訊科技發展所產生的各種新型態的廣告模式，例如於電子遊戲、網站、線上影音、部落格、數位電視等各種載體傳播的內容上置入廣告，都是許多資訊傳播學領域碩士論文研究關注的問題。C29的論文大都是針對部落格相關問題的研究，包括部落格的使用行為分析、部落格內容的可信度與傳播成效、部落格行銷的影響因素與應用。C30包含許多利用內容分析新聞報導的相關論文，且有部分論文是針對媒體傳播所產生或再現的個人及群體形象的研究。

(四)與其他方式的比較

首先比較「刪減相關性較小連結線再進行網路劃分」的方式。本研究建議的方式在進行PFNet處理後共保留786條連結線，所以依據相關性大小刪除產生的新網路上也將保留786條連結線。對新網路進行劃分的處理方法，同樣採用網路分析軟體NodeXL所提供Group by Cluster功能中的Girvan-Newman方法。

僅依據相關性大小刪除連結線的方式會使新網路包含非常多規模很小的成分，所有成分共有379個，不超過5個節點的成分便有355個，節點數量佔全體的69.75%。然而較大成分不足代表全體的研究，可能造成研究主題分析的偏誤。此外，過多的成分也造成資訊視覺化的不便，因此利用Kamada-Kawai演算法繪製網路圖時只選用5個節點以上成分，且分開繪製各成分，所產生網路圖如圖4所示。觀察圖形可發現，利用Girvan-Newman演算法進行網路劃分的結果，僅有最上方左邊第二個成分產生兩個子群，其餘成分都自成一個子群。

接下來，比較「分別進行網路劃分和PFNet處理」的方式。利用Girvan-Newman演算法對網路進行劃分後，總共產生45個子群。圖5說明這種方式和本研究建議的整合方式所產生子群的節點數量分布比率。很明顯可發現，「分別進行網路劃分和PFNet處理」的方式(Independent)所產生子群上面的節點數量差異非常大，共有6個子群包含40個以上的節點，但節點數量小於10的子群也佔非常大比率，反觀整合方式(Integrated)所產生的子群，其節點數量分布相當集中。

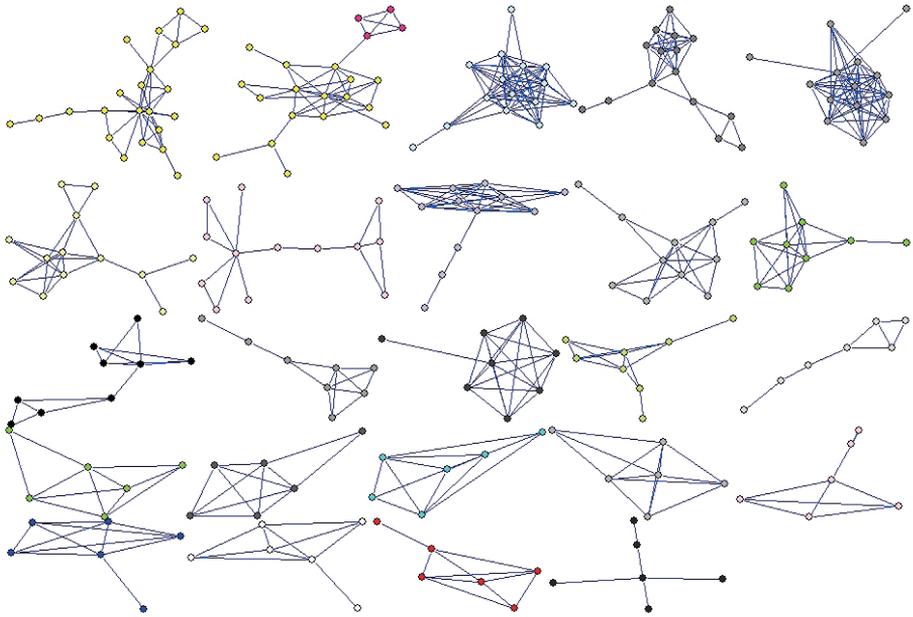


圖4 「刪減相關性較小連結線再進行社群偵測」方式所產生網路圖

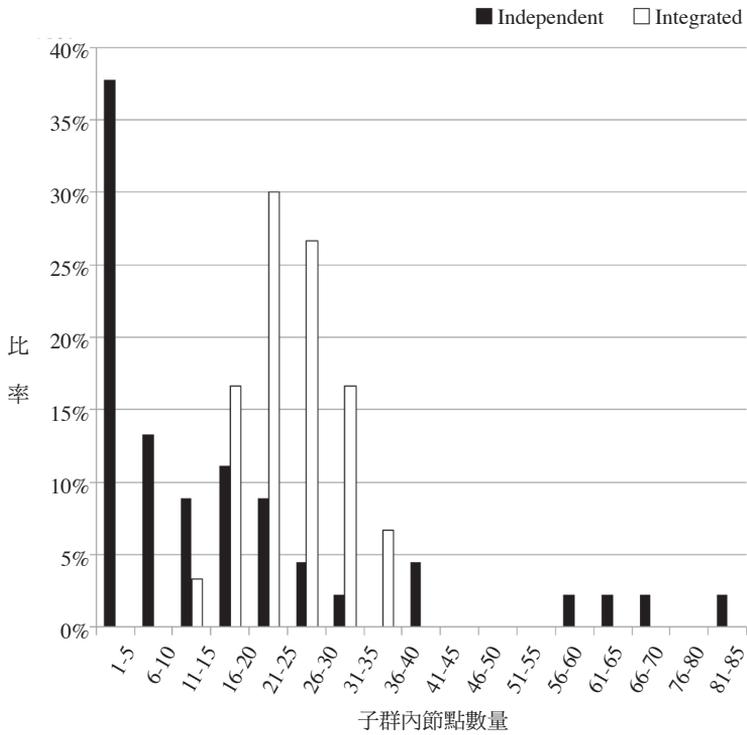


圖5 「分別進行網路劃分和PFNet處理」和「整合PFNet處理和網路劃分」方式所產生子群節點數量分布比率

圖6是將Girvan-Newman演算法輸出的45個子群對應到PFNet處理產生的網路上所繪製成的網路圖，每種顏色代表一個子群。觀察圖6可發現，雖鄰近的節點屬於同一個子群的情形很常見，但也有不少相鄰的一組節點彼此屬於不同子群。這種情形經常出現在網路分支的末端，例如圖6的右下方、中央下方和左下方都可明顯地發現。造成這種情形的原因是否由於位於PFNet處理網路分支末端的節點較難歸類於特定研究主題，因而分開處理的PFNet和網路劃分會產生不一致的結果？由於本研究並非針對「分別進行網路劃分和PFNet處理」的方式提出改善做法，未來的研究可進一步探討此一現象並加以解決。

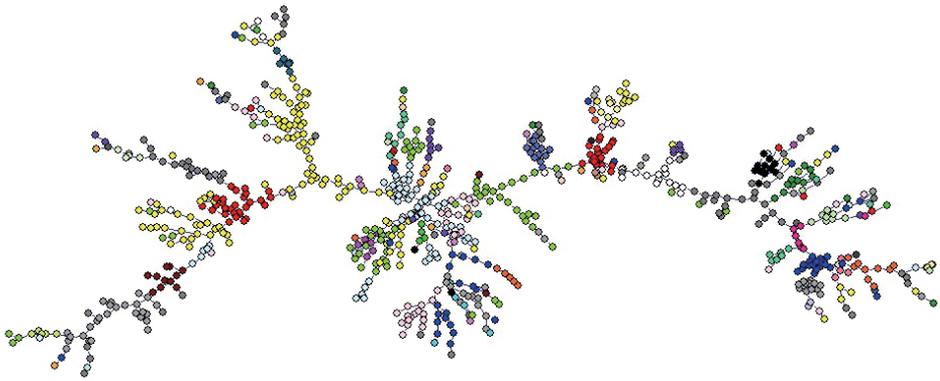


圖6 「分別進行網路劃分和PFNet處理」方式產生的網路圖

以下從圖3、圖4和圖6討論本研究的整合方法與「刪減相關性較小連結線再進行網路劃分」以及「分別進行網路劃分和PFNet處理」等三種方法產生的視覺呈現結果。

1. 「刪減相關性較小連結線再進行網路劃分」的方式，在刪減連結線時並沒有保留網路的型態，產生很多規模相當小的成分。在確認研究主題時，若忽略這些成分，容易產生偏差。將刪減後的網路呈現為圖形時，也無法將數量龐大但彼此不相連成分間的關係表現出來。

2. 「分別進行網路劃分和PFNet處理」的方式利用Girvan-Newman演算法對未經過刪減的網路進行劃分，產生結果的子群數量較多且各子群內的節點數量差異很大，且在網路圖上，彼此相鄰的節點大多屬同一子群，但分支末端則有許多相鄰節點分屬不同子群。就資訊視覺化的目的而言，若子群數量較多且屬同一子群的節點不在距離較近的區域內，一般認知能力的人類很難一一觀察到不同的子群，進而發現各研究主題的分布與關係。

3. 本研究建議先經由PFNet演算法刪除較不重要的連結線，再利用Girvan-Newman演算法對新網路進行劃分，結果產生的各子群其節點數量相當接近，且同一子群內所有節點在網路圖上彼此相鄰近，產生的網路圖較容易觀察。

(五)子群的品質

PFNet演算法是一種維度縮減 (dimensionality reduction) 的方法 (Börner, Chen, & Boyack, 2003)。利用PFNet演算法刪減連結線，事實上是希望藉由捨棄較不重要的資訊，產生容易觀察的網路圖。經過PFNet處理，難免使網路劃分的結果受到影響。為了解連結線刪減對網路劃分造成的影響，可將經過PFNet處理網路所產生的30個子群與原先網路產生的45個子群進行比較。為了方便解釋，以下以C代表本研究對於經過PFNet處理的網路進行劃分而產生的某一個子群，而將對原先網路直接進行劃分而產生的某一個子群稱為C'。以下利用熵值測量C上的每個節點分散在C'的情形。若C的熵值較大，代表它包含的節點

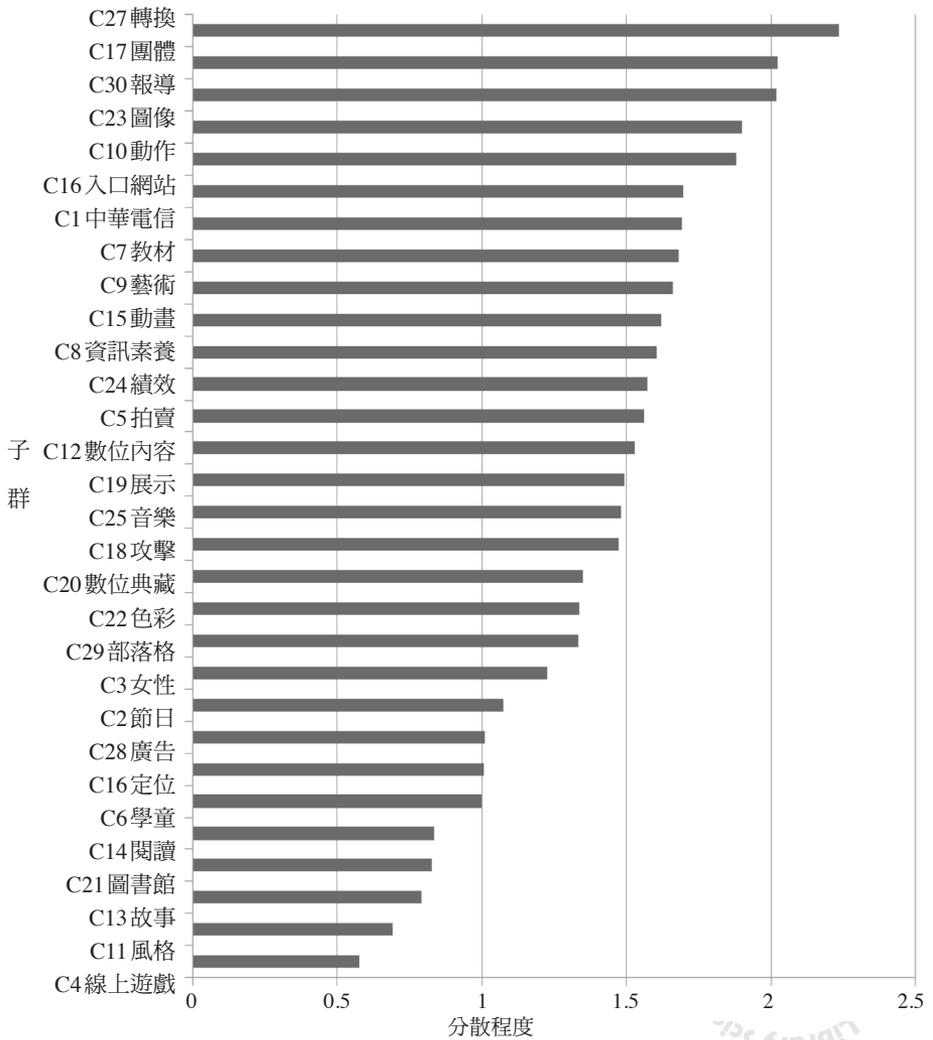


圖7 本研究30個子群在「分別進行網路劃分和PFNet處理」方式所產生45個子群分散情形

分布的C'較多，且節點的分布數量較平均；反之熵值較小，則集中於較少的C'上。各子群依據分散情形大小排序的結果，如圖7所示。

最分散的子群如C27、C17和C30，最集中的子群則為C4、C11和C13。參考表1「各子群內出現頻次最高前五個詞語」，子群的分散程度似乎與它們的最高頻次詞語是否多屬於一般生活用語，或資訊傳播學或其基礎領域的專殊術語有密切關係。以分散情形較大的子群C17為例，「團體」、「家庭」、「電話」、「壓力」和「健康」是這個子群出現頻次最高的五個詞語，這些詞語大多很難看出與資訊傳播學領域之間有明顯的關係，甚至不容易發現五個詞語彼此間的相關性。所以利用這些詞語做為評估論文間相關性的特徵，將會高估內容出現這些詞語但不同研究主題論文之間的相關性。在查看子群C17的論文資料後，也發現這些論文所涵蓋的面向相當廣泛。反之，子群C4最高出現頻次的五個詞語為「線上遊戲」、「玩家」、「online game」、「論壇」，和「女性」，除了「女性」之外的詞語彼此間的相關程度很高，所以出現這些詞語的論文彼此間的相關程度也很高，不論PFNet處理或Girvan-Newman演算法的網路劃分都很容易將這些論文叢集成同一子群，C11和C13也有類似情況，因此這些子群較為集中。

四、討論與結論

利用資訊視覺化方法進行學術領域的研究主題分析，能夠提供分析人員利用圖形有效地觀察與思考，因此有愈來愈多的研究運用各種圖形以及演算法，嘗試找出大量論文資料的最佳視覺呈現方式。本研究建議將PFNet處理和Girvan-Newman演算法整合，產生能夠表現重要研究主題的網路圖。本研究並建議利用最高頻次詞語做為研究主題的標示，讓結果的分析與詮釋更加容易。從研究結果發現，在子群內的節點其對應的論文大多具有相同或類似的研究主題，網路圖上位置相近的子群彼此間的研究主題也很接近。比較其他的視覺化方式，本研究的幾項特色如下：(一)可利用詞語分布相似程度估計論文彼此間的相關性，並根據這項資訊，PFNet方法可將研究主題相同或相似的論文大致對應到網路圖的節點分支上。(二)利用Girvan-Newman演算法進行網路劃分，能夠根據網路的結構特性，將位於同一分支上的節點視為一個子群，結合前項特性可發現資訊傳播學的重要研究主題。(三)利用論文資料中詞語出現的統計資訊，找出子群內最高出現頻次的詞語，這些詞語大多和資訊傳播學以及其基礎領域的問題、方法、理論和技術相關，可做為研究主題的標示。

雖然從表1所列詞語可容易地推敲出大多數子群對應的主題，但有些子群的最高頻次詞語多屬於一般生活用語，很難一眼看出和資訊傳播學或其基礎領域間的關連，甚至也不易發現五個詞語彼此間的相關性，例如C17。利用這些詞語特徵評估論文間的相關性，不易區別不同研究主題的論文。如此一來，便

使得這些少數子群較難判讀。

本研究的網路圖具體展現出資訊傳播學的範疇與內涵，可以和梁朝雲(2011)以及林頌堅(2012)的研究結果相互參考與驗證。梁朝雲(2011)將資訊傳播定義為「應用資訊科技與數位媒體，結合創意設計以產生多元型態的訊息文本，經編製、典藏與加值等歷程，透過傳播管道以提供數位內容與多媒體資訊服務給受眾，以達成有效溝通、資訊交流、知識共享，與文明提升的目標。」本研究的網路圖呈現出資訊傳播學在傳播媒體研究、圖書資訊學、資訊與通訊科技、藝術設計等各基礎領域著重的研究主題，並特別反應了資訊與通訊科技的運用對於傳播媒體與藝術設計在產業經營、內容創作、應用服務與使用體驗的創新與衝擊。本研究結果更指出位於網路圖核心地位並整合多個不同基礎領域研究的數位學習和數位內容是資訊傳播學領域碩士論文研究的重要研究方向，許多的論文研究主題都涉及數位學習和數位內容的技術開發與內容產製。此外，本研究結果也驗證了Lin(1997)、Chen(1998)和Zhao與Lin(2010)對於各種圖形與演算法在資訊視覺化上應用的看法，本研究產生的網路圖與林頌堅(2012)所繪製的自組織映射圖兩者都可呈現出資訊傳播學領域的重要研究主題，並呈現出主題間的關聯。然而由於自組織映射圖方法所產生的結果為一個表現資料型態的聯想網路(Lin, 1997)，訓練初期輸入的每一組詞語特徵資料都會影響到整體圖形的調整，隨著訓練循環的增加而縮小影響的區域範圍與調整幅度，使得結果的圖形上各研究主題會依據本身相關的基礎領域映射在相鄰近的區域，在不同基礎領域間交錯的區域則是由這些領域相互激盪影響後產生的新研究主題的映射區域，因此自組織映射圖善於呈現出領域內所有主題的整體架構。本研究所採用的網路圖則以節點間的連結線與權重表示資料間的相關性，並在經過PFNet處理後進一步刪除較不重要的連結線，並在以圖形呈現時盡量保留距離資訊(Lin, 1997)，細膩地表現出局部的連結結構(Chen, 1998)以及論文和主題間的從屬關係(membership)。因此本研究運用這個特點利用網路圖的型態進行網路劃分，將產生的子群對應到重要的研究主題。同時也可利用論文在網路圖上的映射位置來判斷它的研究主題，映射於網路圖分支前端的論文通常具有較多元的研究主題。兩種圖形所提供的資訊可互補，對學術領域的研究主題分析提供宏觀與微觀的不同觀點。

未來除可將本研究所提出的研究主題視覺化方法應用到更多學術領域外，也將分析不同發表期間的論文在網路圖上的映射分布，探討學術領域的發展趨勢，或將論文映射分布應用在研究單位發表的論文中，探討各研究單位著重的研究主題。此外，本研究以詞語分布相似程度測量論文資料間的相關性，也可嘗試以論文資料的共被引情形做為相關性評估的參考。最後，選擇能夠區別不同主題的詞語(van Eck, Waltman, Noyons, & Buter, 2010)做為評估論文的相關性，也是未來發展的重要課題。

參考文獻

- 李敦仁、余民寧 (2007)。學習表現的知識結構評量研究：以「教育統計學」學科知識為例。《教育研究與發展期刊》，3(4)，113-148。
- 林頌堅 (2012)。資訊傳播學領域的研究主題發展分析。《資訊傳播研究》，2(2)，1-19。
- 梁朝雲 (2011，3月)。尋找「資訊傳播」定義、發現「資訊傳播」研究。在淡江大學資訊與圖書館學系主辦，《教育資料與圖書館學》40週年國際學術研討會，新北市。
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179-255.
- Börner, K., Sanyal, S., & Vespignani, A. (2007). Network science. *Annual Review of Information Science and Technology*, 41(1), 537-607.
- Chen, C. (1998). Bridging the gap: The use of Pathfinder networks in visual navigation. *Journal of Visual Languages and Computing*, 9(3), 267-286.
- Chen, C. (2003). *Mapping scientific frontiers: The quest for knowledge visualization*. London, UK: Springer-Verlag.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of American Society for Information Science and Technology*, 57(3), 359-377.
- Chen, C., & Carr, L. (1999). Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998). In J. Westbomke, U. K. Wiil, J. J. Leggett, K. Tochtermann, & J. M. Haake, (Eds.), *Proceedings of the tenth ACM Conference on Hypertext and hypermedia: Returning to our diverse roots: Returning to our diverse roots(HYPertext '99)* (pp. 51-60). New York, NY: ACM.
- Chen, C., & Morris, S. (2003). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on* (pp. 67-74). Seattle, WA: IEEE.
- Chen, C., & Paul, R. J. (2001). Visualizing a knowledge domain's intellectual structure. *Computer*, 34, 65-71.
- Chen, P., & Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetrics*, 4(3), 278-290.
- Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6), 066111.
- Colliander, C., & Ahlgren, P. (2012). Experimental comparison of first and second-order similarities in a scientometric context. *Scientometrics*, 90(2), 675-685.
- de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek* (Structural Analysis in the Social Sciences, vol. 27). New York, NY: Cambridge University Press.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of*

- America*, 99(12), 7821-7826.
- Groh, G., & Fuchs, C. (2011). Multi-modal social networks for modeling scientific fields. *Scientometrics*, 89(2), 569-590.
- Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Amsterdam, Netherlands: Morgan Kaufmann.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1), 7-15.
- Lin, X. (1997). Map displays for information retrieval. *Journal of the American Society for Information Science*, 48, 40-54.
- Morris, S. A. & Van der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology*, 42(1), 213-295.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577-8582.
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113.
- NWB Team. (2006). *Network Workbench Tool*. Indiana University, Northeastern University, and University of Michigan, Retrieved on November 20, 2012 from <http://nwb.slis.indiana.edu>.
- Schubert, A., & Soós, S. (2010). Mapping of science journals based on h-similarity. *Scientometrics*, 83(2), 589-600.
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595-610.
- Takeda, Y., & Kajikawa, Y. (2009). Optics: A bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, 78(3), 543-558.
- van Eck, N. J., Waltman, L., Noyons, Ed C. M., & Buter, R. K. (2010). Automatic term identification for bibliometric mapping. *Scientometrics*, 82(3), 581-596.
- Wallace, M. L., Gingras, Y., & Duhon, R. (2009). A new approach for detecting scientific specialties from raw cocitation networks. *Journal of the American Society for Information Science and Technology*, 60(2), 240-246.
- Waltman, L., van Eck, N. J., & Noyons, Ed C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629-635.
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of American Society for Information Science and Technology*, 54(3), 423-434.
- Yan, E., Ding, Y., Milojević, S., & Sugimoto, C. R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. *Journal of Informetrics*, 6(1), 140-153.
- Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on ISI category classification. *Journal of Informetrics*, 4(2), 185-193.
- Zhao, H., & Lin, X. (2010). A comparison of mapping algorithms for author co-citation data analysis. *Proceedings of American Society for Information Science and Technology*, 47(1), 1-3.

附 錄

各節點集合對應之論文舉例

集合編號	代表論文題名	出版年度
1	台灣行動上網成功關鍵性因素之探討—以日本NTT DoCoMo i-mode成功經驗來看行動上網業	89
	以價值網觀點探討數位電視產業之競爭優勢—無線、有線、電信網路及衛星四類平台之比較	91
	數位時代下的創新科技管理—探討中華電信MOD經營策略與產業分析	92
2	影音序列組合對於敘事的影響	91
	台灣本土幼教電視頻道節目內容及時段編排對幼兒收視與模仿行為之研究	92
	電視新聞鏡面設計對於新聞產製流程的影響以及現況分析之初探	92
3	兩岸女性網站行銷策略之研究	90
	台妹的消費實踐與自我認同	96
	女性美容資訊節目對女性閱聽人自我建構與展現之探索性研究—以「女人我最大」電視節目為例	97
4	線上遊戲作為網路行銷工具之初探--以大安銀行e-guess猜謎網為例	89
	線上遊戲玩家心智模式及其行銷應用之研究	92
	大型多人線上角色扮演遊戲中個人化服務設計之探討	92
5	消費者對網際網路之信任與線上購物經驗關聯性研究	93
	警察日常工作資訊網絡之研究—以台北縣政府警察局新店分局為例	95
	社交網站人際關係維繫之研究：以Facebook為例	98
6	遊戲式數位教材於電腦輔助教學之應用—以國小五年級因數單元為例	96
	應用Web2.0概念及Moodle平台於國小社會科家鄉單元之教學—以彰化二林地區為例	96
	互動式數位學習網站在國小自然科花的單元學習成效之研究	97
7	應用動態影音內容索引技術建構數位學習教材之研究	92
	符合建構式教學目標之數位學習教材設計之研究—以溥心畬書畫為例	95
	互動多媒體數位學習教材之研製與易用性研究—以九年一貫小學四年級社會科為例	97
8	因應報社生產自動化探討印務部門技術人員專業職能需求分析之研究	92
	企業員工數位學習之學習風格與學習績效及其相關因素之研究：以資訊工業策進會為例	95
	不動產經紀營業人員資訊素養與仲介能力之研究--以台北(縣)市信義房屋為例	97
9	位置感知技術應用於華語文學習之研究	94
	由歐普藝術作品探討形狀文法與數學模型應用於平面造形衍生設計之研究	94
	碎形在小學藝術與人文教材設計之研究	96
10	網站之文化使用性研究—以華文入口網站為例	91
	視覺刺激訊息對於偏好態度之影響研究—以瞳孔測量實驗為例	93
	行動電話圖片檢索服務之設計與使用性研究	96
11	應用形狀文法與動態規劃法於山水畫結構造形之分析—以夏珪與馬遠為例	95
	連環漫畫視覺造形與風格之分析	98
	碎形圖形生成之研究	98
12	建構適用於知識表達之創意階段的知識可視化模式	96
	具安全與效率性的數位版權管理系統之研究	97
	TRIZ創新原則應用於文化商品之創作設計	98
13	大學生之電子書理想出版形式研究	91
	不同電子文本類型之數位閱讀理解比較研究	98
	電子童書文本結構與閱讀效果之研究	98

14	台灣動畫產業核心資源之研究 遊戲產業創意企劃人員專業能力及人才培育研究 台灣遊戲軟體產業經營策略分析與建議—以智冠科技為例	91 91 93
15	無線區域網路之室內定位機制研究 無線區域網路之位置感知系統研究 以定位服務為基礎的行動電子商務系統研究	93 94 95
16	青少年網路使用行為與輔導方案之研究—以新竹市為例 無線區域網路電話發展環境之評析 建構一個運用 RFID 於居家照護與急診室智慧型醫療系統之研究	88 94 97
17	一個可抵擋阻斷服務攻擊之金鑰交換協定設計 在無線環境下的遠端使用者認證機制之研究 分散式網路安全分析及偵測系統之研究	93 94 96
18	多媒體之情境設計探討—以「未來一行的憧憬」創作為例 互動玩具設計與開發—以體感互動裝置增進兒童多元智能之玩具為例 運用擴增實境於多人互動情境之研究—以遊戲式學習為例	90 93 94
19	應用多模式人機介面於互動式展示之探索與應用 互動式媒體之應用表達—以櫥窗資訊展示設計為例 以情境式學習探討互動式展示空間設計應用於地方特色館—以桃園縣中國家具博物館為例	92 93 93
20	歷史性報紙新聞數位典藏商務平台之研究—以北平世界日報為例 數位博物館視覺企畫與設計之實作研究—以故宮博物院「殿本主題網站」為例 臺灣電視公司影音數位典藏之研究	93 96 98
21	大學圖書館圖書編目委外作業書目品質管控及績效評估調查 民眾為什麼不使用公共圖書館：以板橋地區居民為例 無線射頻辨識在圖書館期刊管理系統之研究	92 94 97
22	個人日常生活與色彩之喜好關聯性研究 應用差異視覺函數改進色差辨識效能之研究 書刊印刷業經營環境與策略之研究	92 92 92
23	以檢索記錄分析網路圖像檢索需求之研究 利用最佳區塊配對的高容量影像隱藏方法 調整影像銳利度對視覺偏好影響之研究	92 93 94
24	以平衡計分卡法評估媒體企業數位化績效指標之可行性研究 台灣地區圖書出版產業行銷通路策略與績效之研究 知識管理運用於有線電視產業組織績效之研究	93 93 97
25	當代台灣音樂工業產銷結構分析 當代台灣嘻哈音樂錄影帶之視覺意涵研究—以西元 2000-2004 年為例 台灣地區線上音樂會員使用狀況與業者行銷策略研究	88 93 95
26	線上競標者的涉入感與其對拍賣網站的忠誠度研究 行動加值服務行銷傳播策略與顧客滿意度之研究—以行動入口網站為例 從科技接受模式探討 3G 手機行動上網之使用態度與意願研究	91 92 98
27	群播演算法在點對點網路的研究 在次像素精準度下的邊緣偵測演算法及其應用 台灣股市資訊視覺化介面之設計研究	92 93 96
28	利用投影不變量作影片虛擬廣告插入之研究 數位電視互動廣告發展現況、趨勢與應用策略：以英國、美國個案為例 網路廣告的版面配置對視覺焦點、閱讀動線的影響之研究	92 93 94
29	部落格地圖：部落格上的社交相似性 台灣立法委員部落格之政治資訊行銷研究 報紙影劇新聞版面引用部落格資訊為新聞來源之現象分析	94 94 95
30	政黨輪替後兩岸政治新聞報導之框架研究—以中國時報、人民日報為例 媒體再現學校形象 SARS 新聞報導內容分析之研究-以自由時報與中國時報為例	93 95 96



A Study of Research Topic Visualization Using Network Structure Analysis

Sung-Chien Lin

Assistant Professor
Department of Information and Communications
Shih Hsin University
Taipei, Taiwan
Email: scl@cc.shu.edu.tw

Abstract

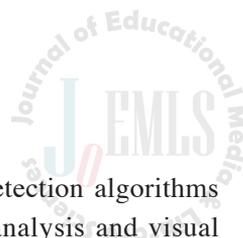
The Pathfinder network (PFNet) method and the community detection algorithms both are methods which have been widely applied to visual presentation and analysis of research topics. The PFNet method can delete a large amount of insignificant links in networks but also retains the structural characteristics of the original networks, while the community detection algorithms are able to partition networks into a set of cohesive subgroups. However, each of the methods has its deficiencies. The PFNet method cannot automatically find out critical subgroups in input networks and the community detection algorithms do not guarantee nodes in the same subgroup able to be mapped in neighboring area. The integration of these two methods provides a way to alleviate the above problems: The output network from the PFNet method is partitioned using community detection algorithm. In addition, this study also suggests that the use of high frequency terms within papers in subgroups as the labels of research topics to make the analysis and interpretation of results easier. This study takes the field of Information Communication as an analytic case to study the application of the integrated methods, and the data of master theses of the related graduated schools are collected to be used for the analysis. The results show that it is effective to integrate the PFNet method and the community detection algorithms to discover subgroups representing important research topics from the network which is constructed upon the relations between papers in the examined field. The terms with high occurring frequency in subgroups are very relevant to the problems, method, theorems and technologies in the field of Information Communications and its fundamental disciplines, and therefore, they are suitable as the representatives of research topics.

Keywords: Analysis of research topics; Information visualization; Path-finder Networks; Community detection algorithms

SUMMARY

Introduction

Pathfinder network (PFNet) scaling and community detection algorithms are both techniques that have been widely applied to the analysis and visual presentation of research topics (Lin, 1997; Chen, 1998; Zhao, & Lin, 2010). To



visualize important research topics in an academic field, representative documents in the examined field are collected and analyzed to establish a network, in which nodes correspond to the documents and edges between nodes are determined by the relevance scores between the documents. PFNet scaling is able to delete a large amount of the insignificant links, while retaining the structural characteristics of the original network. Community detection algorithms are able to partition the network into a set of cohesive subgroups. However, each technique also has its deficiencies. PFNet scaling cannot automatically determine the critical subgroups in input networks and community detection algorithms do not guarantee that nodes in the same subgroup can be mapped in neighboring areas. Groh and Fuchs (2011) proposed a method to combine these two network visualization techniques by performing them separately, but this method did not alleviate the above problems. Instead of using the two techniques independently, this study proposed integration of the two: The output network from PFNet scaling is used as the input for the community detection algorithm to partition nodes in the network into subgroups. Thus, subgroups of nodes in the resulting network represent important research topics in the examined field and the structure of the network is able to express relationships between the research topics. Furthermore, this study also suggests using high frequency terms within papers in subgroups as the labels of research topics, to make the analysis and interpretation of results easier.

Methods

Figure 1 shows the process proposed in this study and the techniques used in each step.

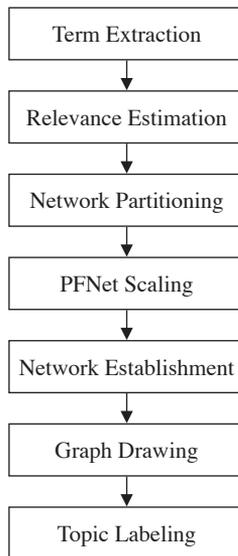


Fig. 1 The Proposed Process for Research topic Visualization

1. Term Extraction: Terms are first extracted from the titles and abstracts of examined documents as features to represent the documents. The method of term extraction used in the study is that used by Lin (2012).

2. Relevance Estimation: A feature vector is assigned to each of the examined documents, and the values of elements in such a feature vector are calculated based on the frequency counts of terms appearing in the corresponding document and the inverse document frequencies (idf) of those terms. The relevance score between two documents is defined as the value: 1 minus the Euclidean distance between their corresponding feature vectors.

3. Network Establishment: This step establishes a network for following processing steps. Each of the documents corresponds to a node in the network. The edge between a pair of nodes is determined by the relevance score of their corresponding documents. It is beneficial to delete edges with very small relevance scores to reduce computational resources needed in the next steps.

4. PFNet Scaling: The method of PFNet scaling used in this step to further delete insignificant edges in the network is the function of “Pathfinder Network Scaling”, provided by the network analysis software, NWB Tool (NWB Team, 2006). To delete as many edges as possible, the parameter r in the algorithm is set to infinite (∞).

5. Network Partitioning: We use the Girvan-Newman method (Girvan, & Newman, 2002), provided by the network analysis software NodeXL (Hansen, Shneiderman, & Smith, 2010), to partition the resulting network of the PFNet scaling into a set of subgroups of nodes. Every subgroup in the output of the algorithm is considered to be an important research topic in the examined field.

6. Graph Drawing: In this study, another network analysis software Pajek (De Nooy, Mrvar, & Batagelj, 2005), is used to visualize the results of PFNet scaling and network partitioning. The visualization method is the Kamada-Kawai algorithm and is very suitable for visualizing the resulting networks generated by PFNet scaling.

7. Topic Labeling: As mentioned above, the subgroups of nodes produced by the Girvan-Newman algorithm are considered to be important research topics in the examined field. To give labels to these topics, we selected the terms with the five highest frequency counts occurring in the documents to correspond to nodes of these subgroups.

Results

The data used in this study

The data used in this study were Masters theses published by graduate

schools related to the field of information communication, retrieved from the database of the National Digital Library of Theses and Dissertations in Taiwan (<http://ndltd.ncl.edu.tw/>). In total, 778 theses were examined in this study. 293 different kinds of terms were extracted from the titles and abstracts of these theses. Feature vectors were generated based on the occurrences of the extracted terms to represent 777 theses, excluding one that was without any extracted term appearing in it, and relevance scores between the theses were estimated to establish a network. The resulting network had 777 nodes and 7168 edges with a density of 2.37×10^{-2} , and there was only one component in the network; in other words, the network was connected.

The results of PFNet scaling

PFNet scaling deleted most edges and the new network had 786 edges (only about 10.97% compared with the original network) with a density 2.61×10^{-3} .

The results of network partitioning and topic labeling

The Girvan-Newman algorithm partitioned the nodes in the new network into 30 subgroups with the condition of maximal modularity. Figure 2 shows the distribution of subgroups in the network. Nodes with the same color indicate that they are in the same subgroup. We assign every subgroup with a code and the term with the highest frequency count in the corresponding theses for convenience of identification. Table 1 shows the number of corresponding theses and the terms with the five highest frequency counts of all subgroups. The largest subgroup is C27 and the smallest subgroup is C15, with 36 and 15 theses, respectively. In Table 1, the terms are almost highly relevant to the field of information and communications or its fundamental disciplines, such as information and communication technology, communication science, and art and design. In addition, terms in each subgroup are also related to each other. Thus, the resulting graph and table can be used as information resources to determine the important research topics and explore how they are related, in the examined field.

Discussion

The resulting graph shows the scope and content of research topics in the field of information communication. Using the graph, we can clearly see that some of the research topics in this field originally came from several fundamental disciplines, which are communication science, library and information science, information communication technology, and art and design. Innovations and impacts from the application of information communication technology on business management, content creation, application services, and user experiences are also important research topics in the field. The above observations confirm Liang's (2011) points. In addition, "e learning" and "digital content", both located

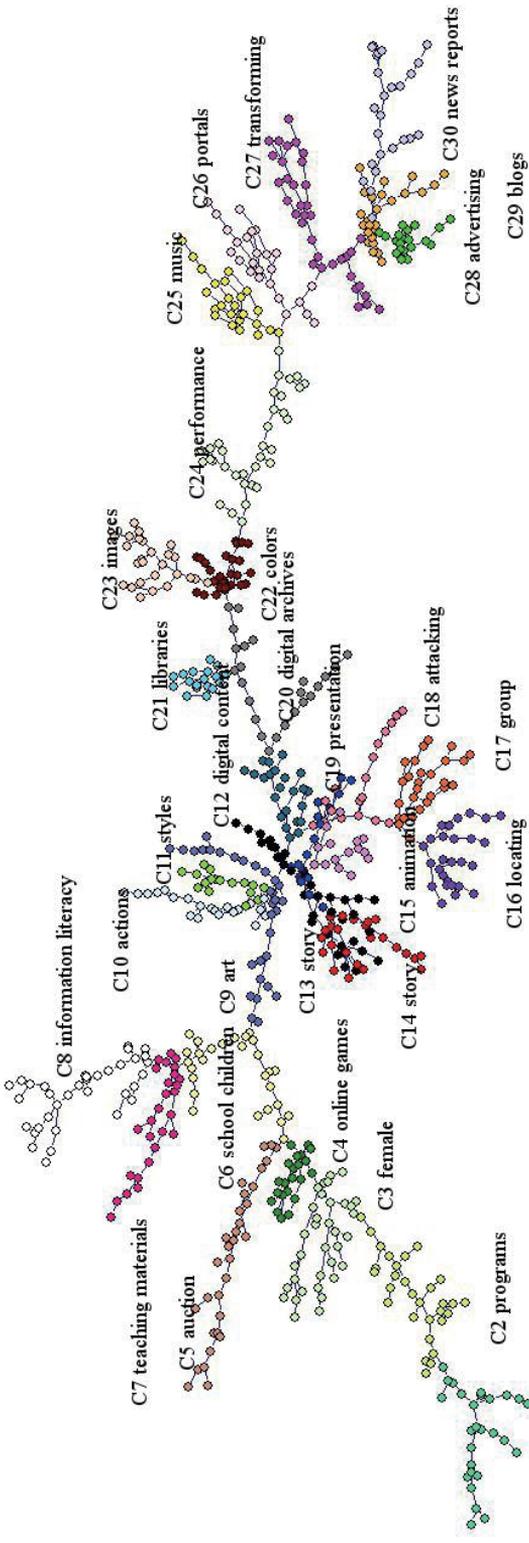


Fig. 2 The Results of Network Partitioning and Topic Labeling

C1 Chungghwa Telecom



Table 1 The Terms with the Five Highest Frequency Counts in Each Subgroup

Sub-group	number of these	Terms with highest frequency count				
C1	27	Chunghwa Telecom	integrated marketing	digital television	magazines	telecommunications
C2	30	programs	video	broadcasting	movies	television stations
C3	35	female	experiences	consumption	Web pages	male
C4	22	online games	players	online game (in English)	forum	female
C5	31	auction	network	transaction	trust	players
C6	28	school children	teaching materials	elementary school	players	grade
C7	27	teaching materials	e-learning (in English)	e learning	web 2 (in English)	e-commerce
C8	35	information literacy	e learning (in English)	e learning	skills	employees
C9	31	art	members	research and development	elementary school	creation
C10	21	actions	features	Chinese	usability	documents
C11	17	styles	graph	shape	art	flat
C12	36	digital content	creative	learner	evaluation	creation
C13	17	story	children	reality	reading	experiences
C14	24	reading	publishing	movies	readers	e-book
C15	15	animation	talents	game industry	colleges	creative
C16	27	locating	signals	dynamic	wireless LAN	tracking
C17	27	group	family	telephone	stress	health
C18	18	attacking	communications	group	protocol	exchange
C19	23	presentation	digital media	projects	digital media (in English)	departments
C20	24	digital archives	navigation	digital archives (in English)	museums	archiving
C21	19	libraries	librarians	library collection	community	readers
C22	27	colors	printing	chromatic aberration	memory	detecting
C23	26	images	retrieval	noise	copying	image quality
C24	34	performance	knowledge management	channels	customers	knowledge management (in Eng.)
C25	25	music	perception	independent	industry	marketing strategies
C26	25	portals	cellular phones	electronic	loyalty	involvement
C27	36	transforming	tags	visualization	algorithms	investment
C28	20	advertising	make-up	blogs	strength	placement
C29	25	blogs	search	people	reliability	search engines
C30	25	news reports	frame	image	China	encoding

in the central region of the graph, were the major research topics of Masters theses in the examined field. “E learning” and “digital content” integrate multiple fundamental disciplines of information communication.

The resulting graph is compared with the self-organizing map generated by Lin (2012) to reveal features of the two methods. Self-organizing maps are good at showing the overall structure of the field. However, the network graph in this study can display the local linking structures of the field, demonstrating the membership among theses and their related topics.

ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT

李敦仁[Lee, Duen R.]、余民寧[Yu, Min-Ning](2007)。學習表現的知識結構評量研究：以「教育統計學」學科知識為例[The assessment of knowledge structures of learning performance in educational statistics]。教育研究與發展期刊[*Journal of Educational Research and Development*]，3(4)，113-148。

林頌堅[Lin, Sung-Chien](2012)。資訊傳播學領域的研究主題發展分析[An analysis of the development of research topics in the field of information communication]。資訊傳播研究[*Journal of Information Communication*]，2(2)，1-19。

梁朝雲[Liang, Chao-Yun](2011, 3月)[2011 March]。尋找「資訊傳播」定義、發現「資訊傳播」研究[In search of the consensus definition and its research agenda of “Information Communication”]。在[In]淡江大學資訊與圖書館學系主辦[Department of Information and Library Science, Tamkang University]，《教育資料與圖書館學》40週年國際學術研討會[*The International Conference of the 40th Anniversary of Journal of Educational Media & Library Sciences*]，新北市[New Taipei City]。

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179-255.

Börner, K., Sanyal, S., & Vespignani, A. (2007). Network science. *Annual Review of Information Science and Technology*, 41(1), 537-607.

Chen, C. (1998). Bridging the gap: The use of Pathfinder networks in visual navigation. *Journal of Visual Languages and Computing*, 9(3), 267-286.

Chen, C. (2003). *Mapping scientific frontiers: The quest for knowledge visualization*. London, UK: Springer-Verlag.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of American Society for Information Science and Technology*, 57(3), 359-377.

Chen, C., & Carr, L. (1999). Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998). In J. Westbomke, U. K. Wiil, J. J. Leggett, K. Tochtermann, & J. M. Haake, (Eds.), *Proceedings of the tenth ACM Conference on Hypertext and hypermedia: Returning to our diverse roots: Returning to our diverse roots(HYPertext '99)* (pp.

- 51-60). New York, NY: ACM.
- Chen, C., & Morris, S. (2003). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on* (pp. 67-74). Seattle, WA: IEEE.
- Chen, C., & Paul, R. J. (2001). Visualizing a knowledge domain's intellectual structure. *Computer*, 34, 65-71.
- Chen, P., & Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetrics*, 4(3), 278-290.
- Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6), 066111.
- Colliander, C., & Ahlgren, P. (2012). Experimental comparison of first and second-order similarities in a scientometric context. *Scientometrics*, 90(2), 675-685.
- de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek* (Structural Analysis in the Social Sciences, vol. 27). New York, NY: Cambridge University Press.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821-7826.
- Groh, G., & Fuchs, C. (2011). Multi-modal social networks for modeling scientific fields. *Scientometrics*, 89(2), 569-590.
- Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Amsterdam, Netherlands: Morgan Kaufmann.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1), 7-15.
- Lin, X. (1997). Map displays for information retrieval. *Journal of the American Society for Information Science*, 48, 40-54.
- Morris, S. A. & Van der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology*, 42(1), 213-295.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577-8582.
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113.
- NWB Team. (2006). *Network Workbench Tool*. Indiana University, Northeastern University, and University of Michigan, Retrieved November 20, 2012, from <http://nwb.slis.indiana.edu>.
- Schubert, A., & Soós, S. (2010). Mapping of science journals based on h-similarity. *Scientometrics*, 83(2), 589-600.
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595-610.
- Takeda, Y., & Kajikawa, Y. (2009). Optics: A bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, 78(3), 543-558.
- van Eck, N. J., Waltman, L., Noyons, Ed C. M., & Buter, R. K. (2010). Automatic term

- identification for bibliometric mapping. *Scientometrics*, 82(3), 581-596.
- Wallace, M. L., Gingras, Y., & Duhon, R. (2009). A new approach for detecting scientific specialties from raw cocitation networks. *Journal of the American Society for Information Science and Technology*, 60(2), 240-246.
- Waltman, L., van Eck, N. J., & Noyons, Ed C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629-635.
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of American Society for Information Science and Technology*, 54(3), 423-434.
- Yan, E., Ding, Y., Milojević, S., & Sugimoto, C. R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. *Journal of Informetrics*, 6(1), 140-153.
- Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on ISI category classification. *Journal of Informetrics*, 4(2), 185-193.
- Zhao, H., & Lin, X. (2010). A comparison of mapping algorithms for author co-citation data analysis. *Proceedings of American Society for Information Science and Technology*, 47(1), 1-3.