

# 尋找虛擬社群中之意見領袖

洪智力

中原大學資管系副教授  
chihli@cycu.edu.tw

葉姵妘

中原大學資管所研究生  
mogufly@gmail.com

## 摘要

找尋口碑傳遞中的關鍵人物是很重要的，可以在網路中散播許多的資訊到其他節點上。而意見領袖就是關鍵的人物之一，其在網路中是屬於橋梁的溝通位置，在口碑傳播中扮演了重要的角色。過去的研究中，較多利用社會網絡分析網絡節點關係的研究相當的豐富，而其中的網絡中心性計算可分析出誰是網路中的重要角色，因此相當多的研究以此分析到的中心性當作為意見領袖。但意見領袖所發表的文章內容也是一個識別是否為意見領袖的重要標的，而過去的相關研究中發現，輔以文章內容加以判斷的不多，且若是單用社會網絡分析找到的節點還是有不足的地方，因此本研究結合了社會網絡分析與文字探勘技術，計算了節點的中心性與識別意見領袖文章的要素。以期在社會網絡中尋找到真正關鍵的意見領袖。

**關鍵詞：**意見領袖，社會網絡分析，文字探勘

## Abstract

Looking for the key person from the delivery of word-of-mouth is very important and it can spread much information to other nodes in the network. Opinion leader is one of the key person and belong to the bridge position. In the past researches, the most of social network analysis study the node's relationship. The centrality can be analyzed who is the important role in the social network. Therefore, many researches would use the centrality to calculate the opinion leader. But the opinion leader's article content is also a major target to identify the opinion leaders. We found the past research and using article content to judge is relatively minority. If only used social network analysis to find opinion leader have some shortcomings. For this reason, we combine social network analysis with text mining to analyze the opinion leader. To anticipate find the real key opinion leaders in social network.

**Keywords:** opinion leader, social network

analysis, text mining

## 1. 研究背景與動機

早期資訊交流以 Web1.0 為主，其概念是單向的資訊流通，使用者與使用者之間能進行的互動較少。然而，隨著時代的變遷與科技的進步，現今網路上 Web2.0 的概念已轉變為雙向的交流，人與人之間的關係變得密切。而雙向的交流也使得社群網絡或是社會網絡(social network)在網際網路上蓬勃發展，就像是現實生活中社區的概念一樣，每一個人代表一個節點，利用網際網路將之連結成一個大型的社群網絡。社群網絡是一個傳遞訊息、知識的媒介[22]。而虛擬社群也是屬於廣大的社群網絡中的應用，虛擬社群藉由網際網路，使得人們可以雙向地溝通、交流、參與[27]。虛擬社群人際間的溝通以互動為主體[23]，正符合了 Web2.0 以使用者為中心的精神。其中國內有名的虛擬社群像是 mobile01 論壇、無名小站、Google 底下的 YouTube 等例子都是虛擬社群平台的運用。

資訊時代的來臨，電子口碑(electronic word-of-mouth)的傳播成為行銷中較為普遍採用的方式，例如有許多消費者會分享資訊或是產品使用心得並發佈在網路上，因此在虛擬社群上的論壇中存在著許多口碑資訊[29]。在虛擬社群中，每個人都可以發佈自己的想法經驗在網路的平台上，但總是會發現一些人氣比較高的節點，這些有影響力節點可以引發他人的迴響，並可將有影響力的節點視為口碑資訊的傳遞者，而接受傳遞者的意見則可視為口碑接受者，口碑傳遞者的專業程度對口碑接收者的購買決策有正向影響 [4][18]。口碑傳遞者很有可能就是這個社群網絡中的意見領袖(opinion leader)，意見領袖的意見無論是正面資訊的影響或是負面資訊影響他人，對整個網路來說都有很高的影響力。意見領袖的意見在社群網絡中具有一定的代表性[31]，也是一個帶來新資訊的節點，因此意見領袖所發表的內容變得相當重要。例如：意見領袖在論壇中發表了一篇對於新手機使用後的開箱文，配合上豐富

生動的圖文說明，很容易就吸引了其他消費者對這隻手機的注意，甚至進而選擇購買這隻手機。因此，我們可以知道意見領袖在口碑的傳遞上是個重要且有影響力的角色，其在網路中傳遞出多元的聲音且具有自然群聚、快速散播與回饋的效用。因此找出意見領袖在口碑行銷上是一個重要的關鍵，除了是傳遞電子口碑的起點之外，也可以使得產品的廠商想要表達的訊息在最短的時間內藉由意見領袖傳達給其他網路上的使用者。

在虛擬社群上的口碑資訊相當的豐富，例如 mobile01 論壇中以 3C 產品與行動裝置的討論區就分成相當多的類別以提供網友間的討論，而個人的經驗與想法可以直接或間接影響的他人思想、感情、與行為[9]。在過去的研究中，偏重用社會網絡分析(social network analysis)對網絡節點的連線數加以判斷[26][41]，或是利用社會網絡的中心性去判別，而較多人去連結的節點就有較大的機率成為意見領袖。而虛擬社群也是屬於社會網絡的一環，因此本研究利用社會網絡的特性進行節點中心度的計算。但從過去的研究發現輔以文章內容或從語義判定意見領袖的方式較少；[6][31][36]，然而意見領袖的文章內容也是判斷的指標之一，可從文章內文去分析文章是否具有意見領袖的特質。

因此本研究除了利用社會網絡分析去分析虛擬社群中有影響力的網絡節點之外，結合了文字探勘的技術分析了意見領袖的文章，進而識別出虛擬社群中的意見領袖。

## 2. 文獻探討

此章節回顧過去在此領域上相關文獻與研究成果，並作為本研究發展之基底。本章節共分為五大部分進行探討回顧，第一部分回顧了使用虛擬社群的動機之文獻，第二部分為意見領袖的特質分析，第三部分為計算網路節點中心性的社會網絡分析(social network analysis, SNA)，第四部分為文字探勘技術：控制字彙、字詞比對、詞頻計算的相關方法進行探討，第五部分為相關研究之討論。

### 2.1 虛擬社群

網際網路的盛行，使得不同地區的人們經由電腦與網路的連結可以共同聯繫在一起，虛擬社群利用了網路組織成一個小型的虛擬社會，藉由網路使人們有共同的目標彼此互動，除了分享資訊也參與了社群的交流[2]。相當多的虛擬社群以知識分享為目的，當有一人發出問題尋求解答時，其他人也會貢獻他們的知識

並加以回應，這種互動的方式就會形成知識的交流[34]。虛擬社群的種類大致上可分為四類：興趣型(interest)、關係型(relationship)、奇幻型(fantasy)、交易型(transaction)[3]。興趣導向的社群，就像是許多人都是攝影的愛好者，那自然就會將喜好攝影興趣的人群聚在一起；關係導向的社群，著重於個人經驗的交換，如看完電影的心得或是使用商品後的感想等交流；奇幻導向的社群，可看作是現今很流行的線上遊戲中的互動模式。交易導向的社群則著重在商品的買與賣[23][42]。而隨著時代的變遷，現今的虛擬社群類別也更加的廣泛了。除此之外，也有企業利用虛擬社群來收集消費者的資訊，從中得知使用者對產品的感想[21]。

綜合相關文獻所述，將使用者使用虛擬社群的主要動機歸類為五大項：尋求專家的解答、知識的交流、情感的交流、買賣交易、收集消費者的情報。

王思峰與劉麗萍在 2007 年的研究中利用韋伯的 idea type 研究方法將虛擬社群中的活動分為商業行為(如：買賣交易、收集消費者情報)和社會性互動(如尋求解答、知識的交流、情感交流)兩大類，而使用者在虛擬社群中的活躍度又分為積極與不積極兩種。如表一，大多數的人都屬於外圍成員，主要的參與目的可能只是想獲得一些資訊或是發問一些問題，業務員在社群中活躍程度是相當積極的，但是因其可能出自於自身的商業利益，就很容易讓人對於他所說的話會呈現一種不甚信任的感覺，但若是出自於非利益相關人所呈現的意見分享上，如同表中的圈內同好，其表達的意見反而會讓人信服，由此可知意見領袖在虛擬社群中的動機就偏向屬於社會型互動，並且意見領袖在虛擬社群中的活躍程度也是相當積極的。

表 1 虛擬社群理想型行為分析

動機 活躍度	商業	社會性互動
積極	業務員	圈內同好
不積極	外圍成員	

資料來源：王思峰、劉麗萍，2007

### 2.2 意見領袖

意見領袖一詞來自傳播學的概念，由 Lazarsfeld 等人在 1957 年提出，其研究發現在

政治選舉中廣告宣傳的影響力不如意見領袖來得大，政治人物利用廣告宣傳僅可能強化原有選民的政治立場，而意見領袖的影響力卻有可能改變選民的行為與態度，進而改變選民政治立場。

意見領袖的定義其實相當的廣泛，意見領袖的特徵根據過去的研究，大致上可分為四種，生活經驗相當的豐富、知識涉獵廣泛、社交能力較強並交遊廣泛、樂於創新[15][32]。而意見領袖也不一定是專家學者，一般人在某些領域有所專精的並善於給與意見者，也可以是意見領袖。

在 Emanuel Rosen 在 2001 年所著作的《The Anatomy of Buzz》提到網路中樞(network hub)一詞，網路中樞與意見領袖的概念相似，也就是網路中樞的影響力會間接或是直接地影響到網路中的其他節點，此書提及的網路中樞有六種特性，分別是：1.先驅特性(ahead in adoption)，並非一定是最早發現新議題，但意見領袖通常會參與在新議題的前端、2.連結特性(connected)，意見領袖通常是一個有強而有力的節點，通常會有許多的連線經由意見領袖的節點而散布到其他的節點中、3.旅行家特性(travelers)，意見領袖會經由瀏覽其他的資訊再發表其知識與想法，就像是到各地去旅行將資訊帶回到網路上分享給大眾、4.渴求資訊特性(information hungry)，在自己擅長的領域中意見領袖會希望能得到更多的知識，也因此對於資訊的渴求度也較他人高、5.自由表達意見特性(vocal)，意見領袖在社群網路中是可以對某些主題暢所欲言、6.媒體曝光特性(exposed to the media)每個人都可以透過網際網路這個媒介去表達自己的想法，意見領袖因為是個受注目的節點，其媒體的曝光度也相對的高。

本研究根據過去文獻中所定義的意見領袖特性，歸納出可能的意見領袖文章的因素[25][28][33]：

- (1) 意見領袖是個有影響力的節點，是會被很多人所連結的。
- (2) 意見領袖是個受人注目的節點，因此其文章的注目度會相對的高，也可能會引發許多人的迴響。
- (3) 意見領袖的文章內容是豐富的，表示文章字數有一定的程度以上並可能輔以圖片加以說明。
- (4) 意見領袖的文章，有一定程度的新穎性並且會切合主題。

### 2.3 社會網絡分析

社會網絡分析的主要目標是從社會網路的潛在結構(latent structure)中，可分析節點與節點之間關係的強弱程度[37]，如：強連結(strong ties)、中介連結(intermediate ties)與弱連結(weak ties)等關係[17][19]，而虛擬社群可視為一種社會網絡，並且可以利用社會網絡分析(social network analysis)的觀點進行分析與探討[35]。從網路上所留下來社會活動的記錄，例如人際網絡連結資訊，便可以轉換成社會網絡分析的資料，可以用來測量並預測虛擬社群成員的可信任度或是中心性等特質[42]。並可運用圖形理論(graph theory)的概念對資料作進一步量化的分析與計算，將龐大的網絡以圖形的方式呈現，將可較容易了解社會網絡中的特性與關係。如表 2，節點 A 到 G 分別代表網路中的節點，以 A 節點為例，表格中的數字 0 表示沒有連結其他節點，數字 1 表示有連結其他的節點。可知節點 A 連結了 B、D、G 三個節點。社會網絡指標的計算與空間的繪圖，可使用應用軟體 Ucinet (University of California of Irvine Network Programs)來分析社群網路所呈現的數據以及中心性[8]，在圖 1 中利用了 UCINET 軟體中 NETDRAW 繪圖功能呈現的表 2-2 節點與節點間連結的圖形。

表 2 節點關係矩陣

	A	B	C	D	E	F	G
A	0	1	0	1	0	0	1
B	1	0	1	1	1	0	1
C	0	1	0	1	0	1	1
D	1	1	0	0	1	1	1
E	0	1	0	1	0	0	0
F	0	0	1	1	0	0	1
G	1	1	1	1	0	1	0

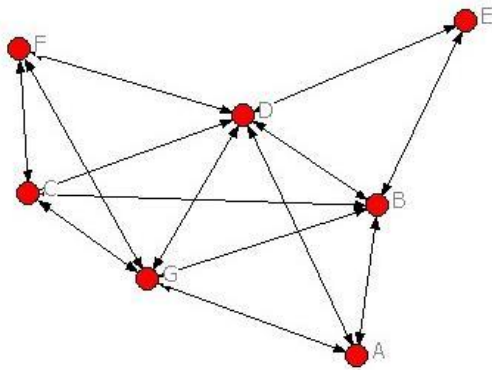


圖 1 模擬社會網絡圖

衡量社會網絡的分析指標大致可分為三種：社會網絡規模(size)、社會網絡密度(density)、網路中心性(centrality) [13][20][30][38]。社會網絡規模是指特定社會網絡中所有節點的數量，這裡節點指的是社會網絡中的成員，社會網絡的規模將是影響節點間關係的重要因素，因為節點彼此間關係的建立會受到資源多寡的限制，若是節點越多表示可用的資源可能也越多。社會網絡密度是用來表示節點間的關係是否緊密或是該社會網絡中節點之間的連結程度，密度越高代表節點與節點之間的關係越緊密。

網路中心性可用來衡量個人在網路中影響力的大小，而一般而言分析網路中心性指標可分為三種：程度中心性(degree centrality)、親近中心性(closeness centrality)、中介中心性(betweenness centrality) [16]。以下針對這三種測量網路影響力的計算做相關的說明：

- (1) 程度中心性(degree centrality)：利用相鄰節點的數量來衡量社會網絡的區域中心性，節點程度中心性值越高表示這個節點的影響力可能越大，就有可能成為這個網絡中的代表人物。以下圖為例，若要計算節點3之程度中心性，節點3相鄰了節點1,2,4，則節點3之計算方式為  $3/4-1=1$ ，表示節點3在這個範例的網絡中具有頗大的影響力

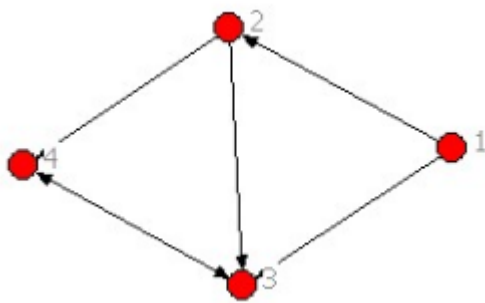


圖 2 範例網絡圖 1

$$C_D(n_i) = \frac{d(n_i)}{g-1}$$

標準化程度中心性公式

$n_i$ ：與欲計算節點相鄰的節點數

$d(n_i)$ ：表示  $n_i$  與其他節點的連帶關係

$g$ ：整個網路的節點數量

- (2) 接近中心性(closeness centrality)：主要是測量節點與其他節點的接近緊密程度，接近中心性值越高表示該節點影響其他節點的速度快也很強烈，而意見領袖對於其他節點的影響力也是相當快且深的。

$$C_c(n_i) = \frac{g-1}{[\sum_{j=1}^g d(n_i, n_j)]}$$

標準化接近中心性公式

$d(n_i, n_j)$ ：表示兩個節點間的距離

$g$ ：整個網路的節點數量

- (3) 仲介中心性(betweenness centrality)：衡量某一點存在於其他任兩點路徑上的重要程度。當仲介中心性的值越高時，表示該節點是位於溝通與橋梁的地位，也代表其在網絡中是個重要的節點，而意見領袖在網絡中就是扮演了這樣的角色。

$$C_B(n_i) = \frac{\sum_{j < k} g_{jk}(n_i) / g_{jk}}{[(g-1)(g-2)]}$$

標準化仲介中心性公式

$g_{jk}(n_i) / g_{jk}$ ：表示通過節點  $i$  連接節點  $j$  和節點  $k$  的最短路徑數/連接節點  $j$  和節點  $k$  的最短路徑數。

$g$ ：整個網路的節點數量

本研究主要這在部分主要是利用社會網絡分析探討有影響力的節點，程度中心性是利用相鄰的節點進行計算而未考慮到一些間接節點的連繫，找出來的意見領袖比較屬於區域性(local)的意見領袖。而接近中心性與仲介中心性會去計算網絡中每一個節點與節點之間的相關距離，找出來的意見領袖比較屬於全域性(global)的意見領袖 [6][7]，因本研究所測量的虛擬社群屬於不對稱型的社會網絡。因此本研究將利用仲介中心性來找出社會網絡中較具影響力的節點。

## 2.4 文字探勘

文字探勘(text mining)主要是針對半結構化或是非結構化的文件資料進行分析，透過文字探勘可以有效的來處理以文字模式存放的大

量資料，挖掘出藏在文字中的資訊與知識 [11][14]。虛擬社群上屬於文章類型的口碑資訊相當地多，因此從文字上去判斷此篇文章是否屬於會影響他人就成為了關鍵，文字探勘技術就可以去挖掘出內部可能隱含的訊息。

可能成為意見領袖文章的因素中，其與發表的內容與主題的切合度應具有一定程度的相關性。因此將使用詞頻(term frequency)計算來進行比對。詞頻(term frequency)計算為詞彙的出現頻率代表某一詞彙出現在一篇文件的次數[24]，計算公式如下：

$$TF_{ij} = n_j$$

#### 計算 tf 值公式

$n_j$ ：表示單字 j 的出現次數。

$TF_{ij}$ ：表示單字 j 在文件 i 的出現次數。

因此若是文章中出現的相關字詞越多的，表示越有可能是切合主題的文章。在詞頻計算之前，必須先建立控制字彙(Controlled vocabularies)的詞庫，因意見領袖屬於某些特定領域上的專家[12]，所以將比對範圍縮小，為特定領域進行關鍵字詞彙的蒐集至詞庫中，例如手機領域的關鍵字為：nokia、3G、費率、S60 系統等關鍵字，在進行比對上會較使用字詞頻率 (term frequency-inverse document frequency, tf-idf) 的方式找關鍵字再進行比對會更加的準確 [5][10]。

## 2.5 相關研究

過去對於網路上意見領袖的測量方法有用量化的方式，例如是使用票選的方式選出該論壇的意見領袖，此種方式利用人為主觀的評測，網友與專家們直接去選出大家心目中的意見領袖，接著再去深入探討其意見領袖的特質是較快也直接的方式[40]。

另一種量化的方式為利用社群網絡分析去分析發文者與回覆者的關係，建構出論壇的社群網絡，並對所建構的網絡進行特徵的分析，並發現其具備有小世界網路的特徵[41]，小世界網路是指網路中的任何一個人，只要幾個步驟就可以將其串連出某種關係。一般被稱為小世界的問題(Small-World Problem)，其演算法與數學模型在 1998 由 Watts 和 Strogatz 兩位學者提出。

也有研究是將在社群網絡中有影響力的部落格作者(Bloggers)分為四群，分別是活躍並且有影響力，活躍但無影響力，不活躍但有影響力，不活躍且無影響力。經由統計去定義主觀

性質的部落格影響力，提出一個初步的量化模型去測量有影響力的部落格作者，會藉由一些統計方法來進行蒐集部落格的發文數、影響力分數來進行分析[1]。

找出意見領袖也有利用一些演算法來輔助識別，Song 等人在 2007 年用影響力排名演算法(InfluenceRank algorithm)找出部落格中有影響力的節點並考慮節點的新穎性(information novelty)與重要性等因素，而新穎性是經由文章萃取內容後進行判定，經由演算法 InfluenceRank algorithm 計算出新穎性的結果，而新穎性也會是成為意見領袖文章的因素之一。或是利用 PageRank 演算法結合些許的情感分析形成一個新的演算法稱為 OpinionRank，此種方法近似於 PageRank 的概念，並找出最高排名的節點當作意見領袖，並利用 HowNet 為文章內的關鍵字給與意見上的分數[36]。

還有一類的研究是將文字探勘與社會網絡分析作一結合，在一網路論壇上顧客的意見與關聯利用社會網絡分析建立，將顧客的文章分成正面、負面與中立三類。利用文字探勘去萃取文章內容的特徵，並利用支援向量機(support vector machines)演算法去定義文章的屬性是屬於哪一類別，並建立文章與顧客間的關係，再利用社會網絡分析尋找中心性(centrality)程度高的節點當作意見領袖[6]。其找出的意見領袖是單純利用社會網絡分析之中心性去判別，進而再去探討其發表的意見趨勢是分屬於哪一種類。但意見領袖的形成，不單只有連線的因素，意見領袖文章影響的要素很多，單考慮一種有些略顯不足。

## 3. 研究流程

研究流程主要分為五大部分：第一部分為實驗資料的蒐集。第二部分為第一階段挑選意見領袖文章，利用社會網絡分析進行節點中心性之計算，再取中心性最高的部分文章當作第一階段挑選出來之意見領袖的文章。第三部分為第二階段挑選意見領袖文章，將文獻探討中之因素利用文字探勘技術將文章計算出個別權重，權重高者當作意見領袖的文章。第四部分為第三階段挑選意見領袖文章，結合了社會網絡分析與文字探勘技術，加總後權重值最高者當作意見領袖的文章。第四部分將分析出來之意見領袖文章，找到其個別的作者，分別將這三階段挑選出來之意見領袖與專家選出的意見領袖，加以評估找出意見領袖之準確率



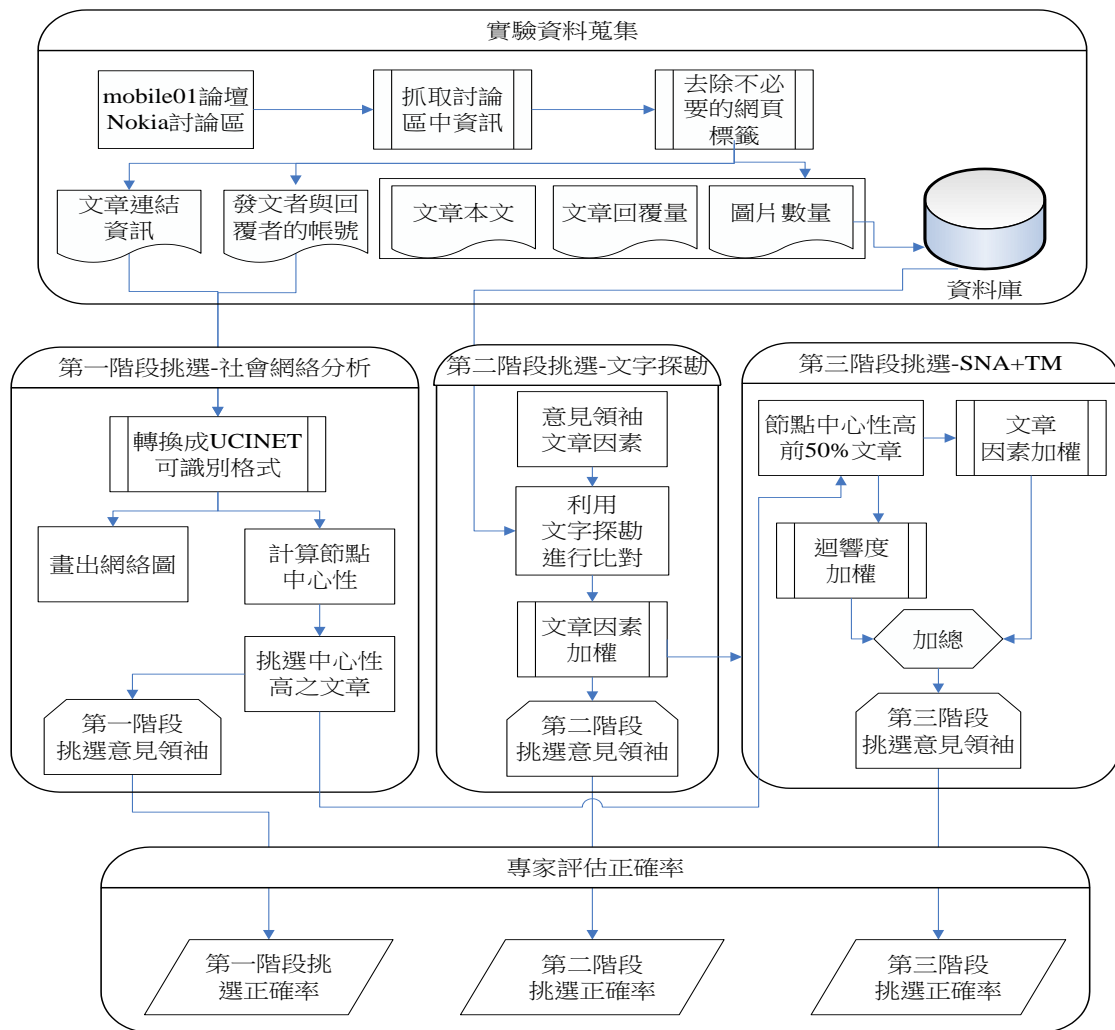


圖 3 研究流程圖

### 3.1 實驗資料蒐集

根據 i-buzz 網站在 2009 年的手機網路口碑資料統計，論壇類網站 mobile01 之行動電話與通訊討論區的口碑資料相當的豐富，如圖 3-2 所示。因此本研究的實驗資料從 Mobile01 網站中行動手機類別中挑選了 Nokia 討論區，並擷取實驗所需之資料。將取網頁內中的文章回覆連結資訊、發文者及回覆者之帳號、文章本文、文章的回覆數、文章字數、文章內的圖片。文章本文需要去除不必要的網頁標籤，例如 CSS 標籤等。最後將實驗所需要的資料儲存至資料庫中。本實驗蒐集了 600 篇文章，共 530 多名發文者，回覆筆數約為 5000 多筆。

### 3.2 第一階段文章挑選

第一階段利用文章回覆的連結資訊輸入至 UCINET 之軟體內，因資料數龐大，無法直接利用 UCINET 提供之介面輸入資料，因此利用程式將實驗資料的發文者與回覆者的關係轉換成 UCINET6 所需要的格式，其轉換格式分為兩種，一種是 node list，另一種為 edge list，因 edge list 的格式方式屬於連結權重的格式，如表 3-1，節點 1 與節點 2 中間的權重為 10，可看作節點 1 被節點 2 回覆了 10 次，因此權重為 10。此兩種方法皆可計算出節點的相同之中心性，因計算節點與節點之間的回覆數總和，須在送進 UCINET 分析之前作加總的動作，而本研究是直接將回覆的連結資訊記錄下來並直接送進 UCINET 分析，是較為快速的作法，因此使用 node list 格式，其格式如表 3-2 可看作節點 1 回覆了節點 2 與節點

的文章。

表 3 edge list 格式

dl n=4,format=edgelist1		
Labels embedded		
data:		
1	2	10
2	3	15
3	4	1
4	3	5

表 4 node list 格式

dl n=4,format=nodelist1		
Labels embedded		
data:		
1	2	3
2	3	4
3	4	
4	3	

接下來，利用 UCINET 計算節點的程度中心性並繪出該網絡之網絡圖，以表 3-2 的例子算出之仲介中心性如表 3-3，nBetweenness 為標準化後的仲介中心性，在此例子中節點 3 與節點 2 計算出來的仲介中心性值最高。圖 3-5 為其網絡之網絡圖。

本研究將根據節點中心性最高之節點，依序當作第一階段利用社會網絡分析挑選出來之意見領袖文章。

### 3.3 第二階段文章挑選

本研究在第二章文獻探討中探討了意見領袖的特質，歸納出四大項兩小類之意見領袖文章因素：主題相關性、文章新穎度、引發迴響、資訊豐富度又分為說明詳細與輔以圖片說明兩小類，其中本研究將引發迴響程度利用社會網絡分析進行計算，其他因素則利用文字探勘技術判斷。並從

資料庫中抓取文章本文、文章字數、文章內的圖片作為文字探勘的內容分析。圖 4 是文章因素內容分析之流程圖。

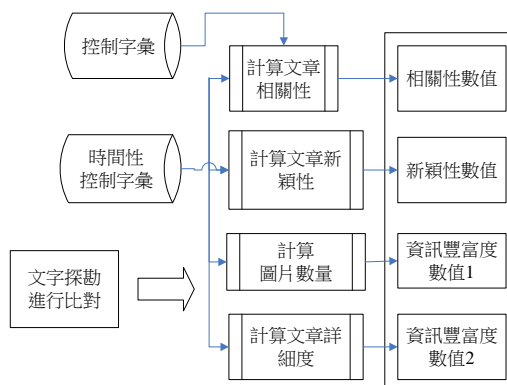


圖 4 文字探勘進行內容比對流程

#### 3.3.1 計算文章相關性

意見領袖所發表的文章，對於主題的相關度不會太低，而是會相當符合該主題的。因此，在文章中出現該主題字數愈多的就有可能愈符合該討論區的主題。計算文章相關性前，要先建立手機 nokia 的相關詞庫，作為文字與文章比對之控制字彙。本研究蒐集了手機 nokia 的相關辭彙，包含了型號、軟體、手機系統、可用服務...等等的詞彙，如圖 5 所示。

機型	軟體	手機系統	照相功能	電池	可用服務
X3	N-Gage	S40	200萬畫素	1000mAh	Oni
NOKIA	Nokia Photo browser	s50	300萬畫素	1320mAh	Google Voice Search
N97	Nokia Sports Tracker	S60	320萬畫素	860mAh	wi-fi
N96	Nokia Widget	S60 FP3.1	500萬像素	890 mAh	Widget
N95 8G	TiVi	Symbian	800萬畫素	1500MAH	口袋bb
N95	滑蓋助手	S60 5th	LED補光燈	BL-4CT	Traveler
N93	AGPS	Symbian3	卡爾蔡司鏡頭	1200mAh	qik
N91	PAPAGO	Maemo	自動對焦	1500mAh	Avatar
N900	Volp		氣氛閃光		Wall Calendar
N90	Handy Blacklist		補光燈		GPS
N86			蔡斯級		Dashboard
N85					Opera Mobile 10 BETA 2
N82					wifi
N81					Lifecasting

圖 5 建置手機詞庫

建立好控制字彙後，開始比對詞庫中的關鍵字在文章出現的次數。假設比對的結果發現文章一出現 nokia 兩次，n97 三次，500 萬畫素一次，S40 詞零次，如表 5，

表 5 計算相關性範例

	nokia	n97	500 萬畫素	S40
文章一	2	3	1	0

其相關性計算方式為：

$$\sum_{ij} f_{ij} = \frac{2+3+1+0}{3} = 2$$

其相關性值會因每個文章的值會隨著該文章的關鍵字愈多其值就愈高，因此本研究會根據所有計算出來之相關性的值，作標準化的動作，標準化後的相關性值設定為  $x_1$ ，而  $x_1$  會利用統計 Z 分數標準化至 0 到 1 之間。

### 3.3.2 計算文章新穎性

文章的新穎性跟文章所發表的時間點有很大的關係，內容若是提及比較新的產品或是服務就有可能是新穎性高的文章。本研究建置了以時間為基礎的控制字彙，利用 google trends 的功能，給予關鍵字一個時間性，例如手機 nokia 字彙中 N97 這個手機型號，在 google trends 中最早被人用關鍵字查詢的時間為 2008 年，就將 N97 註記為 2008 年的辭彙，以此類推其他字彙建立時間性的方式。

裝箱法(binomial method)可用來進行資料的區間分離[39]，本研究將利用等深裝箱法(equal-frequency-interval)將新穎性字彙分為三等級，分別是新穎性強、新穎性中、新穎性弱。並以年份作為切割點，2009 年以後出現的字彙屬於新穎性強，2008 年出現的字彙屬於新穎性中等，2007 年以前出現的字彙屬於新穎性較弱。當文章內的關鍵字屬於強中弱某一類別較多的就加乘權重，文章新穎性程度以  $x_2$  表示之。表 6 表示了每一等級所可加權的百分比。

表 6 新穎性權重百分比對照表

權重百分比	字彙新穎性等級
33%	新穎性弱
66%	新穎性中
100%	新穎性強

因此文章新穎程度  $x_2$  的權重公式為：

$$x_2 = \frac{(33\% * a_1) + (66\% * a_2) + (100\% * a_3)}{a_1 + a_2 + a_3}$$

$a_1$  表示出現新穎性弱的字彙在文章出現次數

$a_2$  表示出現新穎性中的字彙在文章出現次數

$a_3$  表示出現新穎性強的字彙在文章出現次數

### 3.3.3 計算資訊豐富度

本研究將資訊豐富度看作文章字數與文章內圖片數量的結合，因一篇文章要有豐富性，其字數一定不能太少，且能輔以圖片加以說明則其資訊豐富度就相對的高。在此研究中不考慮其文章發表者發文中所發佈的連結資訊，例如發文者所張貼之影片連結或新聞連結等，因抓取資料時其標籤容易與該網站的連結標籤搞混，造成判斷不易的情形，因此未將文章內之資訊連結納入資訊豐富度內。字數與圖片數量將利用相對比率，透過比率的統計和分析，瞭解數據與數據之間的關係。 $x_3$  表示說明詳細度的值， $x_4$  表示圖片數量的值，資訊豐富度  $x_5$  為  $x_3$  與  $x_4$  平均，其分別表示的公式如下，

$$x_3 = \frac{\text{每一篇文章字數值}}{\text{資料集內字數的最大值}}$$

$$x_4 = \frac{\text{每一篇文章有圖片的數值}}{\text{資料集內某篇文章內之圖片數量最大值}}$$

$$x_5 = \frac{(x_3 + x_4)}{2}$$

### 3.3.4 文章因素加權

將文章相關性  $x_1$ 、文章新穎性  $x_2$ 、資訊豐富度  $x_5$  加總之後，並依照權重值排序，權重最高者為第二階段利用文字探勘挑選出之意見領袖文章。

### 3.4 第三階段文章挑選

第三階段文章將結合社會網絡分析與文字探勘技術挑選意見領袖文章，根據第一階段節點中心性最高前百分之五十節點，抓取資料庫內節點程度中心性高之後選意見領袖文章資訊。並將社會網絡分析計算之仲介中心性的值當作迴響程度的權重，以代號  $x_6$  表示之。

綜合上述四大面向相關性、新穎性、豐富度、迴響度之加權值，可得到該篇文章之加權值。則意見領袖文章的總合加權公式則為，

$$\text{意見領袖加權} = \frac{(x_1 + x_2 + x_5 + x_6)}{\text{總篇數}}$$

### 3.5 評估方式

將隨機抽樣本研究判定之意見領袖與非意見領袖之文章，交與專家判斷其自身的答案，在與本研所得到的三階段意見領袖結果作比較，用 recall 值與 precision 值來評估正確率。並利用綜合 recall 和 precision 的 F 值，作為綜合評估的指標。



$$recall = \frac{\text{專家判定正確的意見領袖} \cap \text{樣本中本研究判定為意見領袖}}{\text{專家判定為正確意見領袖}}$$

$$precision = \frac{\text{專家判定正確的意見領袖} \cap \text{樣本中本研究判定為意見領袖}}{\text{樣本中本研究判定為意見領袖}}$$

$$F - measure = \frac{(2 * precision * recall)}{(precision + recall)}$$

#### 4. 結論

本研究提出了結合社會網絡分析與文字探勘的意見領袖文章分析，先利用社會網絡分析挑選了中心性較高的節點當作第一階段的意見領袖，並在文字探勘內容方面採用了文字量化與文章相關要素量化的方式計算出文章面向的權重值當作第二階段的意見領袖，並在第三階段混合社會網絡分析與文字探勘。以提高在虛擬社群中找到意見領袖之準確率。

#### 參考文獻

- [1] Agarwal, N., Liu, H., Tang, L., and Yu, P. S. Identifying the influential bloggers in a community. *In Proceedings of the international Conference on Web Search and Web Data Mining*. WSDM '08. ACM, New York, NY, 207-218, 2008.
- [2] Armstrong, A. G., and Hagel, J. The real value of on-line communities. *Harvard Business Review*, 134-141, 1996.
- [3] Armstrong, A. G., and Hagel, J.. Net Gain: Expanding Markets through Virtual Communities, *McKinsey & Company*, 1997.
- [4] Bansal, H. S., & Voyer, P. A.,. Word-of-Mouth Processes within a Services Purchase Decision Context. *Journal of Service Research*, 3(2), 166, 2000.
- [5] Berardini, T.Z., Mundodi, S. Reiser, L., Huala, E. & et al.,. Functional Annotation of the Arabidopsis Genome Using Controlled Vocabularies1. *Plant Physiology*, 135(2), 745-755, 2004.
- [6] Bodendorf, F., Kaiser, C.,. “Detecting opinion leaders and trend in online social networks” .*Social Web Search and Mining (SWSM2009)*. in conjunction with CIKM2009, Hong Kong, 2009.
- [7] Batagel, V.,. “Centrality in Social Networks” .<http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/CentNet.pdf>, 1996.
- [8] Borgatti, S.P., Everett, M.G., and Freeman, L.C. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard: Analytic Technologies, available in: <http://www.analytictech.com>, 2002
- [9] Cialdini, R.B., (2000). *Influence: Science and Practice*. Glenview, IL: Scott Foresman. 2000
- [10] Cimino, J.J., Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med*, 1998.
- [11] Deogun, J.S., Sever, H., Raghavan, V.V., “Structural Abstractions of Hypertext Documents for Web-based Retrieval”, *Proceeding of Ninth International Workshop on Database and Expert Systems Applications*, pp.385 -390. 1998
- [12] Engel, James F., Roger D. Blackwell, and Paul W. Miniard, *Consumer Behavior*. 6th ed. Chicago: Dryden. 1990
- [13] Faust, K., Centrality in affiliation networks. *Social Networks* 19: 157-191. 1997.
- [14] Fayyad, U., Uthurusamy, R., (1996). “Data Mining and Knowledge Discovery in Databases”, *Communications of the ACM*, 39(11), 24-26. 1996.
- [15] Flynn, L.R., Goldsmith, R.E., Eastman, J.K., “Opinion Leaders and Opinion Seekers: Two New Measurement Scales”, *Journal of the Academy of Marketing Science*, Vol. 24, No. 2, 137-147. 1996.
- [16] Freeman, L.C., “Centrality in networks: I. Conceptual clarification”. *Social Networks* 1, 215-239. 1979.
- [17] Garton, L., Haythornthwaite, C., & Wellman, B. Studying online social networks. *Journal of Computer-Mediated Communication*, 1997.

- [18] Gilly, M. C., Graham, O.L., Wolfinbarger, M. F. and Yale L.J., (1998). A Dyadic Study of Interpersonal Information Search. *Journal of the Academy of Marketing Science*. 26(2), 83-100.1998.
- [19] Granovetter, M., "The strength of weak ties". *American Journal of Sociology*, 78, 1360-80,1973.
- [20] Hanneman, R.A., & Riddle, M., "Introduction to social network methods". *Riverside, CA: University of California, Riverside.From*  
<http://faculty.ucr.edu/~hanneman/>.2005
- [21] Huang, E., Hsu, M.H., Yen, Y.R., "Resourceful virtual communities and revisit intention", *The Proceeding of WSEAS Conferences -E-Activities'07*, Puerto De La Cruz, Tenerife, Canary Islands, Spain, December 14-16. 2007.
- [22] Kempe, D., Kleinberg, J., and Tardos, É. Maximizing the spread of influence through a social network. *In Proceedings of the Ninth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*. Washington, D.C., August 24 - 27.2003
- [23] Lee, F. S. L., Vogel, D. et al., "Virtual community informatics: what we know and what we need to know." *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference*.2002.
- [24] Luhn, H.P., The Automatic Creation of Literature Abstracts, *IBM Journal of Research and Development*, 2, 157-165, 1958.
- [25] Marlow, C., "Audience, structure and authority in the weblog community". *In International Communication Association Conference*, New Orleans, LA,<http://web.media.mit.edu/~cameron/cv/pubs/04-01.pdf>, 2004.
- [26] Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., and Bhattacharjee, B., "Measurement and analysis of online social networks". *In Proceedings of the 7th ACM SIGCOMM Conference on internet Measurement*, 2007.
- [27] Rheingold, H., "The virtual community", Ablex Publishing Corporation.
- [28] Rosen, E., (2001). *The Anatomy of Buzz: How to create word of mouth marketing*. 1st Edition, New York: Doubleday, 1993.
- [29] Schindler, R.M., Bickart, B., "Published Word of Mouth: Referable, Consumer-Generated Information on the Internet", *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World*, 35-61, 2004.
- [30] Scott, J., *Social Network Analysis: A Handbook*. London, UK: Sage, 2000.
- [31] Song, X., Chi, Y., Hino, K., and Tseng, B., "Identifying opinion leaders in the blogosphere". *In Proceedings of the Sixteenth ACM Conference on Conference on information and Knowledge Management*, 2007.
- [32] Valente, W.T., Davis, L.R., "Accelerating the Diffusion of Innovations Using Opinion Leaders" *The ANNALS of the American Academy of Political and Social Science*, Vol. 566, No. 1, 55-67, 1999.
- [33] Watts, D.J., Dodds, P.S., "Influentials, Networks, and Public Opinion Formation", *Journal of Consumer Research*. Vol. 34, No. 4, pp. 441-458, 2007.
- [34] Wasko, M.M., Faraj, S., 'It is what one does': why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems* 9, 155-173, 2000.
- [35] Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C., "Computer net-works as social networks: Collaborative work", *telework and virtual community. Annual Review of Sociology*, 22, 213-38, 1996.
- [36] Zhou, H., Zeng, D., Zhang, C., "Finding Leaders from Opinion Networks", *Intelligence and Security Informatics, 2009. ISI '09. IEEE International Conference*. 266-268, 2009.
- [37] 吳齊殷、莊庭瑞，〈超連結網絡分析：一項分析網路社會結構的新方法〉，《資訊社會研究》，第6卷，頁127-148，2004。
- [38] 唐芷菁、王俊程，以 SNA 社群觀點分析網路口耳相傳與產品銷售相關性之探索性研究—以手機產品為例。第十六屆國際資訊管理學術研討會，2005。
- [39] 曾憲雄，蔡秀滿，蘇東興，曾秋蓉，王慶堯，資料探勘，旗標出版社，台灣，2005。
- [40] 周裕瓊，網路世界中的意見領袖-以強國論壇“十大網友”為例，當代傳播第三期，2006。

[41] 高俊波，楊靜，在线论坛中的意见领袖分析，*電子科技大學學報*，36 卷 6 期，頁 1249-1252，2007。

[42] 王思峰，劉麗萍，虛擬社群中誰比較可以信任？，*中華心理學刊*，49 卷 1 期，頁 1-17，2007