自動鏈結分析演算法在社會網絡之開發與應用

江憲坤
彰化師範大學資訊管理系

陳鴻文
銘傳大學資訊傳播工程系

楊境榮
銘傳大學資訊傳播工程系

摘要

在關聯分析或序列樣式分析之資料探勘研究中，即使採用了多重門檻值的設定來過濾大資料集合，仍會找到過多無用且信度過低的關聯法則，或可能遺漏了頻率較低但實質上卻具有高度價值的資料項目。此外，除了少數特定問題外，以往鏈結分析之研究，都需要仰賴專家來目測已轉換為視覺化的資料，來進行主觀評估，以發現資料之規則性。此種衡量和評估的方式，對於複雜之網絡，往往費事耗時且成效不彰。而一些社會網絡之研究也指出，頻率低的弱鏈結扮演著聯繫不同群體之重要角色。因此，本研究透過基本的圖學理論，提出一個不需要依賴門檻值設定，就能找出存在於網絡中的弱鏈結及關鍵弱鏈結路徑之自動鏈結演算法。本研究再利用真實的安隆企業電子郵件資料，配合NetDraw視覺化的網絡分析工具，以實驗來檢驗本自動化鏈結分析演算法之可行性及正確性。

關鍵字：資料探勘、鏈結分析、弱鏈結、關聯法則

The Development and Application of an Automatic Link Analysis Algorithm for Social Networks

Heien-Kun Chiang
Department of Information Management, National Changhua University of Education

Hown-Wen Chen
Department of Computer & Communication Engineering, Ming Chuan University

Jing-Rong Yang
Department of Computer & Communication Engineering, Ming Chuan University

Abstract

Even with the settings of multiple thresholds when screening large data sets, using link analysis or sequential patterns analysis, many data mining studies obtain lots of not-very-useful low-confidence association rules or miss the low-frequency but actually highly valuable data items. In addition, except for some specific problems, previous link analysis researches mostly rely on experts' subjective visual investigations of analyzed data which are transformed into visual form in order to find the data's regularities. This kind of assessment and evaluation is usually time-consuming and inefficient for complicated networks. Prior studies of social networks have revealed that low-frequency (weak) links play important roles in connecting different cliques in a social network. Therefore, utilizing the topology in graph theory this study proposes an automatic link analysis algorithm without depending on the thresholds to discover the weak links and the key weak link paths in a network. To check and see the feasibility and accuracy of the proposed algorithm, empirical studies on the well-known Enron e-mail data sets using the NetDraw network visualization tool are conducted, and the results are found to be positive.

Key words: Data mining, link analysis, weak link, association rules

壹、緒論

一、研究背景

現今企業普遍面臨到的問題已不再是資訊的不足，而是資訊超載(information overload)，過多的資訊往往難以用人工方式來慢慢地分析處理。因此往往需要先以資料倉儲(data warehouse)的方式，來將龐大的資料進行有效的分類及儲存，再透過資料探勘(data mining)技術的應用，來協助業界從龐大的資料庫中找出潛藏的資訊，以呈現給決策者。而鏈結分析(link analysis)的探勘技術，主要是將許多具有關聯性的資料，以節點(node)及鏈結(link)表示成圖形(graph)之後，再針對此圖形進行分析，以期檢視和發覺資料間的互動關係，和找出其中功能殊異的資料個體，甚至一些隱含的特性。分析所得的鏈結資訊可被應用於犯罪偵察(Xu & Chen 2004)或探勘人際網絡方面，例如美國聯邦調查局採用鏈結分析來追查奧克拉荷馬市(Oklahoma city)炸彈案的兇嫌(Berry et al. 1997)；亦可以應用於網絡分析、追蹤，及行銷決策的基礎，如美國電信業者利用鏈結分析與資料屬性，找出專用傳真的號碼及語音、傳真共用之號碼；及收集顧客使用電話的時間與頻率，進而推斷顧客的使用偏好，提出有利於公司的方案，用以行銷或提高服務品質(Berry et al. 1997; Westphal & Blaxton 1998)。而英國廣播公司(BBC)則用來分析收視分佈，以提供節目製作的參考(Adriaans & Zantinge 1999)。

然而目前的鏈結分析技術卻仍遭遇到下列的問題：

（一）鏈結分析難以自動化處理

與一般資料探勘自動萃取資料特性的方式不同，現存絕大部份鏈結分析是屬於半自動的視覺化探勘方法；其方式是將多維度的資料投影到二維或三維空間，以視覺化的物件呈現出圖形不同的色彩、形狀及大小等，藉著人類視覺與智慧辨認特性及對複雜圖形的快速認知能力的輔助，讓使用者及領域專家可以動態檢視及探索其有興趣的部份，進而分析出資料的模式。但缺點在於需要耗費專家人力，尤其在大量資料的情況下，由於資料呈現上的不易，容易產生視覺上的混淆，進而造成主觀性的誤判，且難以應用在即時性的系統分析中。

目前最常見的自動化鏈結分析之運用，則是搜尋引擎用來衡量網頁或超鏈結價值性的網頁超鏈結分析(hyperlink analysis)，亦稱為connectivity-based ranking (Henzinger 2001)或link analysis ranking (Borodin et al. 2005)。較為廣為人知者如Google所採用的「PageRank」技術 (Brin & Page 1998)，原理是運用學術界論文引述的概念，因為一篇論文被其它論文引用參考的次數，大致可視為該篇論文的重要性及品質。此方法目的是在計算某網頁本身之重要性價值，主要偏重在考慮鏈結數量的多寡和參照網頁的重要性；然而由於並未考慮到圖形連接的拓樸特性，故未必適用於所有鏈結分析問題的處理。

（二）資料特性門檻值的不易界定

常見的資料探勘技術在使用時，大都會遭遇到門檻值的設定問題：例如在關聯分析

的研究中，為了發掘出稀少但具高價值的項目集，大都需利用到多重門檻值(Chung & Lui 2000; Liu et al. 1999)，但須事先對於資料進行適當的分類，再個別地設定適當的最小支持度；或是由Yun (2003)所提出的RSAA演算法，藉由另外設定第二個門檻值，得以分階段進行探勘。然而對於真實世界複雜多變的問題，這些著重於頻率門檻值的研究，大都會遭遇難以順利解決的時刻。

門檻值選擇之高複雜性不妨參考社會學家Granovetter提出的「暴動門檻」案例（胡守仁 2002）：假設每個人都有一個參加暴動的「門檻」，亦即大多數人不會無緣故的參與暴動，但在特殊的情勢(如在被「逼瘋了」的狀況)之下，就可能會參加。每個人的門檻值高低可能與他的性格有關，或與他是否在乎觸法有關，而團體行為的複雜度則更難以預知。例如在酒吧裡旁觀的一百人當中，每個人的暴動門檻值分別為0到99，也就是有一個人的門檻為0，第二個人為1等等；在這種狀況之下，暴動是無法避免，因為那個門檻值為0的「極端份子」會掀開暴動，然後門檻為1的人就會加入，而後暴動就像野火般蔓延開來。但如果單單只是將門檻值為1的那個人調高為2，其他人都不變，那麼在第一個人開始搗毀東西時，其他的人只會旁觀，那麼暴動的連鎖反應將不會產生。藉由此例不難發現，對於同一問題，很可能由於單一筆資料屬性的些許改變，導致原本適用的門檻值無法再發揮作用；廣義而言，解決各個問題適切性之門檻值的確難以制訂，因為影響的變數層面太過於複雜；故很難保證藉由門檻值的設定，就能獲得我們想要的分析結果。所以在現實多變的生活中，我們應該思考如何加入或使用其它的考量因素，來解決或避免這個問題。

（三）弱鏈結優勢理論

1960 年代美國心理學家Milgram曾試圖做了一個研究人際網絡聯繫的有趣實驗，他在內布斯加州(Nebraska)和堪薩斯州(Kansas)隨機選取並寄信給一些人。Milgram在沒有提供這些人他在在波士頓一位朋友地址的情況下，請這些人設法透過自身的人際網絡來轉寄他的信給他在波士頓的朋友。Milgram的研究結果出人意料，大多數的信都安全寄到了他在波士頓的那位朋友的手中，更令人訝異的是這些信件的轉寄次數大約僅需六次左右；此種現象後來被稱為「六度分隔」(six degrees of separation)理論（胡守仁 2002），亦即世界上任何兩個看似毫不相關的人，只要透過不超過六個人際關係的鏈結，便可將這兩個人串連在一起。

而社會學家Granovetter (1973)的研究則進一步指出，促使全世界人際關係形成一個有效網絡(或稱為小世界：意指能以類似六度分隔形態組成的網絡)的關鍵因素並不是來自於強鏈結(strong link)　來自家人、親朋好友間的綿密關係；相對地，而是來自弱鏈結(weak link)　就是點頭之交的朋友關係。經由這些泛泛之交的朋友做為橋樑(bridge)，人們即可聯繫原本互無關聯的不同群體，使得我們的人際關係得以延伸。如圖1所示，即使其它鏈結頻率再高，節點C與D間的鏈結仍具有不可替代之重要地位，而這也正是現有利用頻率門檻設定的演算法極難找出的特徵。因此，Granovetter (1973)提出「弱鏈結優勢(the strength of weak link)」理論，即弱鏈結在不同叢集 (clique)間，扮演的是訊息交換的橋樑。倘若去除了這些弱鏈結，那麼整個世界將可能支離破碎，形成無數個孤單的個

體。因此可知弱鏈結具有著不可替代的重要性。

有鑑於弱鏈結優勢理論和節點間最短路徑（Geodesic distance）的觀念，本研究提出以圖學理論及拓樸關係為基礎，加入社會網絡研究中弱鏈結的考量要素，以降低對於資料頻率門檻值的依賴，和免除進行視覺化分析時人力的耗費和過於主觀的兩大問題。本研究亦實際處理真實的美國安隆(Enron)企業電子郵件資料庫，並佐以視覺化的分析工具，以說明本自動鏈結分析演算法之可行性和效能。本研究之主要貢獻為(1)利用圖形拓樸特徵來找出圖形連結中具有特殊拓樸意義的個體，如弱鏈結、接合點及關鍵弱鏈結路徑，此不同於一般網頁超鏈結分析利用連結頻率來計算其重要性；(2)提出一個不需依賴頻率門檻值的新鏈結分析演算法，來自動找出社會網絡研究中具有高價值卻不易被發現的弱鏈結，以解決先前著重在頻率門檻值分析之關聯法則研究，卻無法找出弱鏈結之缺失；(3)本研究提出的自動化鏈結分析演算法，可自動判斷出圖形中弱鏈結、重要節點與叢聚等特徵，將大大提升分析效率，解決視覺化資料探勘需要耗費大量人力在鏈結資料的主觀分析判斷之效率問題。

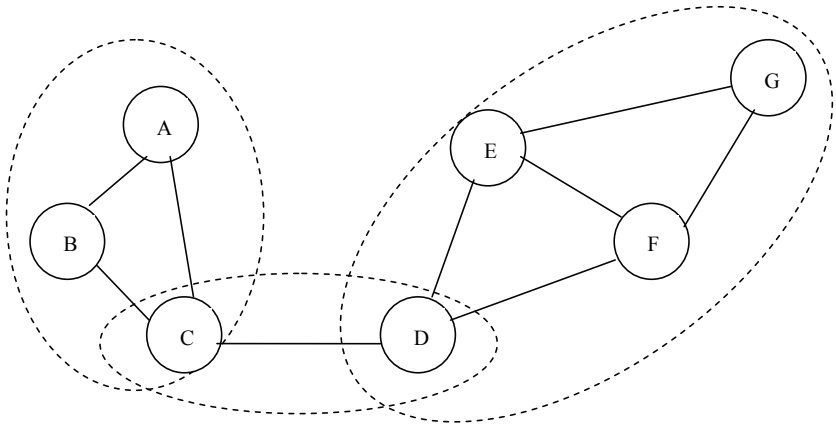

圖1：節點C與D間存在具有橋樑功能的弱鏈結

貳、文獻探討

一、社會網絡分析及關係鏈結

社會學者對於「網絡」(network)曾下過幾種定義：Knoke 與 Kuklinski (1982)解釋「網絡」是一群人、個體或事件間關係鏈結的特定型態；學者吳寶秀（民79）指出「網絡」就是幾個節點(nodes)和節點與節點之間的連線所構成的結構，而節點可為個人、角色或組織團體。蕭新煌（1998）則將網絡視同「關係」或「鏈結」的同義詞。

在進行人際網絡分析時，一般將社群中的個人視為節點，而人際關係則為連接社群中個體所造成之鏈結關係；其結構與動態過程，即為「社會網絡」概念之基礎。而社會網絡的定義與人際關係雖很接近，但其所連接的不僅是個人與個人間之關係網絡，而更是家庭與家庭間，甚至不同社群之間的互動聯繫關係，可以比人際關係的涵蓋面更廣。

社會網絡包括以下四個要素：

行動者(actor)：行動者為社會網絡中擔任不同角色的個體節點，可為不同之社會實體(social entity)，如個人、社會及組織等。

關係(relationship)：行動者間的關係可以內容(content)、強度(strength)與方向(direction)衡量。

鏈結：學者Garton et al. (1997)指出鏈結是指兩行動者間的關係組合。Granovetter (1973)將關係鏈結區分為：強的(strong)、弱的(weak)以及不存在(absent) 三種，而鏈結強度的衡量有「互動時間」(interaction time)、「情感強度」(emotional intensity)、「親密程度」(intimacy)或「互惠行動」(reciprocal service)四項屬性。具體而言，在人際網絡裡，經由強關係鏈結所獲得的訊息，多為重覆性高或類似之訊息，較少創新之訊息；而透過弱關係鏈結形成的網絡，因其異質性較大，所提供的訊息將較為豐富有用。

叢聚(clique)：行動者基於彼此的相似性組合而成的次團體(sub-set)，所屬成員間關係緊密結合，並與團體外的行動者沒有緊密的鏈結關係存在。所以扮演兩個或多個次團體間溝通的弱鏈結關係，經常在整體網絡中居於重要地位，因為次團體間的溝通將隨著弱鏈結的消失而消滅。

社會網絡分析(social network analysis)主要透過網絡圖形分析，找出各種不同社會網絡的功能與目的，將有助於瞭解人際聯結的過程。

二、圖形理論

（一）名詞定義

在無向圖形G = (V, E)中，V與E分別代表圖中的節點及鏈結所形成的有限集合(finite set)，其中$V = \{v_1, v_2, \cdots, vn\}$，$E = \{e_1, e_2, \cdots, em\}$；此外，以TCC(G)代表圖G中的連通元件的總個數。

1. 二元連通元件(biconnected component)：對於一個至少具有兩條鏈結的無向連通圖G'中的任意兩個不同節點vi及vk之間，若皆存在至少兩條不同的簡單路徑，亦即存在一包含vi及vk的簡單迴路(simple cycle)，則稱G'為一個二元連通圖(biconnected graph)。若二元連通圖G'為原圖G的最大子圖集，即稱圖G'為圖G中的一個二元連通元件；如圖1中的{A,B,C}、{C,D}及{D,E,F,G}皆為二元連通元件。
2. 接合點(articulation point)：若將圖G 中某一節點v_i以及與v_i有相連接的鏈結移除後，形成的新圖稱之為$G' = (V',E')$，其中$V' = V - \{v_i\}$，且$E' = E - \{(vi, vk) \mid k \neq i\}$。若TCC (G') > TCC (G)，亦即G'被分割成更多的連通元件，則稱v_i 為圖G的一個接合點，亦可稱為切點(cut point)，如圖1中的節點C、D皆為接合點。因若將C、D任一點移除，圖G都會從原本的一個連通元件，變成兩個連通元件。
3. 橋樑(bridge)：將圖G中某一個鏈結 e_i 移除所形成的新圖$G' = (V, E')$，其中$E' = E - \{e_i\}$。若TCC(G') > TCC(G)，則稱e_i為圖G 的一個橋樑；如圖1中的鏈結<CD>。

定理：二元連通元件將形成圖形鏈結之完全分割(partition)

此定理保證利用二元連通元件來進行圖形分割時，可包括所有的節點及鏈結，而不會有所遺漏，且每一鏈結僅會包含在單一個二元連通元件內部（僅出現一次）(Biconnected Component Partition 2003)。

4. 強連通元件(strongly connected component)：對於一個至少具有兩條鏈結的有向連通圖G'中的任意兩個不同節點vi及vk之間，若皆存在一條以上路徑，則稱G'為一個強連通圖(strongly connected graph)。若強連通圖G'為原圖G的最大子圖集，即稱圖G'為圖G中的一個強連通元件(Cormen et al. 2001; Diestel 2000)。

（二）基於深度優先搜尋的二元連通元件演算法

深度優先搜尋(depth-first search) 演算法是先任意選擇一個節點v為起始點，然後以所到達之相鄰節點，繼續做為新的起點，如此以縱深的方式拜訪節點，直到無法再前進時，才回溯到上一層改採寬橫的方式，繼續探索其它尚未拜訪的節點，直到所有節點都走完為止。利用深度優先搜尋演算法，可以搜尋出無向圖形中所包含的所有二元連通元件；本文所採用的詳細步驟，可以參閱基礎演算法(Weiss 1993)或圖學理論習相關文獻(Diestel 2000; Cormen et al. 2001; Valiente 2002)。

（三）基於深度優先搜尋的強連通元件演算法

由於每個強連通元件，必定是一個循環圖(a graph with cycle)，因此也可使用深度優先的觀念來處理。此搜尋強連通元件的演算法(Weiss 1993)與先前所述「基於深度優先搜尋的二元連通元件」演算法相似，可以保證每一個節點及鏈結只經過一次，時間複雜度亦為O(|E| + |V|) (Valiente 2002)。

三、鏈結分析技術在Enron資料庫之應用

一般處理Enron資料庫的研究，大致上可區分為

利用自然語言剖析技術:：嘗試經由內文的分析(Keila & Skillicorn 2005; Berkeley 2006)，或以統計方法、matrix factorization (McCallum et al. 2005; Berry & Browne 2005)等，透過不同特徵的比重，來依照郵件內容分類其主題。

依照程序的先後關係：形成在通訊或業務上的序列關係（chain）(Lawu et al. 2005; Shetty & Adibi 2005)。

將多維度的資料投影到二維或三維空間：以視覺化的物件呈現出圖形不同的色彩、形狀及大小等，讓使用者及領域專家可以動態檢視，進而分析出資料的模式(Visual_complexity 2006)。

PageRanking：在計算某論文或網頁的重要性價值時，主要偏重在考慮鏈結數量的多寡和參照論文本身的重要性（Duan et al. 2005; Borodin et al. 2005; Henzinger 2001; Brin & Page 1998）。

以統計特徵值來觀察：隨著時間改變之使用者郵件數量的變化是否合理 (Priebe 2005)。

利用圖學理論：來計算圖形所呈現出的特徵數值，如Between-Centrality, Closeness-

Centrality, In_Degree Centrality, Out_Degree Centrality, Cluster Diameter, Component數目等（Diesner & Carley 2005; Chapanond et al. 2005）。這也是一般的社會網路工具如UCINet (UCINet 2006)，常採用的分析方式。

本研究採取類似(六)之理論模式。但相較之下，這些研究及分析工具偏重於以平均數值來顯現整體或某群體的平均特性，而本研究方法則是偏向於直接找出具有特殊拓樸意義的個體。

參、處理流程暨演算法

由前述的鏈結分析相關研究之探討可得知，目前尚缺乏一個能夠自動分析資料中鏈結關係的演算法。根據「六度分隔」之理論，本研究將以圖論中最短距離之觀念來找尋圖形中的弱鏈結。使用的方法是首先將關係鏈結圖中的鏈結逐一移除，然後計算各節點間的最短距離值；以此再與原來圖形之最短距離值集合來比較，判斷鏈結移除前後是否呈現「顯著」差異，並藉此來決定此鏈結是否重要。但此演算法所需的計算時間過久，且需定義差異顯著的門檻值，故較難以實際應用在大量資料的分析。為了改良這些缺點，本研究應用前述尋找連通元件之演算法，將所求得二元或強連通元件 (節點數大於2者)內的鏈結及枝葉鏈結去除之後，所剩下的鏈結便是所有可能成為結構上的橋樑候選者。

一、資料處理流程

如圖2所示，本研究主要架構可分為以下幾個部份：

資料庫處理：選擇適當的資料庫及資料來源，並進行資料淨化(data clean)篩選，以產生適當、可供分析的資料。

圖形結構分析（本研究重點）：

1. 圖形轉換：依據來源資料及問題的特性，將之轉換為有向或無向圖呈現。
2. 相鄰矩陣轉換：將項目間關係次數以相鄰矩陣(adjacency matrix)的方式來記錄，以利後續的探勘步驟。
3. 求得所有的二元或強連通元件：以基於深度優先搜尋演算法，在無向圖中，找尋出所有的二元連通元件。而在有向圖中，則找尋出所有的強連通元件。
4. 結構橋樑分類：去除所有二元或強連通元件(節點數大於二者)內部的鏈結，得到所有結構上可能的橋樑關係，並進行適度過濾及分類。

頻率分析（非本研究重點）：

特別針對加權圖形，可用關聯法則分析演算法，經由預先設定的最小支持度，來得到大二項集合(large 2-itemsets)，而集合中項目間的關係即為強鏈結。

整合解釋：分別將兩演算法所得到的結果，經由社會網絡理論的整理解釋，便可得到使用者所關心的鏈結關係及其闡釋。

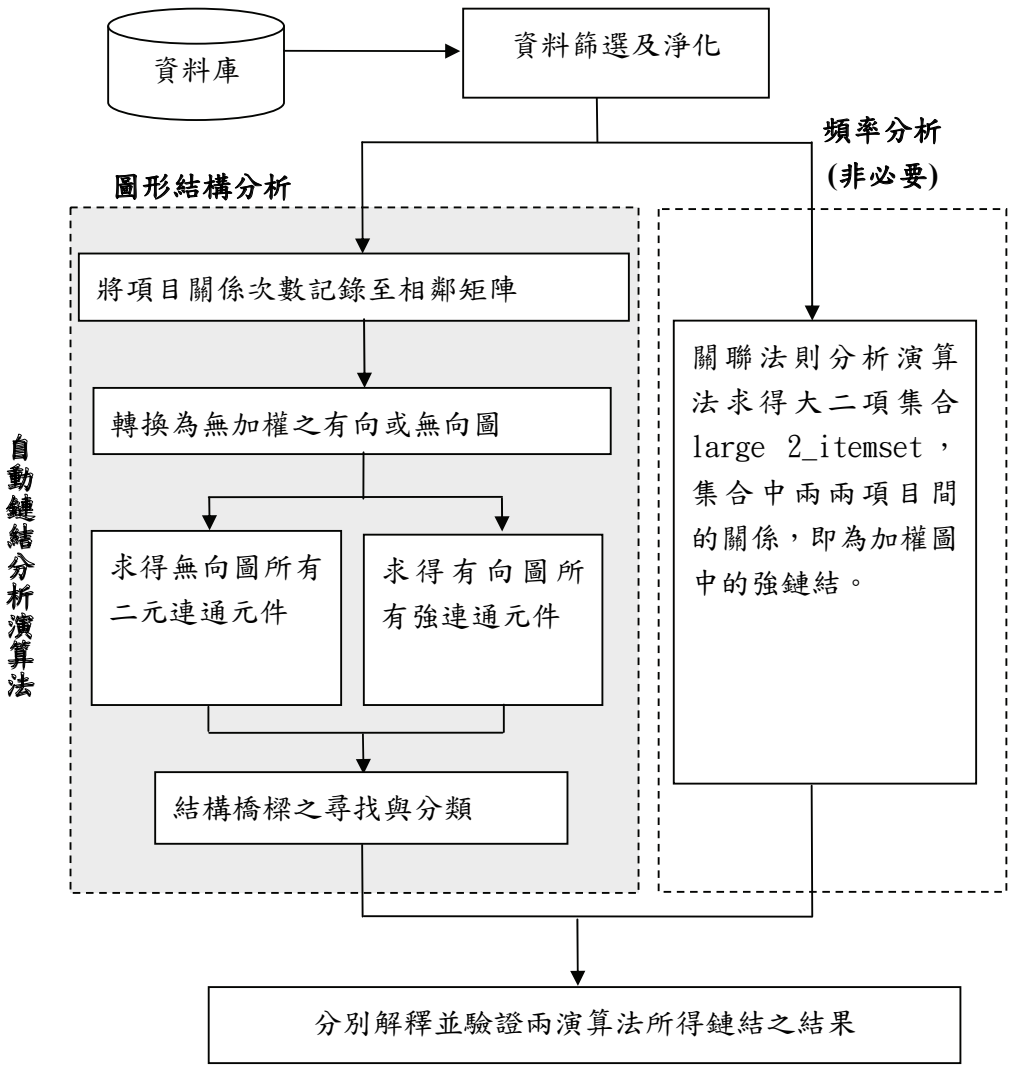

圖2：本研究之資料處理流程架構

二、本研究之鏈結分析基本運作過程

以圖3為例，說明本研究之鏈結分析的運作方式：首先執行一次枝葉節點及鏈結之清除，故捨棄鏈結<HJ>。接著利用前述尋找二元連通元件的演算法，找出{ABC}及{DEFG}兩個節點數目大於2之二元連通元件，亦可視為將此圖形節點主要區分為此兩大群組；在尋找二元連通元件的同時，也標示了節點A及G是接合點。若分別以接合點來代表此兩個連通元件，則很容易觀察出其間主要是靠鏈結集合{<AH>,<HI>,<IG>}來形成一條關鍵弱鏈結路徑，其功能宛如一條弱鏈結。同理，在圖1中，則是由弱鏈結<CD>負責來連接兩個二元連通元件。而本研究的主要工作，就是自動地標示出圖形中之連通元件、關鍵弱鏈結路徑及弱鏈結這些重要特徵。

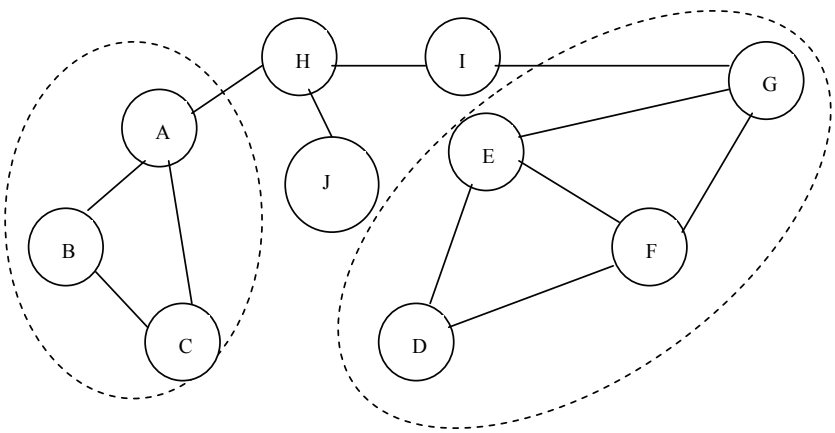

圖3：無向圖鏈結分析範例

三、 自動鏈結分析演算法（圖形結構分析）

假設交易資料庫中包含n個個體，且具有m筆關連項目。本研究提出的自動鏈結分析演算法步驟說明如下：

步驟 1：掃描資料庫的m筆資料，並將涉及相同個體的關係次數記錄至n*n相鄰矩陣之中。

步驟 2：一般情況下，直接將此相鄰矩陣中非零者視為無加權圖中的鏈結即可。故掃瞄資料庫完成後，轉換成二元(binary)相鄰矩陣，可以用來表示有向或無向簡單圖。以下圖形仍以G = (V,E)的方式呈現。

步驟 3：計算各節點的分支度(degree)。

步驟 4：去除圖形G中僅有一個相鄰節點的枝葉節點及鏈結（僅執行一次），且將這些枝葉鏈結記錄至集合LL之中。

步驟 5：選取圖形G中任意一個分支度不為0的節點為起點，尋找包含此節點的二元連通元件BC_i（但不包含節點數剛好為二者），並將元件內之鏈結記錄到BC_i集合之中；或搜尋有向圖包含此節點的強連通元件，並將元件內之鏈結記錄到集合CC_i之中。

步驟 6：若為無向圖，則另將搜尋出的接合點記錄到集合AP之中；若處理的是有向圖，則嘗試逐一將同時具有入分支及出分支鏈結的節點移除，以深度優先搜尋來判斷與之相鄰的節點是否因此無法透過其它路徑到達；若是即為類接合點，將之記錄到集合PAP之中，直至所有節點皆處理完畢為止。

步驟 7：重覆步驟5及6。直至處理完圖形G中所有的連通子圖，即所有鏈結都拜訪過為止。

步驟 8：對於集合BC_i或CC_i內部的所有節點，皆可視為同一個虛擬節點。又此時若處理的是有向圖，可針對每一強連通元件的內部，找到連接局部資訊集中點集合LISink與局部資訊來源點集合LISource的鏈結，將之記錄到集合$LISL_i$之中。

步驟 9：以虛擬節點表示原圖形G，亦即不考慮圖形G在集合BC_i或CC_i內部的所有鏈結，其餘即為包含潛在弱鏈結的集合G'。若在圖形G'中某一鏈結的兩端為接合點或隸屬不同BC_i或CC_i元件，則此鏈結即為弱鏈結，並將之記錄到集合WL中；若此弱鏈結兩端隸屬不同BC_i或CC_i，則需進一步進行此兩個連通元件的合併。

步驟 10：之後再針對任兩虛擬節點及接合點之間，進行Dijkstra最短距離路徑(陳會安2002)搜尋，可得到一關鍵弱鏈結路徑CP_j。

步驟 11：重覆步驟10，直至圖形G中任兩虛擬節點間皆處理完畢為止。

步驟12：在處理完上述所有步驟後，所剩下之鏈結即為一般性的潛在弱鏈結，將之記錄至集合PWL中。

步驟 13：若處理的是有向圖，最後再找出全域資訊集中點集合GISink與全域資訊來源點集合GISource的鏈結，記錄到集合$GISL_i$中。

步驟8中虛擬節點的形成及步驟9中連通元件的合併處理範例，可見圖4的說明。圖4中四個虛線橢圓區域，都為二元連通元件，故皆可視為虛擬節點。而圖4中實線橢圓區域顯示弱鏈結(v_a,v_b)直接連接了不同的二元連通元件，故可再進一步將其合併成單一虛擬節點。而各步驟執行之確切成效將於後續實驗中詳加說明。

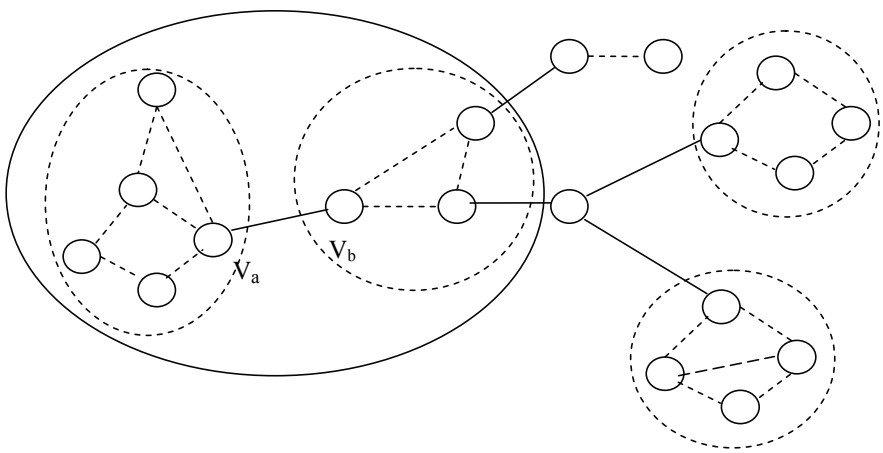

圖4：二元連通元件形成虛擬節點暨合併示意圖

綜合上述演算法步驟，此演算法所使用到名詞意義如下：

資訊集中點(information sink)：有向圖中，資訊集中點可分為兩種：

1. 局部資訊集中點(local information sink)：具有兩個以上直接來自同一強連通元件且皆為進入鏈結的節點。
2. 全域資訊集中點(global information sink)：具有兩個以上直接來自不同強連通元件且皆為進入鏈結的節點。

資訊來源點(information source)：有向圖中，資訊來源點可分為兩種：

1. 局部資訊來源點(local information source)：具有兩個以上直接指向到同一強連通元件且皆為輸出鏈結的節點。

2. 全域資訊來源點(global information source)：具有兩個以上直接指向到不同強連通元件且皆為輸出鏈結的節點。

類接合點(pseduo articulation point)：在有向圖中，若將類接合點移除，會使得原本僅能依靠此點連接的兩相異節點，無法經由其它路徑到達。

綜合上述演算法步驟，此演算法所使用到各集合的直觀意義如下：

LL：原始連接枝葉節點與其唯一相鄰節點的鏈結集合。

BCi或CCi：在無向圖中BC_i為二元連通元件鏈結集合，而在有向圖中相對應CC_i則為強連通元件鏈結集合。

AP或 PAP：無向圖中的接合點集合，或有向圖中的類接合點集合。

PWL：不為枝葉鏈結，亦不包含在任何連通元件之中的鏈結（潛在弱鏈結）集合。

WL：用以連接接合點或連通元件的單一鏈結，所形成之弱鏈結集合。

CPj：用以連接兩個不同連通元件的非單一鏈結(路徑)集合。

LISink及LISource：局部資訊集中點及來源點集合。

LISLi：對於局部資訊集中點或局部資訊來源點，與其對應的強連通元件所連接的鏈結集合。

GISink及GISource：全域資訊集中點及來源點集合。

GISLi：對於全域資訊集中點或全域資訊來源點，與其對應的強連通元件所連接的鏈結集合。

肆、實驗與結果評估

本研究採用所提出之自動化鏈結分析演算法，來分析真實的安隆(Enron)公司之電子郵件資料集，並配合視覺化的分析工具，來輔助說明結果。

一、實驗資料來源

安隆公司電子郵件資料集最早由美國聯邦能源管制委員會(Federal Energy Regulatory Commission)在調查該公司時所公佈，而後CALO 學會(Cognitive Assistant that Learns and Organizes)整理並解決資料之完整性後，由卡內基美隆(Carnegie Mellon)大學的學者William將結果公佈在網站(Enron Email Dataset 2006)上。其中包括151名員工在1998年至2002年間所收發的517,431封電子郵件(不包含附件檔)，分散在約3500個資料夾之中。學者Shetty與Adibi在整理上述版本資料集，去除資料重覆性後，剩下的郵件共有252,759封，並將所整理的結果以MySQL資料匯出檔公佈在網站(Enron Dataset 2006)上，以供各相關領域的學者應用；其中記錄了員工姓名、電子郵件信箱、郵件標題和收件者等有用資訊，其資料庫schema如圖5所示。

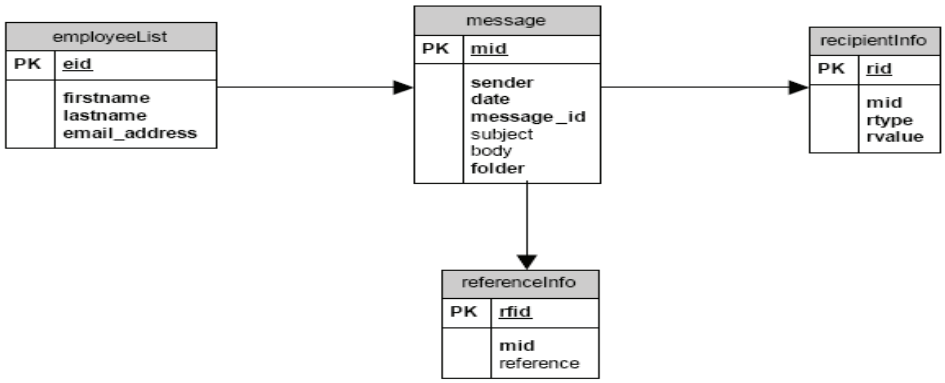

圖5：Enron 電子郵件資料庫shema (Enron Dataset 2006)

二、前置處理

（一）資料庫前置處理

在電子郵件傳遞的過程中，寄件者與收件者之間存在寄信的有向鏈結關係。在此資料集中的151位員工，可個別視為人際網絡中的行動者節點，對於兩節點之間的鏈結強度則以信件寄送的次數來計算。但因為在電子郵件內容中一般會包含著寄件者的電子郵件信箱，可供收件者直接回覆，因此本研究為了簡化由寄送電子郵件而產生的關係鏈結，將其視為具有一雙向對等的鏈結關係，以形成無向加權圖形。亦即直接將原關係相鄰矩陣中的上下三角矩陣值相加，並存回相對應的位置，可形成一對稱矩陣。擷取其部份矩陣內容，如表1所示。

表1：員工間相互寄送郵件次數之部份關係矩陣

	1. Badeer	2. Hyatt	3. Geaccon	4. Lokey	5. Ring	6. Taylor	7. Staab	8. Pereira	9. Panus	10. Allen
1. Badeer	0	0	0	0	0	0	0	0	0	0
2. Hyatt	0	0	14	29	0	0	0	0	0	0
3. Geaccon	0	14	0	14	0	0	0	0	0	0
4. Lokey	0	29	14	0	0	2	0	0	0	0
5. Ring	0	0	0	0	0	0	0	0	0	0
6. Taylor	0	0	0	2	0	0	0	0	94	0
7. Staab	0	0	0	0	0	0	0	0	0	0
8. Pereira	0	0	0	0	0	0	0	0	0	0
9. Panus	0	0	0	0	0	94	0	0	0	0
10. Allen	0	0	0	0	0	0	0	0	0	0

（二）視覺化時參數之選擇

因本資料集節點與鏈結數量過多，若將原始資料所形成之無向圖，直接以視覺化的方式呈現，如圖6所示，會有視覺混亂(visual clutter)的狀況；亦即針對原始資料，若想要以視覺化工具來輔助說明本研究提出演算法計算結果的正確性，將難以目視發現圖形中的弱鏈結及叢聚性。因此需要設定一參數，來減低圖形上的複雜度，以方便我們進行目視辨識。但此一參數之選擇，並不會影響到鏈結分析演算法之執行過程。

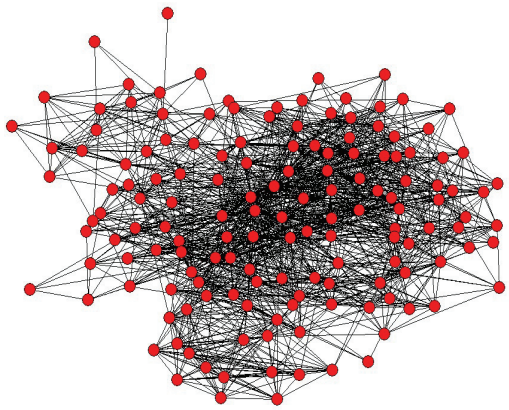

圖6：員工間郵件寄送之完整關係網絡圖

其中若寄件次數為0，表示節點間的鏈結及關連性不存在，因此首先把鏈結權重為0的資料刪除，剩下的資料共計2149筆，而寄件次數與信件累計次數統計圖，如圖7。在無向網絡圖分析的部份，我們欲留下前5%最有意義的資料，所以將參數設定為第95個百分位數，其值為91。在有向網絡圖分析的部份，則留下10%的資料，參數設定為第90個百分位數，其值為81。

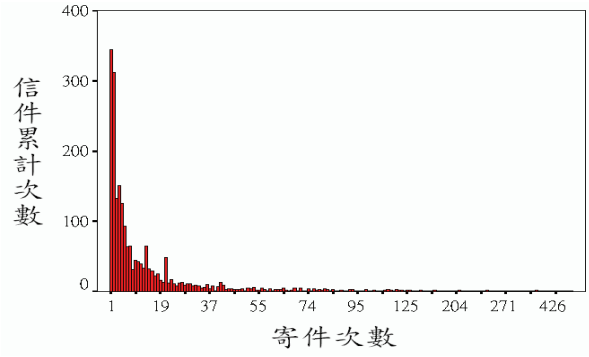

圖7：Enron電子郵件資料集寄件次數與信件累計次數統計圖

三、實驗結果分析

為了檢驗所提出的演算法執行結果的正確性，我們使用視覺化網絡分析軟體NetDraw 1.48，來進行資料的視覺化呈現。將151位員工以節點的方式呈現，並依據前述的統計方式，分別保留寄送電子郵件次數超過90及80的鏈結，以產生員工間寄送電子郵件的無向及有向關係相鄰矩陣，再進行自動化的資料鏈結分析，與關聯法則分析演算法的頻率分析。最後再綜合兩者所得結果，進行更進一步的解釋。

（一）無向網絡圖的自動化資料鏈結分析

資料以相鄰矩陣形式準備好後，依演算法步驟4尋找分支度為1的節點，並將連接此

節點之枝葉鏈結記錄至集合LL之中，之後將這些鏈結由圖形中刪除。

LL={ <v105,v112>, <v100,v111>, <v100,v104>, <v100,v102>, <v88,v24>, <v105,v112> , <v88,v24>, <v133,v86>, <v114,v110>, <v114,v52>, <v114,v96>, <v114,v63>, <v114,v81>, <v114,v95>, <v20,v30>, <v68,v3>, <v139,v140>, <v11,v12>, <v17,v41>, <v17,v21>, <v17,v10>, <v17,v25>, <v17,v26>, <v17,v71> , <v61,v125>, <v17,v40>, <v17,v59>, <v17,v56>, <v76,v142>, <v150,v151>, <v107,v78>, <v107,v92>, <v107,v54>, <v107,v126>, <v107,v122> }，如圖8之黑色鏈結所示。

在去除枝葉鏈結之後，依演算法步驟5～7，此次我們選取節點編號最小者（節點1）為起點，尋找接合點集合AP與二元連通元件(節點數大於二者) BC_i，結果如下：

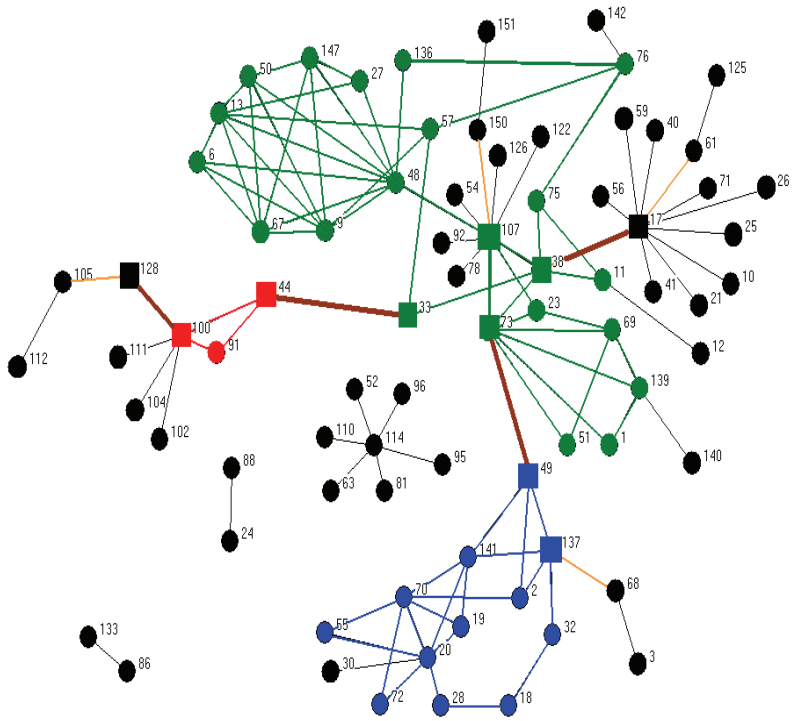

圖8：員工間郵件寄送關係分析結果之無向網絡圖

AP={ v128, v100, v44, v33, v107, v38, v17, v73, v49, v137 }，如圖8之方形節點所示。

(1)BC_1={ v6, v67, v9, v48, v27, v147, v50, v13, v136, v76, v57, v33, v75, v38, v11, v73, v23, v69, v51, v1, v139, v107 }，如圖8之綠色節點與鏈結所組成。

(2)BC_2={ v49, v141, v137, v2, v70, v19, v55, v20, v72, v28, v18, v32 }，如圖8之深藍色節點與鏈結所組成。

(3)BC_3={ v44, v100, v91 }，如圖8之紅色節點與鏈結所組成。

依演算法步驟8及9，去除BC_1、BC_2、BC_3與集合LL所包含的鏈結後，所剩下的即為潛在弱鏈結集合。再依照定義，可得弱鏈結集合WL={ <v73, v49>, <v33, v44>, <v128, v100>, <v38, v17> }，如圖8之棕色鏈結，其中<v73, v49>連接BC_1與BC_2，<v33, v44>連接BC_1與BC_3，而<v128, v100>及<v38, v17>則各自連接兩個接合點。

依演算法步驟10及11，將二元連通元件視為虛擬節點後，來尋找關鍵弱鏈結路徑CP_j；在本例中找到CP_1=<v_{73}, v_{49}>，CP_2=<v_{33}, v_{44}>，皆為已發現之弱鏈結。再依演算法步驟12，剩下的即為重要性較低的潛在弱鏈結集合PWL={ <v_{105},v_{128}>, <v_{128},v_{100}>, <v_{107},v_{150}>, <v_{17},v_{38}>, <v_{17},v_{61}>, <v_{137},v_{68}> }，如圖8之橘色鏈結。

（二）有向網絡圖的自動化資料鏈結分析

首先依步驟4尋找分支度為1的節點，並將連接此節點之枝葉鏈結記錄至集合LL之中，之後將這些鏈結由圖形中刪除。

LL={ <v_{100},v_{111}>, <v_{100},v_{102}>, <v_{100},v_{104}>, <v_{100},v_{91}>, <v_{24},v_{88}>, <v_{105},v_{112}>, <v_{105},v_{128}>, <v_{128},v_{105}>, <v_{114},v_{81}>, <v_{114},v_{63}>, <v_{114},v_{96}>, <v_{114},v_{95}>, <v_{76},v_{142}>, <v_{107},v_{122}>, <v_{122}, v_{107}>, <v_{107},v_{103}>, <v_{107},v_{78}>, <v_{107},v_{92}>, <v_{107},v_{126}>, <v_{107},v_{54}>, <v_{151},v_{150}>, <v_{11},v_{12}>, <v_3,v_{68}>, <v_{68},v_3>, <v_{61},v_{125}>, <v_{125},v_{61}>, <v_{32},v_{18}>, <v_{18},v_{32}>, <v_{17},v_{26}>, <v_{17},v_{41}>, <v_{17},v_{59}>, <v_{17},v_{56}>, <v_{17},v_{40}>, <v_{17},v_{25}>, <v_{17},v_{21}>, <v_{17},v_{10}>, <v_{17},v_{71}> }，如圖9之黑色鏈結所示。

在去除枝葉鏈結之後，接下來依步驟5～7尋找強連通元件(節點數大於二者)CC_i如下：

1. CC_1={ v_9, v_{13}, v_{48}, v_{50}, v_{67}, v_{147} }，如圖9之深藍色節點與鏈結所組成。
2. CC_2={ v_2, v_{19}, v_{20}, v_{70}, v_{72} }，如圖9之棕色節點與鏈結所組成。
3. CC_3={ v_{23}, v_{69}, v_{73}, v_{139} }，如圖9之綠色節點與鏈結所組成。
4. CC_4={ v_{11}, v_{75}, v_{38} }，如圖9之紫色節點與鏈結所組成。

之後找尋類接合點，並記錄至集合PAP={ v_2, v_{17}, v_{20}, v_{32}, v_{33}, v_{38}, v_{44}, v_{48}, v_{61}, v_{67}, v_{70}, v_{73}, v_{75}, v_{107} }，如圖9之方形節點所示。

依步驟8及9去除CC_1、CC_2、CC_3、CC_4與集合LL中鏈結後，接下來則是找尋弱鏈結集合WL；及依步驟10來尋找關鍵弱鏈結路徑CP_j。可得結果如下：

WL={ <v_{17},v_{38}>, <v_{17},v_{61}>, <v_{107},v_{17}>, <v_{33},v_{38}>, <v_{44},v_{33}>, <v_{73},v_{38}>, <v_{107},v_{38}>, <v_{48},v_{107}>, <v_{73},v_{107}> }，如圖9之橘色鏈結。

1. 連接CC_3與CC_4：

CP_1={v_{73}, v_{38} }

CP_2={v_{73}, v_{107}, v_{38} }

CP_3={v_{73}, v_{107}, v_{17}, v_{38} }

2. 連接CC_1與CC_4：

CP_4={v_{48}, v_{107}, v_{38} }

CP_5={v_{48}, v_{107}, v_{17}, v_{38} }

在形成虛擬節點的同時，可找出局部資訊集中點集合LISink、局部資訊來源點集合LISource及其與CC_i相連接的鏈結記錄到集合$LISL_i$中，如圖9之灰色鏈結。但此例中僅能發現局部資集中訊點。

LISink={ v_{27}, v_{57}, v_{141}, v_{55}, v_{51}, v_1}

1. 對於在CC_1中

$LISL_1$={ <v_{147},v_{27}>, <v_{48},v_{27}>, <v_{13},v_{27}>, <v_{13},v_{57}>, <v_9,v_{57}> }

2. 對於在CC_2中

$LISL_2$={ <v_{20},v_{141}>, <v_{70},v_{141}>, <v_{19},v_{141}>, <v_{70},v_{55}>, <v_{20},v_{55}> }

3. 對於在CC_3中

$LISL_3$={ <v_{69},v_{51}>, <v_{73},v_{51}>, <v_{73},v_1>, <v_{139},v_1> }

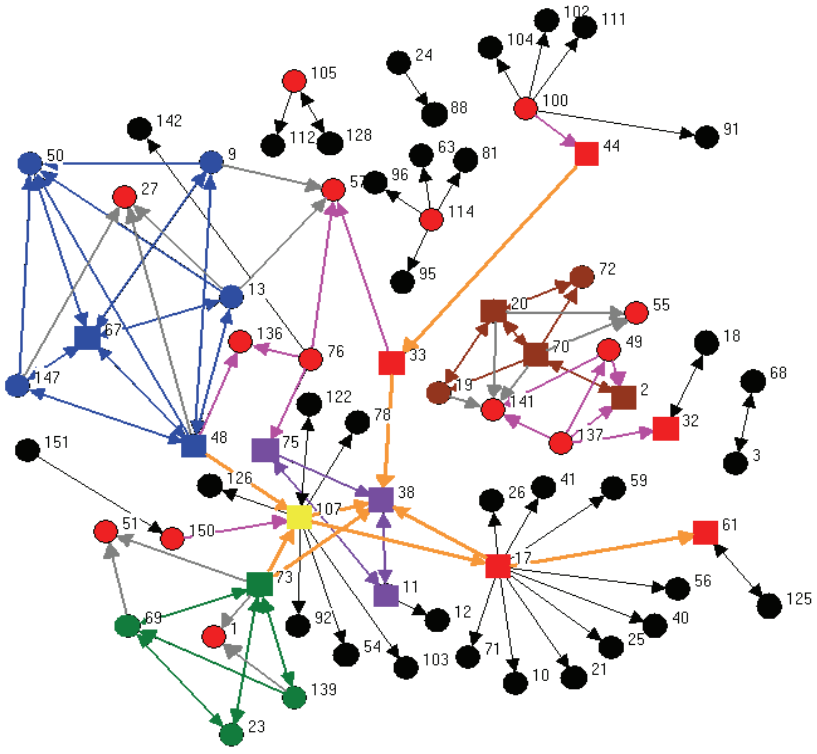

圖9：員工間郵件寄送關係分析結果有向網絡圖

將弱鏈結WL分離出後，剩下的即為重要性較低的潛在弱鏈結集合PWL如下：

PWL={ <v_{33},v_{57}>, <v_{49},v_{141}>, <v_{76},v_{57}>, <v_{76},v_{136}>, <v_{100},v_{44}>, <v_{137},v_{32}>, <v_{137},v_{49}>, <v_{137},v_{141}>, <v_{150},v_{107}>, <v_{137},v_2>, <v_{49},v_2>, <v_{48},v_{136}>, <v_{76},v_{75}> }，如圖9之粉紅色鏈結。

之後尋找可接收來自兩個以上不同強連通元件資訊的全域資訊集中點集合GISink，發現節點 v_{107}可直接接收來自CC_1與CC_3，如圖9之黃色節點。但此圖形資料中並未發現全域資訊來源點。

（三）關聯法則演算法分析

最後再透過關聯法則分析演算法(Agrawal et al. 1993)，以頻率分析出強鏈結關係。若有一規則A->B，其信度計算是將鏈結<vA,vB>的強度除以所有與vA相連的鏈結強度。結果發現若在無向圖中要找出所得的弱鏈結<v_{33},v_{44}>、<v_{73},v_{49}>、<v_{128}, v_{100}>與<v_{38}, v_{17}>，最小信度門檻值必需設得非常低(0.29 %)，才可以找出33->44 (31.78 %)、44->33 (43.61 %)、49->73 (2.86 %)、73->49 (0.29 %)、100->128 (3.65 %)、128->100 (5.0 %)、

17->38 (9.37 %)及38->17 (28.32 %)等規則。同樣地，若在有向圖中要找出所得的弱鏈結<v_{17},v_{38}>、<v_{17},v_{61}>、<v_{33},v_{38}>、<v_{73},v_{38}>、<v_{107},v_{38}>、<v_{48},v_{107}>與<v_{73},v_{107}>，最小信度門檻值亦須要設定得很低(5.60%)，才可以找出17->38 (36.51%)、17->61 (54.72%)、33->38 (10.37%)、73->38 (5.60%)、107->38 (6.63%)、48->107 (8.99%)與73->107 (39.33 %)等規則，但同時卻又會找出太多沒有意義的規則。由此可知，倘若僅靠頻率門檻值的設定，在現實世界中，是很容易遺失掉使用者所關心的資料法則。

四、實驗結果討論

（一）實際上本研究使用的圖論元件，皆可從社會網絡分析的觀點來進行解釋，如下所示。表二為重要特徵名詞在圖形理論、社會網絡和本研究上之定義對照表。

1. 連通元件：包含二元連通元件及強連通元件，如同人際網路中的叢聚團體，因為任兩個內部節點之間具有二條(含)以上的連通路徑，意即此二者具有某種程度的緊密性連接。
2. 接合點：即團體間溝通的重要行動者，在資訊傳播的行動中，是具有著關鍵性的角色。
3. 潛在弱鏈結：具有橋樑的功能，若將之移除會影響到資訊傳播的進行。
4. 弱鏈結：除具有與潛在弱鏈結相同的功能外，更強調連結的是兩個叢聚。若以公司組織而言，即為兩部門的對話或業務窗口。
5. 關鍵弱鏈結路徑：人際網路中兩團體間傳遞訊息的路徑。
6. 全域資訊集中點：可直接獲取兩個以上叢聚團體的資訊。因此可取得的訊息比起其它行動者更加豐富、多元。
7. 全域資訊來源點：可將訊息直接傳遞至兩個以上叢聚團體的行動者，對於資訊傳播有一定的效力，也可能具有較多影響他人的能力。

表2：重要特徵名詞在圖形理論、社會網絡和本研究上之定義

定義 / 項目	圖形理論	社會網絡	本研究
二元連通元件	任意兩個不同節點間存在至少兩條不同的簡單路徑(參閱p.7)	無明確定義，僅以叢集(clique)來表示一群行動者，基於彼此的相似性組合而成的次團體（參閱p.7）	與圖形理論之定義一樣，本文採用Weiss (1993)之演算法及程式片斷，並改寫之以找出二元連通元件(參閱p.8)
接合點	將接合點及與其相連之鏈結移除，會分割成更多的連通元件 (參閱p.7)	在不同團體間，扮演資訊傳播之關鍵性行動者	與圖形理論定義一樣，本文採用Weiss (1993)之演算法及程式片斷，並改寫之以找出接合點 (參閱p.8)

定義 / 項目	圖形理論	社會網絡	本研究
弱鏈結	無	弱鏈結扮演不同叢集(clique)間訊息交換的橋樑(參閱p.5)	參照社會網絡的觀念，本文自訂之操作型定義：某一鏈結的兩端為接合點或隸屬不同二元連通元件，即為弱鏈結(參閱p.12，步驟 9)
關鍵弱鏈結路徑	無	無	本文自訂之操作型定義：針對任兩二元連通元件及接合點之間，進行最短距離路徑搜尋，可得到一關鍵弱鏈結路徑(參閱p.12，步驟 10)
潛在弱鏈結	無	無	本文自訂之操作型定義：去除枝葉鏈結、連通元件內部鏈結、弱鏈結及關鍵弱鏈結路徑後之剩餘鏈結(p.12，步驟 12)
強鏈結	在有向圖中，定義任意二個不同節點u和v皆可互達時，稱為強連通(strongly connected)；也就是說，u->v和v->u	強鏈結的衡量計有「互動時間」較長、「情感強度」較強、「親密程度」較深或「互惠行動」較多(p.7)	本文參照社會網絡「互惠行動較多」的觀念，自訂之操作型定義：以關聯法則分析演算法求得大二項集合（即頻率高於最小支持度者）後，此集合中兩兩項目間的關係，即為加權圖中的強鏈結(參閱p.10，圖2)。
全域資訊集中點	無	可接受來自二個以上叢聚團體訊息之行動者，可能扮演資訊傳達之重要角色	參照社會網絡的觀念，本文自訂之操作型定義：具有兩個以上直接來自不同連通元件且皆為進入鏈結的節點(參閱p.13)
全域資訊來源點	無	可將訊息直接傳遞至兩個以上叢聚團體的行動者，對於資訊傳播有一定的效力，也可能具有較多影響他人的能力	參照社會網絡的觀念，本文自訂之操作型定義：具有兩個以上直接指向到不同連通元件且皆為輸出鏈結的節點(參閱p.13)

（二）為避免視覺化顯示時造成視覺混亂(visual clutter)的情況，在本實驗中仍設定了門檻值，亦即以小圖形(節點、鏈結)的資料來進行檢驗。但實際操作本自動化分析演算法時，使用者可依需求進行不同的門檻值設定，或是直接使用大量原始資料(不設定門檻值)，皆可以快速地得到結果。

（三）本研究著重在自動化的鏈結分析上，可快速分析出二元連通元件、強連通元件、接合點及弱鏈結等特徵，可省去人工耗費的大量時間。因所提出鏈結分析方法主體是使用二元連通元件演算法（同步找出接合點），為O(V+E)；及找出虛擬節點及接合點兩兩間最短距離路徑，為O(V3)。故本自動化的鏈結分析方法之時間複雜度為O(V3+E)。以本實驗圖形具有151個節點而言，在P4 3.0G電腦上執行的時間，針對110條無向鏈結來運算，僅需要約0.81秒便可分析出結果。若不設定門檻值，即針對1526條無向鏈結來運算，執行時間需要約0.98秒，但僅能發現一條枝葉鏈結(v135, v144)，及一個二元連通元件(除了枝葉節點v144，包含其它所有剩餘的150個節點)。而在有向網絡關係圖實驗中，即針對117條有向鏈結來運算，僅需要約1.95秒便可得出結果。若不設定門檻值，即針對2149條有向鏈結來運算，執行時間需要約4.56秒，但只發現一條枝葉鏈結<v135, v144>，及一個強連通元件。可知在效率可較先前視覺化處理快速許多，分析結果也更為客觀精確。

（四）本實驗中雖可利用自動化鏈結分析演算法，求得弱鏈結、潛在弱鏈結及關鍵弱鏈結路徑等重要鏈結關係，但仍需測試更多的資料庫，及有待社會學者或熟知組織關係者，進行更深入的探究。

伍、結論與後續研究建議

本研究提出一種將鏈結分析自動化的作法，主要是先去除枝葉鏈結，再基於圖學理論，使用深度優先方式尋找二元連通元件演算法及強連通元件演算法，將所找出的二元連通元件或強連通元件(節點數大於2者)鏈結從圖形中刪除後，即可找出存在於圖形結構中的重要橋樑成份-潛在弱鏈結(包含弱鏈結)及關鍵弱鏈結路徑。並佐以實驗檢驗的方式，利用真實的美國Enron企業電子郵件資料庫，配合視覺化的網絡分析工具，來檢驗本研究提出的自動化鏈結分析演算法。經由實驗結果，本演算法的確可以自動且快速有效地發掘出網絡圖中的叢聚，及在圖形結構上的弱鏈結橋樑關係，並依據其在結構上的重要性作一適當的分類，足以節省許多的人力支出。

一、研究結論

以下針對主要的研究成果進行說明及探討：

（一）探討對象為圖形拓樸特徵

不同於一般網頁超鏈結分析是在計算某網頁本身之重要性價值，亦可視為計算圖形中的節點及其權重。本研究則偏重於找出圖形連接的拓樸特徵，考慮的是圖形中的鏈結

位置的重要性，如弱鏈結、弱鏈結路徑及接合點等。

（二）避免以往偏重於頻率分析，使得隱性高價值訊息無法呈現的問題

由於先前關聯法則分析研究主要著重在頻率門檻值的探討，但卻仍難以解決真實複雜世界多變的問題。本研究在加入圖形結構拓樸的考慮因素，與社會網絡分析研究中，對於弱鏈結重要性的探討之後，提出此一自動化鏈結分析演算法；並佐以Enron電子郵件資料集分析所得的結果，可發現本自動化演算法確實執行迅速並且有效。因為若欲以關聯法則分析演算法探勘出所得的弱鏈結關係，則最小信度門檻值需要設定在 0.29% 以下，但如此將會得到大量不重要的規則，而需要花費大量人力時間在後續的規則驗證之上。

（三）演算法分析的自動化，可減少人力的耗費

由於視覺化的資料探勘，需要耗費大量人力在鏈結資料的主觀分析判斷，而本研究提出的自動化鏈結分析演算法，可自動判斷出圖形中弱鏈結、重要節點與叢聚等特徵，將大大提升分析效率。以Enron電子郵件資料集為例，以P4 3.0G所需要的執行時間僅需1秒，便可以獲得到分析結果。

二、後續研究建議

由於研究時間與資源上的限制，研究過程中仍存有待改進之處：

（一）社會網絡分析研究的解釋

本演算法所得之分析結果，如弱鏈結、重要節點與叢聚現象，皆須進一步地以社會網絡分析研究的觀點來進行解釋及驗證。

（二）未來的擴展性

本研究著重在弱鏈結的探勘及分析，希望對於以往僅著重於頻率門檻值、而忽略結構性因素的作法能有所突破，並能廣泛地運用到更多領域的研究中。如以交通網路管理為例，可預先依據道路不同的屬性，如路寬、路面材質及速限等，或是透過即時路況通報所提供的資訊，如車禍、道路坍塌、交通壅塞、管制等，計算後給予此道路適當的權重值，之後可藉由設定不同的門檻值，若低於門檻值則將道路視為不連通，再經由本演算法的分析，便可以迅速地將關鍵道路及路徑提供給交通管理機關，以做為進一步的流量控管及道路維護的參考。本研究之成果將可提供相關研究，一個更為強大的鏈結分析工具。

參考文獻

1. 吳寶秀，民79，台灣製造業員工個人社會網絡分析，東海大學社會學研究所碩士論文。

2. 胡守仁譯，2002，連結：混沌、複雜之後，最具開創性的小世界理論，台北：天下文化。
3. 陳會安，2002，資料結構理論與實務，台北：學貫行銷。
4. 蕭新煌、龔宜君，1998，『東南亞台商與華人之商業網絡關係』，華商經貿，第三八一期：19-38頁。
5. Adriaans, P., and Zantinge, D. *Data Mining* (1st ed.), Addison-Wesley, New York, 1999.
6. Agrawal, R., Imielinski, T., and Swami, A. "Mining Association Rules between Sets of Items in Large Database," *Proceedings of the ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.
7. Berkeley "UC Berkeley Enron Email Analysis Project," 2006 (available online at http://bailando.sims.berkeley.edu/enron_email.html).
8. Berry, M. J. A., and Linoff, G. *Data Mining Techniques: For Marketing Sale and Customer Support* (1st ed.), John Wiley & Sons, California, 1997.
9. Berry, M. W., and Browne, M. "Email Surveillance Using Nonnegative Matrix Factorization," *Proceedings of Workshop on Link Analysis*, Counterterrorism and Security, Newport Beach, California, USA, 2005, pp. 45-54.
10. Biconnected "Biconnected Component Partition," 2003 (available online at http://csci.biola.edu/csci480spring03/biconnectedComponents.pdf).
11. Borodin, A., Roberts, G. O., Rosenthal, J. S., and Tsaparas, P. "Link Analysis Ranking: Algorithms, Theory and Experiments," *ACM Transactions on Internet Technology* (5:1), 2005, pp. 231-297.
12. Brin, S., and Page, L. "The Anatomy of Large-Scale Hypertextual Web Search Engine," *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, 1998, pp. 107-117.
13. Chapanond, A., Krishnamoorthy, M. S., and Bülent, Y. "Graph Theoretic and Spectral Analysis of Enron Email Data," *Proceedings of Workshop on Link Analysis*, Counterterrorism and Security, Newport Beach, California, USA, 2005, pp. 15-22.
14. Chung, F. L., and Lui, C. L. "A Post-analysis Framework for Mining Generalized Association Rules with Multiple Minimum Supports," *Workshop Notes of KDD'2000 Workshop on Post-Processing in Machine Learning and Data Mining*, Boston, MA, USA, 2000, pp. 9-14.
15. Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. *Introduction to Algorithms* (2nd ed.), 2001.
16. Diesner, J., and Carley, K. M. "Exploration of Communication Networks from the Enron Email Corpus," *Proceedings of Workshop on Link Analysis, Counterterrorism and Security*, Newport Beach, California, USA, 2005, pp. 3-14.
17. Diestel, R. *Graph Theory* (2nd ed.), Springer, 2000.
18. Duan, Y., Wang, J., Kam, M., and Canny, J. "A Secure Online Algorithm for Link Analysis

on Weighted Graph," *Proceedings of Workshop on Link Analysis, Counterterrorism and Security,* Newport Beach, California, USA, 2005, pp. 71-81.
19. Enron Dataset 2006 (available online at http://www.isi.edu/~adibi/Enron/Enron.htm).
20. Enron Email Dataset 2006 (available online at http://www-2.cs.cmu.edu/~enron/index.html).
21. Garton, L., Haythornthwaite, C., and Wellman, B. "Studying Online Social Networks," *Journal of Computer-Medicated Communication* (3:1), 1997, pp.124-132.
22. Granovetter, M. S. "The Strength of Weak Ties," *American Journal of Sociology* (78), 1973, pp. 1360-1380.
23. Henzinger, M. R. "Hyperlink Analysis for the Web," *IEEE Internet Computing* (5:1), 2001, pp. 45-50.
24. Keila, P. S., and Skillicorn, D. B. "Structure in the Enron Email Dataset," *Proceedings of Workshop on Link Analysis, Counterterrorism and Security*, Newport Beach, California, USA, 2005, pp. 55-64.
25. Knoke, D., and Kuklinski, J. H., *Network Analysis*, Sage Publications, California, 1982.
26. Lauw, H. W., Lim, E. P., Tan, T. T., and Pang, H. H. "Mining Social Network from Spatio-Temporal Events," *Proceedings of Workshop on Link Analysis, Counterterrorism and Security*, Newport Beach, California, USA, 2005, pp. 82-93.
27. Liu, B., Hsu, W., and Ma, Y. "Mining Association Rules with Multiple Minimum Supports," *Proceedings of the 1999 International Conference on Knowledge Discovery and Data Mining,* San Diego, CA, USA, 1999, pp. 337-341.
28. McCallum, A., Corrada-Emmanuel, A., and Wang, X. "The Author-Recipient-Topic Model for Topic and Role Discovery in Social Networks, with Application to Enron and Academic Email," *Proceedings of Workshop on Link Analysis, Counterterrorism and Security,* Newport Beach, California, USA, 2005, pp. 33-44.
29. Milgram, S. "The Small World Problem," *Psychology Today* (2), 1967, pp. 60-67.
30. Priebe, C. "Scan Statistics on Enron Graphs," *Proceedings of Workshop on Link Analysis, Counterterrorism and Security,* Newport Beach, California, USA, 2005, pp. 23-32.
31. Shetty, K., and Adibi, J. "Discovering Important Nodes through Graph Entropy the Case of Enron Email Database," *Proceedings of 3rd International Workshop on Link Discovery,* ACM Press, New York, 2005.
32. UCINet 2006 (available online at http://www.analytictech.com/downloaduc6.htm).
33. Valiente, G. *Algorithms on Trees and Graphs*, Springer, 2002.
34. Visual_complexity 2006 (available online at http://www.visualcomplexity.com/vc/project_details.cfm?id=39&index=39&domain).
35. Wasserman, S., and Faust, K. *Social Network Analysis: Methods and Application,* Cambridge University Press, New York, 1997.
36. Watts, D. J., and Strogatz, S. H. "Collective Dynamics of Small-World, *Networks* (393),

1998, pp. 440-442.
37. Weiss, M. A. *Data Structures and Algorithm Analysis in C, Addison-Wesley,* Boston, 1993.
38. Westphal, C., and Blaxton, T. *Data Mining Solutions, John Wiley & Sons, New York,* 1998.
39. Xu, J. J., and Chen, H. C. "Fighting Organized Crimes: Using Shortest-Path Algorithms to Identify Associations in Criminal Networks," *Decision Support Systems* (38), pp. 473-487.
40. Yun, H., Hwang, D., Ha, B., and Ryu, K. H. "Mining Association Rules on Significant Rare Data Using Relative Support," *The Journal of Systems and Software* (67:3), 2003, pp. 181–191.