

國立臺灣大學文學院圖書資訊學系



碩士論文

Department of Library and Information Science

College of Liberal Arts

National Taiwan University

Master Thesis

以網絡書櫃資料建構讀者閱讀偏好多樣性之指標研究

The Estimation of aNobii User Reading Diversity

Using Book Co-ownership Network

柯逸凌

Yi-Ling Ke

指導教授：唐牧群 博士

Advisor: Muh-Chyun Tang, Ph.D.

中華民國 102 年 8 月

August, 2013



國立臺灣大學碩士學位論文
口試委員會審定書

以網路書櫃資料建構讀者閱讀偏好多樣性之指
標研究

The Estimation of aNobii User Reading Diversity
Using Book Co-ownership Network

本論文係柯逸凌君（學號 R98126021）在國立臺灣大學
圖書資訊學研究所完成之碩士學位論文，於民國一〇二年七
月十九日承下列考試委員審查通過及口試及格，特此證明

口試委員：

唐淑靜

（指導教授簽名）

林頌堅

吳怡瑾

系主任、所長

柯逸凌

（簽名）





本論文獲國立公共資訊圖書館博（碩）士論文研究獎助



謝辭



論文的完成，象徵碩士班學習的休止符，感覺就像是自己設計的房子完工，很有成就感而且心中充滿感激。

感謝唐牧群老師在研究中給予的幫助，您在我研究過程中熱心的指導與協助，讓我能夠放手去做屬於自己的研究，而您時常迸發的靈感也讓我的研究變得更精采可期，謝謝您。

感謝論文計畫書與學士論文口試委員林頌堅老師與吳怡瑾老師，你們給予的建議讓我的研究得以與實際應用接軌。

感謝國立台灣大學圖書資訊學系全體教授長達八年的指導，你們在各個領域帶給我的啟發與教導帶領我到達現在。

感謝我的研究室同學謝宜瑾與董采維，謝謝你們在每個過程中的熱心幫助。

感謝參與這個研究的 aNobii 讀者們，謝謝你們提供豐富的資料，讓我有機會做這樣有趣的研究。

感謝在研究過程中協助我的每一個人，台大電機所的唐立宇同學，台大資工所的彭維謙同學與台大資訊系的曹穎欣同學，謝謝你們在研究所需程式與資料運算上的實質幫助。

感謝我的男朋友聞浩凱，謝謝你在各種安全與危險的時刻，不論是生活或課業，一切因為有你而順利圓滿。

最後要感謝我的父母，感謝你們在學生生涯的照顧與包容，讓我能無後顧之憂的快樂學習。

柯逸凌 2013.08.16



摘要



網路愛書人社群 aNobii 的興起，提供了讀者閱讀活動記錄的資料，讓我們得以研究讀者的閱讀偏好，並因此提升系統的書籍推薦表現。過去研究（Ross, 1999; Tang et al., 2012）藉由了解讀者的閱讀偏好結構，分析讀者的閱讀活動。其中，閱讀偏好結構包含偏好發展、閱讀偏好多樣性與閱讀涉入性。閱讀偏好多樣性在過去研究中顯示，對推薦書籍策略有影響，相較於其他兩個閱讀偏好結構的面向，相較容易量化，因此本研究選擇閱讀偏好多樣性為目標，希望藉由網路書櫃資料的建置，找出能呈現讀者閱讀偏好多樣性的指標。

閱讀偏好多樣性的計算，我們抽樣 50 個網路書櫃，並建立書籍的共現網絡，計算書籍間的相似性。接著，在 5 種相似性計算下，選擇 3 種分群方式，針對抽樣的書櫃做分群，分群數量則作為讀者的閱讀偏好多樣性，分群數量愈多，閱讀偏好愈多樣，反之亦然。由於過去研究顯示作者為重要的選書因素（Mikkonen & Vakkari, 2012; Tang et al., 2012），我們同時建立作者共現網絡，選擇書目計量學中計算研究跨領域多樣性的指標，以作者在個別書櫃中的種類與其所占比例做為計算多樣性的依據。此外，利用作者共現網絡計算作者間的相似性，將作者相似性加入多樣性指標的計算。為檢定個別書櫃分群數量與作者多樣性指標是否能呈現讀者的閱讀偏好，研究針對 50 個抽樣書櫃的擁有者進行問卷訪談，訪談結果則做為評量閱讀偏好多樣性的標竿。

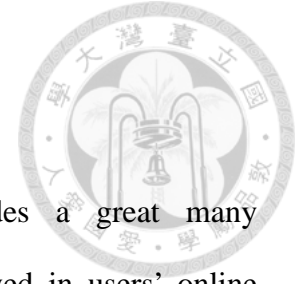
研究結果顯示，Interminus 相似性測量的分群結果與讀者自評多樣性呈顯著正相關，而作者多樣性指標結果則皆無顯著相關。Interminus 相似性測量可大幅去除書籍僅被一個書櫃所擁有，而造成書籍相似性被扭曲的結果，而其餘 4 種相似性測量之個別書櫃分群結果，除與使用者自評多樣性皆無顯著相關，反之卻與個別書櫃之書籍總數呈高度正相關。綜上述，一般認為讀

者閱讀偏好多樣性與閱讀書籍總數有關，然而，本研究閱讀偏好多樣性結果顯示，單從書籍總數無法判別讀者閱讀偏好多樣性，應進一步考量其閱讀書籍的相似性，才能準確地呈現讀者的閱讀偏好多樣性。



關鍵字：閱讀偏好多樣性、社會網絡分析、書籍共現網絡、作者共現網絡

Abstract



Usage data available through social media provides a great many opportunities to capture users' preference. Using books saved in users' online bookshelves, the study set out to explore social network analytical methods to capture the diversity of a reader's reading interests. "Reading diversity" denotes how widely scattered one's reading interests are. Drawing from data from aNobii, a social networking site for booklovers, users' reading diversity was defined by the number of components created by the book co-ownership network of the books in their bookshelves. A total of 50 user's bookshelf data were collected, resulting in a total of 21,199 distinctive books. They were also asked to fill out a questionnaire designed to elicit three dimensions of their preference: "reading diversity", "preference insight" and "involvement."

Networks of the books were created where each node represented a book and the strength of their linkages were determined by five co-ownership based similarity measures: cosine, correlation coefficient, "normalized interaction", and "intersection-minus-1." The thresholds for the dichotomization of the five respective similarity measures were then determined by a level above which where the greatest percentage of the disappearance of the edges, which were then applied to the individual bookshelves so the number of the components in each bookshelf could be determined. Correlation analyses were then performed between the user's self-assessed reading diversity and the number of the components in her/his bookshelf. One of the proposed similar measures, "intersection-minus-1" produced a clustering result that was significantly correlated with users' self-assessed diversity. Furthermore, multiple regression

analysis showed the proposed measure was able to provide explanatory power over and above mere counting the number of books in the bookshelf.



Keywords: Reading Preference Diversity, Social Network Analysis, Book Co-ownership Network, Author Co-ownership Network

目次



摘要	I
ABSTRACT	III
目次	V
表次	VI
圖次	VII
第一章 緒論	1
第一節 研究動機	1
第二節 研究目的與問題	4
第三節 名詞解釋	5
第二章 文獻分析	7
第一節 社會媒體與休閒閱讀	9
第二節 推薦策略與偏好結構	15
第三節 以網路分析法探討閱讀多樣性	21
第三章 研究方法與步驟	32
第一節 研究方法與架構	32
第二節 研究實施	36
第三節 研究範圍與限制	48
第四章 結果與討論	49
第一節 研究結果	49
第二節 綜合討論	66
第五章 結論與建議	70
第一節 結論	70
第二節 建議	71
第三節 進一步研究之建議	72



表次



表 1	推薦策略 (Senecal & Nantel, 2004).....	16
表 2	使用者偏好發展程度分組	19
表 3	顧客偏好發展與推薦策略(Kwon et al., 2009)	20
表 4	研究論文詞語高頻分析網絡分析閾值使用之結構特性(林頌堅, 2010)	28
表 5	selected measures of diversity (Rafols & Meyer, 2010)	30
表 6	資料收集與研究結果比較架構	35
表 7	偏好結構問卷之平均數與標準差	37
表 8	抽樣書櫃統計資料	49
表 9	Jaccard 最適閾值	50
表 10	Inter 最適閾值	51
表 11	Interminus 最適閾值	53
表 12	Cosine 最適閾值	55
表 13	Correlation 最適閾值	56
表 14	Region 分群敘述性統計	59
表 15	Region 分群相關性分析	59
表 16	Block 分群敘述性統計	60
表 17	Block 分群相關性分析	60
表 18	Girvan-Newman 最佳分群最小值分群數敘述性統計	62
表 19	Girvan-Newman 最佳分群最小值 Q 值敘述性統計	62
表 20	Girvan-Newman 分群最小值相關性分析	63
表 21	Girvan-Newman 最佳分群最大值分群數敘述性統計	63
表 22	Girvan-Newman 最佳分群最大值 Q 值敘述性統計	64
表 23	Girvan-Newman 分群最大值相關性分析	64



圖 次

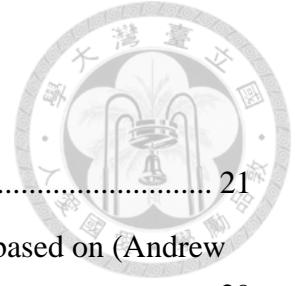


圖 1	本研究之研究架構	21
圖 2	Schematic representation of the attributes of diversity, based on (Andrew Stirling, 1998),p. 41)	29
圖 3	aNobii 讓使用者建議該書分類之畫面	34
圖 4	擷取抽樣書櫃中每一本書籍的所有書櫃擁有者，以「Whoops!!大債時代」一書為例，這本書有 1195 個書櫃擁有者將其加入書櫃	39
圖 5	book-owner 矩陣	39
圖 6	Region 分群演算法	42
圖 7	Block 分群演算法	43
圖 8	Girvan-Newman 分群演算法	44
圖 9	Jaccard 最適閾值	51
圖 10	Inter 最適閾值	52
圖 11	Interminus 最適閾值	54
圖 12	Cosine 最適閾值	56
圖 13	Correlation 最適閾值	57



第一章 緒論

第一節 研究動機



網路社群的盛行讓有類似興趣或嗜好的人們得以在其中進行交流或是進一步成立社團，以愛書人網路社群為例，國內有 aNobii、iReading 等，國外則有 Library Thing。這些社團可讓讀者記錄其擁有的書籍、評分或發表書評等，透過瀏覽閱讀偏好類似的其他讀者書櫃，或藉由社群網友推薦類似的書籍，愛書人可以從自身的閱讀出發尋找延伸閱讀的目標。網路讀者社群創造了愛書人資訊偶遇的機會，讀者可藉由社團中書與書、人與人、人與書之間的連結作為導覽


(navigation) 途徑進行延伸閱讀。網路社群內讀者閱讀活動的紀錄連結可做為了解閱讀活動的重要線索，並進一步成為閱讀推薦的重要參考資料。然而過去關於書籍推薦相關研究，多以改善系統預測準確度為目標做改進(Bell & Koren, 2007; Bennett & Lanning, 2007)，較少關注使用者間的互動情形以及使用者偏好結構對系統效能的影響，而在許多電子商務網站推薦系統相關研究指出(Kwon, Cho, & Park, 2009; Simonson, 2005)，使用者的偏好結構與推薦策略之間有交互關係。

過去閱讀偏好結構研究關注的面向有偏好發展程度(Simonson, 2005)、偏好異質性與閱讀涉入程度(Tang, Ting, & Sie, 2012)。其中，偏好發展程度為消費者對自己於某類型產品的了解與偏好的穩定程度(Franke, Keinz, & Steger, 2009; Kwon et al., 2009; Simonson, 2005)，如果應用於閱讀，所指則為讀者是否已清楚發展出自己偏好的閱讀類型，或是並未發展出特定偏好、不清楚自己喜歡閱讀的是什麼；除了自我了解與穩定度之外，本研究提出閱讀多樣性作為偏好結構的另一面向，多樣性指讀者偏好閱讀類型的多寡，讀者是否執著固定於少數閱讀文類，亦或不限於某一類作品或作者。我們認為閱讀偏好的多元或單一性可能影響書籍的推薦策略。閱讀偏好單一的讀者，由於其閱讀集中於固定文體、追隨固定作者，若給予新奇性較高的書籍，讀者的接受程度可能不高；而之若為閱讀偏好較多元

的讀者，通常較願意嘗試新事物，因此非顯而易見（non-obviousness）符合其閱讀偏好的書籍推薦可能更吸引該類讀者。

然而，過去關於使用者閱讀偏好結構之研究僅使用問卷調查作為研究方法 (Tang et al., 2012)，若欲對休閒閱讀讀者之閱讀偏好結構多樣性進行一般性之調查，因讀者數目過多，實不適合採用使用者調查研究。回顧使用者偏好之研究，以音樂、電影為主題之研究，新近大量利用社會網絡分析，從使用者資料分析使用者的興趣偏好(Bell & Koren, 2007; Bennett & Lanning, 2007; Buldú, Cano, Koppenberger, Almendral, & Boccaletti, 2007; Tang et al., 2012)。本研究嘗試以社會網絡分析方法探索讀者偏好結構多樣性的指標。因考慮閱讀偏好結構的眾多面向中如涉入性、閱讀發展程度，多樣性為最容易被量化測量之指標，亦可能對閱讀推薦策略選擇有影響，因此本研究於閱讀偏好結構上，將以閱讀偏好多樣性為分析目標。

若欲分析讀者偏好結構之多樣性，需要藉由在資訊豐富的平台收集資料，並於其中探索適合作為分析讀者閱讀偏好結構多樣性之指標。aNobii 為結合科技性與社會性的個人書櫃管理平台，不但記錄了個別讀者的藏書，也記錄了人與人、人與書之間的關聯，且使用者眾，是目前最多華人使用的閱讀分享社會網絡。讀者可以在 aNobii 中建立自己的閱讀清單，並且能夠撰寫評論、給予閱畢書籍評分，更能進一步藉以網路書櫃作為媒介與其他讀者互動（如傳送訊息）、分享資訊（如邀請參加閱讀社團）。此外，在選書的過程中，讀者亦能參考資訊線索如書評、評分、系統推薦輔助其選書決策。於現今之資訊環境中，讀者之閱讀偏好結構多樣性為讀者選書之關鍵，若欲進行客觀、量化之閱讀偏好一般性研究，需大量收集讀者的閱讀記錄，因此本研究選擇 aNobii 作為探索閱讀偏好多樣性之研究場域。



本研究於 aNobii 利用網絡分析法中使用節點與連線的方式蒐集資料，本研究將網路書櫃的書籍作為網絡中的節點（node），書籍之間共同出現在同樣的書櫃作為連結（edge）的依據，共同出現的次數則做為連結的權重（weighted），以連結的強度作為書籍間的相似性（similarity）。接著，對個別書櫃利用 UCInet 之 Region、Block 與 Girvan Newman 進行個別書櫃的分群，最後以個別書櫃中的集群種類與數量作為書櫃多樣性（diversity）的判斷依據，以個別書櫃內集群數量做為判別多樣性之依據。

除使用網絡分析方法測量多樣性之外，亦可使用書目計量學分析方式。書目計量學中常使用多種多樣性指標如 Simpson's Diversity、Gini 與 Shannon's Entropy 測量學術跨領域的情形(Rafols & Meyer, 2010)；於閱讀偏好相關研究中，尚未存在普遍採用於測量多樣性之指標。書目計量學中藉由分析引用文獻之多樣性，測量期刊文章或作者研究跨領域的情形。而論文作者引用其他多篇不同領域論文與讀者閱讀多本不同類型書籍，兩種情形於形式上相似，因此有可供參考之處，故本研究嘗試類比書目計量學之多樣性至閱讀偏好中，採用書目計量學之多樣性指標作為測量閱讀偏好多樣性之依據。

綜言之，本研究欲從網路書櫃中書籍的分配、書籍網絡的集中與分群情形，並採用上述兩種方法，探討讀者的閱讀偏好多樣性，期未來能夠對書籍推薦與相關個人化服務有所助益。

第二節 研究目的與問題



本研究欲從讀者的網路書櫃出發，目的為利用網絡分析法與書目計量學多樣性指標分析呈現讀者的閱讀偏好多樣性。

一、本研究欲分析書籍間的關聯性，利用共現矩陣進行相似性測量（**similarity measurement**）後之相似性矩陣（**proximity matrix**）分析書籍間的關聯性，並進一步利用網絡分析法針對個別書櫃進行分群。接著，利用個別書櫃內分群數量作為判斷多樣性之依據，呈現不同書櫃（代表讀者偏好結構）多樣性。根據以上目的，本研究的研究問題如下：

（一）進行相似性測量之共現矩陣，呈現讀者閱讀偏好多樣性之最適閾值為何？

（二）選擇閾值後，經由分群之結果，是否能呈現讀者閱讀偏好多樣性？

二、學術文獻分類中常使用主題詞與作者作為分類的依據，在休閒閱讀的範疇內重要的分類方式則為作者。由個別書櫃作品集中於特定作者的程度判別多樣性，藉此呈現不同讀者偏好結構多樣性。

此外，過去判別多樣性僅以不同種類作為判別多樣性之依據，由於不同種類間亦有較相似與較相異種類的差別（**degree of difference**），因此本研究嘗試加入種類相似性（**disparity of similarity**），即以作者與作者間的相似程度，探索加入種類相似性之後的多樣性指標是否更能準確測量讀者閱讀偏好結構多樣性。

（一）作者作為多樣性分析依據，**Simpson's Diversity**、**Gini**、**Shannon's Entropy**三種多樣性指標是否能呈現讀者的閱讀偏好結構多樣性？

（二）**Stirling's Revised Diversity** 加入作者共現矩陣作為作者（類別）間距離作為多樣性指標，所得結果是否更接近讀者自我評量之多樣性指標？

第三節 名詞解釋



一、休閒閱讀

與一般常見資訊行為不同，閱讀的動機非起源於完成任務。讀者帶著自發性的閱讀渴求，依據情緒、情境、想獲得的閱讀經驗等尋找閱讀材料，進行閱讀行為。

二、社群媒體

社會媒體（Social Media）是人們能夠分享與交流意見、心得、見解、想法與觀點的平台。與過去傳統媒體不同處為社會媒體是大眾可參與發聲的媒體，對於媒體內容擁有更多選擇與編輯的權力，並能自行集結成閱聽社群。

三、網絡分析

於此本研究所指為社會網絡分析（Social Network Analysis）。社會網絡分析利用網路理論（network theory）來分析社會關係（social relations），常把關聯（relation）利用節點（node）和連線（tie）來分析。它是一種跨領域研究，結合多種概念如圖論、理論模擬、認知分析等了解不同的網絡特性。本研究嘗試用社會網絡分析的概念來分析 aNobii 使用者的書櫃。

四、偏好結構

由消費者行為領域的研究延伸而來的使用者變項，泛指人們於某類型產品之個人偏好特質，涵括個人偏好特質的各種不同層面。過去研究之閱讀偏好結構包括偏好異質性、偏好發展程度、產品涉入程度等三種面向。於本研究中僅採用偏好異質性（homogeneity）之概念，但與書目計量學中之多樣性指標結合，稱為多樣

性 (diversity) 描述閱讀偏好結構中的多樣性程度。



五、多樣性

多樣性是書目計量學中分析研究跨領域常用指標，用以測量學術研究跨領域的程度。於本研究中，多樣性 (diversity) 可由三個性質來考慮：種類 (variety)、平衡 (balance) 與差距 (disparity)。種類是指不同類別 (distinctive categories) 的數量，平衡為類別的分布或是平均程度，差距是指類別與類別中的差別。

第二章 文獻分析



與任務導向的搜尋行為不同，休閒閱讀讀者所追尋的是閱讀過程中能獲得的經驗，而非特定主題的知識。因此，讀者所處的環境或個人情緒等主觀因素可能導致閱讀經驗與選書因素的改變，故休閒閱讀讀者的資訊行為，和取用與學習某領域知識或解決特定問題等任務導向的資訊行為有相當大的差異。在過去休閒閱讀讀者的資訊環境，讀者常從自己的信賴來源如朋友處找到感興趣的書籍。另外也有其他管道，譬如讀者可在圖書館的 OPAC 查找瀏覽、或藉由瀏覽網路書店找到感興趣的書籍。而隨著社會媒體的興起，也有愈來愈多的讀者透過這個管道獲得新書的訊息，除提供更豐富的尋書管道，使用者在社會媒體上的活動資料，也成為分析使用者偏好的一個重要來源。現在的社會媒體就像是富含資訊的資訊生態系統，網際網路打破了距離的藩籬，數位化儲存的資料記錄了使用者的各種活動歷程，對資料進行適當分析後將可探勘出使用者的使用情形並提供多種應用，例如若能改善資訊系統的表現將能讓使用者能更容易互動或進行資訊交換，或是在了解使用者偏好後使用適當的策略進行推薦或行銷等。讀者可加入閱讀社群或在社群中藉由朋友（的書櫃）遇見喜愛的書籍，或在社群網站中書籍之間的關連找到「你可能感興趣的書」。讀者本身可能希望搜尋的過程能輕鬆方便，而圖書館員、書籍出版與銷售者，尤其是電子商務網站為了提供更好的服務與銷售，皆希望能在了解讀者在資訊環境的尋書過程後進行自動化分析而成為推薦系統，因此本研究希望能探索非干擾式的資料分析方法了解讀者的偏好結構，讓休閒閱讀的尋書推薦能夠更符合需要。

而讀者喜歡推薦書籍與否與其閱讀偏好與偏好結構有關，(Franke et al., 2009; Kari & Hartel, 2007; Kwon et al., 2009)其中偏好結構包含偏好穩定度或自我察覺度等因素影響(Simonson, 2005; 謝宜瑾, 2012)。近來於行銷研究的文獻中也有學者發現偏好結構對於個人化推薦工具效能的影響(Kwon et al., 2009)。本研究欲透



過休閒閱讀讀者常用的社會媒體觀察休閒閱讀的讀者閱讀偏好結構，研究結果期能在未來的休閒閱讀推薦系統發展上，對不同的閱讀偏好與偏好結構使用適當的策略，推薦使用者可能感興趣的書籍。基於上述目標，首先必須尋找合宜的社會媒體並選用適當的方法來進行分析。本研究選擇了網路書櫃服務 aNobii.com 作為分析讀者偏好的場域。關於讀者的偏好結構分析，本研究將採取兩種嘗試，第一是社會網絡分析，利用書櫃內書籍共現矩陣（co-occurrence matrix）建立書籍間連結，並使用正規化（Normalize）共現原始資料（raw）形成正規化後之共現矩陣（normalized co-occurrence matrix）與多種分析網絡結構（structure）之相似性測量（similarity measure）形成之相似性矩陣（proximity matrix），建立書籍連結的強弱關係，並根據關係將其分群（clustering）呈現該讀者之偏好結構。其二是使用書目計量學指標分析作者的多樣性，並利用作者相似矩陣考慮作者間的相似性以探究個別書櫃內的作品多元程度。最後本研究將上述兩結果和使用者自我評估（user self-assessment）的閱讀偏好多樣性做驗證，期望能找出最適之偏好結構分析方法，方能在未來使用非干擾式（non-intrusive）之資料收集分析使用者之休閒閱讀偏好結構，並對休閒閱讀推薦系統之推薦策略選擇、個人化推薦有所助益。以下就研究主題相關之社會媒體與休閒閱讀、推薦系統偏好結構與網絡分析法分析閱讀多樣性三面向分述。

第一節 社會媒體與休閒閱讀



一、休閒閱讀

傳統上，圖書資訊領域資訊行為的研究多著重於使用者解決特定問題（problem-solving）、執行特定任務（task）的資訊尋求與資訊利用，其中更以協助學術、教育環境中的資訊使用者為主。然而在讀者進行休閒閱讀活動的情境下，並無任務導向的檢索目標。(Xu, 2007)將非問題導向的資訊檢索與過去問題導向的研究做比較，研究目標為了解使用者在非問題導向的情境下，其資訊檢索的相關判斷是否有所不同。研究結果發現，使用者在非問題導向的情境下確實著重不同的相關判斷標準。此種非以完成特定任務為目標之資訊行為(Kari & Hartel, 2007; Ross, 1999)，在圖資領域也較少被討論。休閒閱讀與一般常見資訊行為不同，在休閒閱讀的過程中，讀者並非尋找某本明確的書籍，有時候只是一個單純的閱讀慾望，或欲獲得特定閱讀經驗等難以清楚描述的動機，驅使讀者產生閱讀需求。由於休閒閱讀需求起源目的性較薄弱，更由於休閒性的資訊需求應考慮使用者偏好及作品內容，讀者常需要結合多種不同的相關判斷方式作為選書標準(Pejtersen & Austin, 1983)，如主題、背景架構、作者目標、可及性等，此種資訊行為與任務導向的資訊行為有相當大的差異。

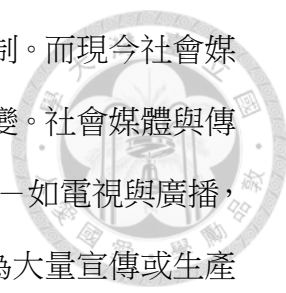
(Lancaster, 1998)提及休閒閱讀：「相較於一般在書目資料庫中使用主題搜尋以尋找期刊文章類的資訊，讀者在搜尋休閒作品時的判斷標準會更為個人化，也更為主觀而難以預測。」因此圖書館欲提供良好休閒作品的主題檢索有困難，也讓休閒閱讀讀者在尋找讀物時更倚靠非正式管道與人際網絡取得資訊(林珊如, 2001)。(Hopper, 2005)在青少年的休閒閱讀研究中指出朋友間口耳相傳對於青少年休閒讀物的選擇是一重要因素，而(Ross, 1999)也指出讀者通常仰賴其他資訊來源如朋友或評論來幫助他們發現和決定閱讀目標。此外，愛書人不僅會進行大

量的閱讀，亦會支持助長與拓展他人的閱讀經驗。因此提出閱讀是社會活動的觀點：我們會影響朋友的閱讀；而朋友的閱讀也會影響我們。於 aNobii 的長期研究結果更為社會網絡在閱讀上的社群面向提供了實證。(Aiello, Barrat, Cattuto, Ruffo, & Schifanella, 2010)此外，其中更發現，隨著書櫃間連結間的增加，書櫃之間的相似性也會增加。而使用者在與其他使用者互動的過程之中，能夠找到新的閱讀材料、參與新的閱讀社團，而這些活動更造成讀者閱讀檔案的變化，可能使得這些人的閱讀檔案更接近，而書櫃間的相似性會提高(Ross, 1999)。

(Ross, 1999)針對休閒閱讀的質性研究中整理了下列五點休閒閱讀讀者選書時著重的因素 1. 當下欲獲得之閱讀經驗 2.取得新書資訊之管道 3.書籍本身要素 4. 書籍提供讀者可能獲得之閱讀經驗的線索 5.書籍取得成本。綜合以上五點，休閒閱讀讀者重視的是在閱讀過程中能獲得的經驗，而非得到特定主題或問題的相關資訊。因此讀者當下所處的環境或個人情緒等因素可能不僅導致選書因素的改變，亦可能改變閱讀經驗。故作品本身的內容主題與其能帶給讀者的閱讀經驗並非絕對相關。因此面對讀者從事著重閱讀經驗的休閒閱讀而非對特定主題尋求資訊與知識時，要連結作品內容資訊與讀者閱讀後可能得到的感受顯得較為困難。(Tang et al., 2012)指出儘管閱讀通常被視為一項單獨存在的活動，而事實上它通常發生在社會關係網絡上，且現在可以透過社會媒體網絡連結大量拓展。因此下段將就人際網絡與資訊交換發生的重要場域—社會媒體進行回顧分析。

二、社會媒體


社會媒體 (Social Media) 是人們能夠分享與交流意見、心得、見解、想法與觀點的平台。現今主流社會媒體主要是基於網路或移動科技將溝通方式轉變為組織、個人或社群之間的互動對話。而社會媒體的發展，讓大量實體的社會活動可在社會媒體上實現，也創造了新的傳播工具及方法。一般而言，由於具有無遠弗屆的特性，社會媒體讓社會交流活動的進行變得更便利、豐富。因此社會媒體



可讓社會媒體上的資訊傳達變得容易，也更不受時間地點的限制。而現今社會媒體通常可夾帶多媒體資訊，使得資訊的傳達方式變得更豐富多變。社會媒體與傳統媒體的差異在於傳播結構、近用能力與專業要求，傳統的媒體一如電視與廣播，組織結構為中央集權；媒體的內容走向由業主掌握，目標通常為大量宣傳或生產銷售；而通常接受較高的專業要求。社群媒體則通常扁平無階層或較少階層，依照使用的需求而定；社群媒體通常免費或低價供社會大眾使用，如眾多網路部落格可免費申請；而社群媒體的內容發佈者通常不需專職，只需貢獻社群有需要的內容即可。其中社群媒體的重要轉捩點在 Web 2.0 觀念的興起，使用者可以上傳自己的數位內容與其他人分享。在 Web 2.0 後的社會媒體上，資訊發佈者與資訊接收者之間的角色可以隨時轉換，更多人可以參與社會媒體上資訊的發佈，讓社會媒體逐漸成為人與人之間重要而獨特的互動方式，社會媒體已成為資訊行為與交流的重要場域。

透過網際網路與社會網絡，人際關係的發展更可以不受地理空間的限制。如(Preece, 2001) 對網路社群的定義即為「在虛擬空間從事資訊交換、相互學習、相互支援或找尋同伴的一群人，無論其規模大小為何，而此社群可能為當地、國內或是國際性。因此在打破距離的限制後，透過網路上的社會媒體，不但可以聯絡因距離而分開的朋友，也可能認識實體生活圈外的同好。如(Armstrong & Hagel Iii, 1996) 從成員組成的參與需求做區分網路社群的類型中的興趣社群 (interest community)，(Armstrong & Hagel Iii, 1996)定義興趣社群為對某一主題有共同興趣或專長的人為討論該興趣或專長，而聚集起來之群體。如網路上喜好閱讀者可能會瀏覽閱讀相關討論區並在討論區發表，或是透過社會媒體記錄與分享他們擁有或閱讀的書籍，網路書櫃服務即是此例。透過網路書櫃的分享，網路書櫃的使用者可能有不同選書與閱讀行為(丁培涵, 2011)。

從之前對於休閒閱讀的討論可以得知，休閒作品分類囿於讀者需求的多層面複雜本質而有其困難之處(Pejtersen & Austin, 1983)，而書籍在排架上亦只能依照



最能反應讀者需求的小說文類主題（Genre）作為分類系統，但分類標準依據亦難以界定(Pejtersen & Austin, 1983)，由於傳統的主題式資訊檢索在面對休閒性資訊需求時所能達到的效能不高，相對於主題檢索人際網絡往往成為讀者取得閱讀資訊來源的管道。(Dourish & Chalmers, 1994)提出了社會導覽(Social Navigation)的概念。社會導覽是指物品與物品間關連參考的資訊蒐集的導覽(瀏覽過程)是透過社會媒體社群中的成員行為而達成的。(Dieberger, Dourish, Höök, Resnick, & Wexelblat, 2000)擴大社會導覽的解釋，讓透過社群成員中的相互協助而來的導覽(瀏覽過程)亦是屬於社會導覽的範疇。如利用 **Del.ici.ous** 分享網路書籤即是一種對於網站的廣義社會導覽。而(Svensson, Höök, Laaksolahti, & Waern, 2001)則進一步區分直接社會導覽與非直接社會導覽。直接社會導覽需包含資訊需求者與提供者的對話。非直接社會導覽則是資訊需求者從提供者留下的訊息或線索先幫助使用者。不僅社會導覽在社會媒體中變得益加受歡迎，社會導覽配合社會媒體更發展出了社會過濾(social filtering)的導覽行為。社會過濾並不主要關注使用者喜歡什麼，而是去關注該使用者的朋友喜歡什麼。因此過去電子商務網站從使用者的決定(選擇)推斷其偏好，發展成可透過使用者的朋友或相似偏好者一起分析其偏好。

三、網路書櫃服務與 aNobii


網路書櫃服務是社會媒體應用在書籍及閱讀的一個實例，使用者不僅可以在網路上記錄他擁有的書籍，亦可發表書評或對書籍評分，更可能透過找尋相似的書櫃認識新的朋友或是找到感興趣的休閒閱讀圖書，這是傳統實體互動較難達到的互動方式，因此為本研究非常適合的場域。而目前網路上較知名的網路書櫃服務有 aNobii、BookArmy、douban(豆瓣)、LibraryThings 與 Selfari 等。

在網路書櫃服務中，本研究選擇了 aNobii 作為分析的資料來源，aNobii 為創立於香港的網路書櫃服務，但因網路無遠弗屆，主要使用群來自義大利、台灣、

西班牙和香港(Aiello et al., 2010)。aNobii.com 為台灣讀者使用率最高的網路書櫃，在(李書萍, 2012)的研究中將目前台灣常見的網路書籍管理與書評發表平台做調查，該調查中使用網路書籍管理（網路書櫃）中的使用者使用 aNobii 的比率約佔 83%，由此可見 aNobii 在台灣網路書櫃的代表性，會有如此高佔有率的原因是因其中文支援度佳、且書目資料更新快又豐富，幾乎不需自行輸入資料，也 aNobii 不僅提供的藏書登錄管理系統，亦有評分與書評撰寫的功能，可在同一網站下完成多項書及相關紀錄與筆記功能，因此廣受愛書人歡迎。

aNobii.com 的內容主要由個人檔案 (profile)、我的書櫃 (book collection) 與聯絡人 (relations) 3 種元素所組成，個人檔案可讓使用者選擇性的揭露性別、年齡、地點等個人資訊；我的書櫃為主要網路書櫃服務部分，使用者可以利用 ISBN 大量匯入個人書籍，管理個人閱讀進度、評分並利用標籤分類，並可整理發表個人閱讀心得等。聯絡人可設定朋友與鄰居書櫃，鄰居書櫃是指非朋友但關注的書櫃。個人書櫃為此平台之服務核心，使用者將欲閱讀、正在閱讀或已閱畢之書籍加入其網路書櫃，使用者在書櫃中點選某一書籍除可瀏覽書籍基本資料外，亦可看到有哪些書櫃擁有此書；而點選他人書櫃，可以看到該使用者擁有的書籍，亦可知道該書櫃與自己書櫃的「品味相似度」，品味相似度為 aNobii 提供兩書櫃是否相似之指標，但其計算方式為統計兩書櫃中共同擁有書籍的數量，然而此種方法過於簡單，未考慮書櫃數量大小差異等基礎因素，僅能作讀者參考基礎，未能準確真實描述書櫃間的相似程度。除了相似書櫃之外，使用者在進入首頁時讀者也可以看到朋友或其他 aNobii 使用者目前新加入的書籍或書評，讀者或可從其中挑選其可能會喜歡的書籍作為休閒閱讀目標。因此目前 aNobii 使用者的尋書流程是以書櫃為基礎，可能是看到其他使用者的新活動，或是瀏覽朋友或鄰居書櫃，亦或是從自己的書籍中看到其他同時擁有此書的陌生人，進入上述對象的書櫃後，瀏覽該書櫃可能感興趣的書籍。

四、小結



休閒閱讀的讀物尋找過程並非回答特定問題或滿足特定目的，而是受讀者心中的情境、人際網路及信賴來源等影響，因此較難用目錄服務如圖書館線上公用目錄（OPAC, Open Public Access Catalog）或是網路關鍵字搜尋找到喜好的休閒閱讀讀物，而往往仰賴朋友或同好的推薦；網際網路環境下發展的社會媒體，讓社會媒體有了無遠弗屆、多媒體、互動等特性。因此過去休閒閱讀尋書資料來源將不受實體生活環境中時間地點的限制，可以結交不同地區朋友或得知不同國家的專家意見。社會媒體應用在書籍及閱讀的一個實例即為網路書櫃服務，使用者在自己書櫃中瀏覽書籍基本資料、查閱擁有此書的書櫃；在他人書櫃查看該使用者擁有的書籍，讀者可透過以上方式尋找休閒閱讀目標讀物。因休閒閱讀與讀者偏好息息相關，而從朋友或他人書櫃中挑選可能喜好的書籍，廣義來說亦是一種推薦。綜言之，社會媒體對愛書人的影響可分為兩個面向：取用面與決策面。取用面方面愛書人藉由社會媒體獲得了直接（從朋友書櫃）與間接（藉由社會式過濾）的方式取得閱讀的材料；決策方面則受到社會媒體中資訊如書評的影響，兩面向關連複雜多樣，於下節回顧。

第二節 推薦策略與偏好結構



一、推薦策略

在生活中許多時候我們必須在欠缺足夠資訊或經驗下做出決策，此時往往會倚賴其他人的推薦，而推薦有許多種形式—可能是親朋好友的經驗與口碑，亦可能是報紙或名人的評論。書店中銷售排行榜或是新上架書籍陳列平台是熱門商品推薦及新奇商品推薦，詢問愛看書或電影的親朋好友最近是否有好看的新書或電影，或是針對特定的研究領域請教該領域的老師或教授列出書單則是專家推薦。

「熱門商品推薦」或「專家推薦」為某一種推薦策略(Recommendation Strategy)，即是為何推薦某一商品的依據或方法。推薦即是整合並轉換各種資料來源，使用不同推薦策略將資料變成可供使用者參考之資訊。傳統之推薦是銷售人員藉由詢問顧客需求，根據其需求給予顧客適當的建議。而推薦系統則是利用程式處理資料，從資料中嘗試分析使用者的偏好結構，進而瞭解使用者個人對產品偏好後根據推薦策略進行推薦。因採用程式處理資料，可同時處理大量的使用者需求。在網路購物平台中，基本的推薦策略如銷售排行等可以方便且即時的透過銷售程式統計，亦不需要額外的成本，因此幾乎所有的電子商務網站都有如銷售排行等基本的推薦策略。而隨著社會媒體與電子商務的發展，使用者的購物歷程與個人基本資料（如年齡、性別、居住區域等）能夠以數位化的形式儲存，根據這些歷程與個人資料，可以利用電腦分析計算後可採用更複雜或個人化的推薦策略來進行對不同消費者的推薦。(Senecal & Nantel, 2004)將推薦資訊來源區分個人資料與非個人資料，亦將推薦策略分為個人化與非個人化，形成四種不同面向，如下表。

Provided information Source	Personalized	Non-personalized
Personal	Personal source providing personalized information (e.g., “My sister says that this product is best for me.”);	Personal source providing non-personalized information (e.g., “A renowned expert says that this product is the best.”);
Impersonal	Impersonal source providing personalized information (e.g., “Based on my profile, the recommender system suggests this product.”);	Impersonal source providing non-personalized information (e.g., “According to <i>Consumer Reports</i> , this is the best product on the market.”).

表 1 推薦策略 (Senecal & Nantel, 2004)

(Simonson, 2005)指出若銷售者能進行適切的客製化推薦、推薦能貼近使用者的個人偏好，將能大幅增進銷售者與顧客的關係，發揮潛在的顧客價值，這也是各網站紛紛進行推薦策略研究的原因。此一現象也從各大購物網站，包括以書籍銷售出發的亞馬遜書店（Amazon.com）與線上影片出租的奈飛公司（Netflix.com）都紛紛引入推薦系統並投入推薦系統的研究可見一斑。

而現今推薦系統常使用的個人化推薦策略大致上可區分為內容式推薦（content-based）與協同過濾（collaborative filtering）。內容式推薦分析使用紀錄，比對使用者曾經使用之產品或服務推薦具相似性的商品，將使用者的使用的單一產品記錄視為偏好的象徵；協同過濾則是使用全部群體的資料集合，將使用者所使用的所有產品紀錄作為偏好的代表，從群體中找出相似的使用者集合，並對集合中使用者普遍擁有但推薦目標使用者未擁有的商品進行推薦。此種從使用者群中找出行為相似的使用者進行推薦，又稱作社會式過濾法。

過去休閒閱讀選書行為相關文獻，閱讀偏好是重要的選書影響因素，在產品涉入較少、閱讀經驗較不豐富的讀者，可能因對類型、作者或資訊來源較不了解

或確定，較難做出選書抉擇；而經驗豐富的讀者則有更高的閱讀動機，想要興趣的閱讀目標，因此也更常認為推薦系統是有價值的(Ross, 1999; Tang et al., 2012)。一般而言，對於推薦系統皆採用下列假設(Peppers, Rogers, & Rogers, 1999)：經由使用者給予的外顯或內隱式回饋，系統可以建立使用者之個人偏好，故獲取使用者偏好的效果與效率可決定推薦系統的能力。

因此，讀者在網路書櫃如 aNobii 找尋休閒作品時，若推薦系統與策略能參考使用者個別偏好結構與群體資料，推薦符合使用者需求的書籍，能在讀者尋找、判斷閱讀材料時產生助益。因此下小節將回顧關於偏好結構之文獻。

二、偏好結構

無論是要推薦消費者購買商品，或是廠商要設計產品或行銷，都需要知道消費者喜歡什麼，因此消費者的偏好結構是一項被廣泛研究的主題。在讀者從事休閒閱讀時亦同，讀者會根據他本身的閱讀偏好結構挑選喜歡或是可能喜歡的書籍來閱讀。有學者指出進行休閒閱讀與推薦相關的研究，可能必須關注消費者或讀者的偏好結構。在早期經濟學界對偏好結構的研究中通常假設使用者的偏好為固定不變的，並且使用者知道其偏好為何。推薦系統或銷售人員是逐漸了解該使用者已發展的固定偏好，而使用者會了解其偏好，且無偏好的發展過程，此種為穩定性偏好，(Simonson, 2005)認為人們面對熟悉、簡單、較直接經驗的狀況時通常有穩定的偏好；。這種情形與人們對經驗的存取有關，若是熟悉的情境，人們傾向依賴經驗法則進行判斷，不會仔細評估。

然而研究指出，使用者對自己的偏好或需求不一定清楚明確，在早期資訊系統中常需要中介者的角色（例如：圖書館員的參考服務）來輔助使用者釐清自己的需求並將其轉換至資訊系統中查詢(Taylor, 1962)。此外偏好或需求亦會受到環境或決策過程影響，如(Bettman, Luce, & Payne, 1998)認為偏好並非固定不變的狀

態，而是會經由刺激產收，或受到任務類型、時間各種因素的影響。消費者行為領域也對此做出研究，產品的選擇或偏好往往是與當下目標、環境與情境、個人特質與自我認知等等因素交互影響下的結果。因此，同一使用者在不同的環境與情境，或是不同使用者在相同的情境與環境下，其偏好與選擇皆會有不同。隨著對產品的了解或使用者對自己的了解增加，可能會做出不同的決定，因此產生了偏好建構：消費者的偏好是發展而來，而非過去所認為總是固定且清楚的。

(Bettman et al., 1998; Haubl & Murray, 2003; Kramer, 2007)以休閒閱讀領域為例，此一假設亦非常適用，讀者並非帶著明確的閱讀需求，而往往在接觸某類型作品後才找到該類型中自己最喜歡的作家或書籍，且在實際閱讀前其實並不確定該作家或某本書是否會讓他著迷。在休閒閱讀的情境下，商品（書籍）的價值是基於主觀的判斷，在使用者尚未親自使用、經驗性之前，都無法了解商品的價值，此種商品稱為經驗性商品。經驗性商品能滿足情緒感官經驗(Holbrook & Hirschman, 1982)，但無法用明確屬性和因素做比較。文化事業相關活動或商品為偏向經驗性商品，在使用商品之前，使用者較無法確認對此商品的評價。書籍亦是一種經驗性商品，每個人對書籍的價值看法有所不同，在尚未閱讀前也無法確認自己是否喜歡一本書。此種面對不確定性高之經驗性商品的歷程，由於難以藉由客觀條件或主觀判斷確認商品是否符合需求，傾向尋求他人的評價。

(Simonson, 2005)針對上述偏好變動的現象進行觀察並提出了偏好發展的理論。此理論挑戰過去偏好既存且不變的假設，亦即消費者可以判斷何種商品最能符合其需求。當消費者面臨了商品或是推薦商品時，往往因為受到主觀與情境因素影響其決策反應與結果：偏好發展程度、與銷售人員的信任關係、商品的展示形式等。針對偏好建構與偏好發展（preference development），(Simonson, 2005)將消費者依據偏好固定程度（stability）與自我認知程度（self-consciousness）兩面向分成四個區塊(見下表)。偏好固定程度意指消費者擁有固定的偏好的程度，

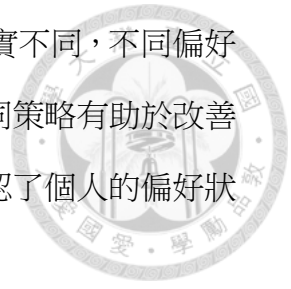
偏好發展愈完整偏好愈穩定。而偏好自我認知程度則是消費者清楚了解自身偏好的程度，包含了解偏好穩定的程度。若消費者偏好不穩定且缺乏偏好自我認知，雖看似因自我偏好不穩定而難以進行推薦，事實上卻因為此一類型消費者欠缺偏好自我認知，加上不穩定的偏好，反而是四種類型消費者中，最易接受推薦系統的一般推薦，亦可能受到影響而認知推薦系統所做的推薦符合其偏好，而感到滿足。下表為(Simonson, 2005)提出之使用者偏好發展程度分組：

偏好穩定 偏好自我認知	不穩定	穩定
模糊	組一	組三
清楚	組二	組四

表 2 使用者偏好發展程度分組

在(Simonson, 2005)提出此一偏好結構概念後，之後有許多研究依此概念試圖將此概念做實際驗證與應用。(Franke et al., 2009)針對偏好結構與客製化商品之效果進行研究。除了偏好自我認知程度外，(Franke et al., 2009)等加入了偏好表達能力與產品涉入程度兩因素之影響，綜合上述所有條件比較大眾化商品與客製化商品對消費者達成的滿意程度與願意購買金額 (WTP, Willing To Pay) 的影響。綜上所述，客製化產品推薦對於擁有下列特質的消費者最能發揮其效益：1. 清楚了解自我偏好 2. 能清楚表達自我偏好 3. 對於產品涉入程度 (involvement) 高。此一研究清楚地指出了使用者偏好結構與其特質確實對推薦系統與客製化商品帶來的效益有顯著的影響。(Kwon et al., 2009)則嘗試對不同組的使用者使用不同推薦系統策略，針對前述(Simonson, 2005)四組偏好固定程度與偏好自我認知了解程度組合分別採用不同推薦策略，使用之策略包括上節所述協同式過濾、內容式推薦、高分推薦與專家推薦，交互比較評估不同策略在不同偏好結構狀態下

的表現。研究結果發現不同推薦策略在不同偏好狀態下效果確實不同，不同偏好狀態可能分別適合不同的策略，證實使用不同偏好結構使用不同策略有助於改善推薦系統表現，亦即偏好結構屬性與推薦策略有交互作用，確認了個人的偏好狀態與推薦策略成效有關。



		Preference Stability	
		Unstable	Stable
		H1:Non-personalized recommendation	H2:Personalized recommendation
Insight into preference	Poor H3:User-collaborative recommendation	Segment 1 H5:Average opinion (non-personalized and user collaborative)	Segment 3 H7:Neighborhood-based CF(personalized and user collaborative)
	Good H4:Knowledge or expertise based recommendation	Segment 2 H6:Expert opinion (non-personalized and expertise-based)	Segment 4 H8:Content-based filtering (personalized and attribute-based)

表 3 顧客偏好發展與推薦策略(Kwon et al., 2009)

綜言之，使用者偏好可依不同觀點分析，本身亦由多種特質構成，又包含多個面向，因此在不同研究中的定義亦有不同。在過去社會媒體蓬勃發展前，在休閒閱讀之情境下，使用者僅能透過倚賴相同偏好朋友的推薦或推專家的書評等找尋與自己閱讀偏好結構類似之書籍；而透過新興社會媒體—如本研究之研究場域 aNobii 網路書櫃服務，使用者除能在此平台管理其閱讀歷程，亦能在此瀏覽不同書櫃，在書櫃間尋找符合閱讀品味之書籍，並能與品味相投之使用者分享閱讀經驗。因此本研究將以偏好發展觀點出發(Simonson, 2005)，期能藉由休閒閱讀社會媒體 aNobii 產生之休閒閱讀歷程記錄，觀察使用者的閱讀偏好結構，進一步對休閒閱讀推薦系統之策略選擇與個人化推薦有所助益。本研究嘗試用社會網絡分析方法來分析 aNobii 使用者社群的書櫃，而相關方法內容將於下節詳述。

第三節 以網路分析法探討閱讀多樣性

本研究以社會網絡分析方法作為資料收集之方法，藉由建立書籍於抽樣書櫃中共現之情形建立共現矩陣呈現讀者閱讀偏好結構。由於讀者之閱讀偏好多樣性與推薦策略有交互關係(Kwon et al., 2009; Simonson, 2005)，本研究以呈現讀者閱讀偏好之多樣性為研究目標，而判別多樣性前必須定義在閱讀偏好（讀者個人書櫃）中的種類為何，因此首先必須計算書籍間的相似性，方能依此分析閱讀偏好之多樣性。多樣性的概念早於書目計量學中跨領域研究即發展成熟，並發展出多樣性指標(Rafols & Meyer, 2010)；而近年來社會網絡分析方法則是分析關聯常用且熱門的方法(Buldú et al., 2007; 林頌堅, 2010)。本研究利用社會網絡分析方法與多樣性指標，探索呈現讀者閱讀偏好多樣性的方法，下述文獻分析之架構如圖所示。

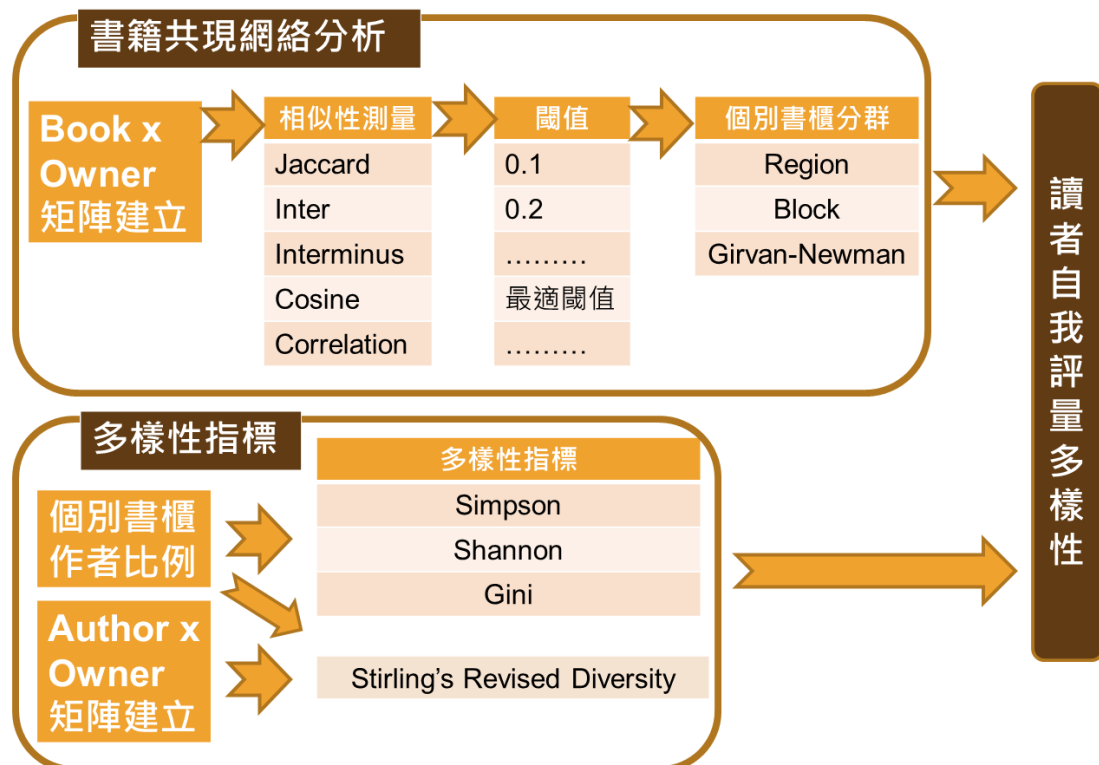


圖 1 本研究之研究架構

一、社會網絡分析與共現矩陣



社會網絡分析是一種跨領域研究，將數學與資訊工程領域上網路理論 (network theory) 應用在分析社會網路上的資料。通常將資料視為節點 (node)，而資料間的關係則用連結節點的邊 (edge) 來表示，邊的連結又可分為有方向性有向圖或無連結方向性的無向圖。除方向外，邊亦可能有權重 (weighted)，來表示連結關係的強度，若邊包含權重，則稱為 Weighted Graph，若無權重則為 Binominal Graph，Binominal Graph 亦可視為 Weighted Graph 的一種特例。建立網路後可用多種圖形理論指標如 Cardinality、Centrality 等研究圖形連結關係。或是將圖型資料已適當資料結構儲存並做進一步計算。如(Buldú et al., 2007)的音樂品味 (music taste) 研究。其以 The art of the Mix Project 的資料作為研究平台，從社會網絡的變化分析使用者的音樂品味。該研究蒐集 1998 年 1 月 22 日至 2005 年 6 月 4 日之間使用者播放清單中的歌曲，藉由分析共同出現在同一播放清單的歌曲探討音樂品味網路特性，該研究假設出現在同一歌單中的歌曲可能屬於同一種音樂品味 (music taste)，即使歌曲們可能分屬於不同的音樂類型 (music genres)，因此，常常出現在同一播放清單之 A 曲與 B 曲，若使用者將 A 曲加入播放清單，則 B 曲很有可能也是使用者喜愛的歌曲。故當兩首歌同時出現在一個播放清單，以歌曲作為節點就產生連結，而兩首歌同時被收在同一播放清單的次數代表兩個歌曲 (節點) 的關聯強度。此外，(Buldú et al., 2007)亦於研究中採用歌曲加入播放清單的時間 (input time)，藉此觀察音樂網路特色參數 (characteristic parameter) 在時間演進下如何變化。研究結果發現 92% 的歌曲會被包含在連結形成的巨大組件 (giant component) 中，亦即只有 8% 的歌曲與其他歌曲幾乎無連結，而熱門歌曲常常是中介歌曲 (bridge songs)，中介歌曲連結兩個獨立的播放清單網絡，可利用中介歌曲連結的播放清單進行撥放清單 (歌曲) 的推薦。

在書目計量學中，Small & Sweeney (1985)利用共同引用 (co-citation) 並分



群 (clustering) 來分析領域論文的引用 (citation) 做了整理與回顧，此一概念逐漸擴展到多種共同出現的情形，如文字共現 (co-word)、連結共現 (co-link)，因此類似概念現今用共現 (co-occurrence) 來包含之。共現網絡 (co-occurrence network) 則是將統計目標物共同出現於同一集合的情形製成圖形網路，通常以目標物為節點而共同出現情形次數為 **Weighted Edge**，如(林頌堅, 2010)利用共現網路計算圖書與資訊學刊論文的高頻詞語抽取與分析，將詞語視為節點，而將詞語共同出現在論文的次數為連結強度。共現矩陣 (co-occurrence matrix) 為一常用於共現網路的資料儲存與表示方式，每一節點視為一個欄 (Column)，而與其他節點或性質的連結情形使用該欄的不同列 (Row) 表示。若為 **Binominal Graph**，則連結與否常用 0 與 1 代表，若為 **Weighted Graph**，則用不同數值代表連結強度。如(Leydesdorff & Vaughan, 2006)將共現矩陣的使用進行了分析，文中提到在過往研究中如(H. Small & Sweeney, 1985)關注的是共同引用 (co-citation)，是將共同引用的發生視為該資料的性質 (attribute)，此時可使用正規化方法如 **Jaccard Index** 直接加以分析比較，此種方式關注兩筆資料性質的相似度，著重資料在區域間的表現。而在社會網絡分析的脈絡下，共同引用的資料可作為相似性之依據，因此資料可透過與其他資料間的連結關係比較相似度，著重兩資料在全體網絡間的相互關係，因此會將資料的所有連結關係透過常用的相似度計算如餘弦相似度 (cosine similarity) 或關聯性係數 (correlation) 做比較。

本研究將應用共現矩陣計算書籍在全體 aNobii 書櫃中共同被擁有的次數。並同時透過上述區域間與整體的概念進行計算，將在下一小節回顧。

二、結構等同與數值距離 (Structural Equivalence and Distance Measurement)



若兩書籍在全體書櫃中共現次數較高，則兩書可能較常被同一讀者選擇，若讀者具固定偏好，則兩本書應屬相似，但「相似」為一模糊之概念，在日常生活中會形容兩本書「很像」，即是相似的概念，但「很像」，可能是閱讀感受，由題材、敘事風格、段落或劇情安排等種種性質構成。多本書若依主觀判斷來界定相互相似性關係，則每個人意見可能均不同。因此首先需要將性質以數值記錄後，以結構分析的方式，使用不同計算方式計算出數值距離 (Distance Measurement) 來表現出相似性，再從多種相似性測量方式中選出較適合者。至於如何了解測量方式的適合，則會使用相似性測量結果加以分群並以使用者的自我認知加以驗證。

相似情形可從區域間結構 (local structure) 與書籍間整體結構 (global structure) 兩方面觀察。共現矩陣表示了書籍間的共現情形，書櫃的主人可能有某一種的偏好，因此同時放入了 A、B 兩本書籍，造成了 A、B 共現情形發生，若 A、B 共現情形發生次數多，則兩本書可能因皆符合某偏好的同一需求所以形成所有該偏好的讀者都同時擁有此兩本書，因此此兩本書可能較相似。此特性是觀察單一書櫃情形再累計總數，亦即注重區域間結構 (local Structure)。但此一方式，須注意書籍的受歡迎程度可能影響到書籍間相似性的計算。任意兩本受歡迎的書籍，可能因為同時都是暢銷書，即使分別屬於差異極大的類別，仍可能因為共同出現的次數大，而造成書籍相似性計算結果偏高。

因此在考慮相似度的數值計算時，需觀察了解計算中必要涉入但造成扭曲的因素與造成之影響，在計算結束前將結果做校正，此一過程為正規化 (normalize)。本研究將使用三種不同的正規化方式。方法一為 Jaccard Index 為統計學中常用於

比較兩集合相似度的方式，不直接使用共現數值，而是採用共現除以書籍總數和，亦即 $\frac{A \cap B}{A \cup B}$ ， A 、 B 為任意兩本書， $A \cap B$ 為書籍 A 與書籍 B 之共現次數， $A \cup B$ 為書籍 A 在所有書櫃中總數加上書籍 B 在所有書櫃中總數扣掉共現部分，進而避免 A 、 B 書籍受歡迎程度的差異造成相似性的扭曲。方法二採用共現次數除以兩本書籍中被持有次數較少者，亦即 $\frac{A \cap B}{\text{Min}(A, B)}$ ， $A \cap B$ 為書籍 A 與書籍 B 之共現次數， $\text{Min}(A, B)$ 為書籍 A 與書籍 B 之整體書櫃出現次數（被持有數）較小者，亦是避免 A 、 B 書籍受歡迎程度差異造成相似性的扭曲，但方法一著重兩本書全部出現的書櫃範圍與相互共現次數間的比率，方法二較注重較少被持有的書籍與共現次數間的比率，林頌堅(2010) 計算詞語間共現比重公式亦採用此一方式，但此方法有一問題為當兩本書 A 、 B 偶然在全部書櫃中共現一次時， $A \cap B$ 與 $\text{Min}(A, B)$ 均為 1，造成相似計算結果為本方法之結果的極大值 1，兩本書為相互最相似的書籍之一，但共現僅一次，代表兩本書都符合某一偏好的情形並不多，因此兩本書最相似之結果並不合理。因此本研究嘗試改良此一方法，將於第三章研究方法中討論。

另一觀點關注書籍在書櫃中相互關係的的整體結構 (global structure)，透過結構分析為在圖型中的節點與其他節點的關係尋找相同 (equivalence) 的結構 (structure) 或樣式 (pattern)，並以相同的結構與樣式為分析的基礎。結構性相同著重的概念為：定義兩標的物之間的關係並非觀察兩標的物之特徵值，而是觀察兩標的物在整體觀察母體中的結構相似性，亦即某個體於整體結構中的位置，亦可說是透過該個體與其他個體的關係定義該個體在群體中的角色。若以人與人之間的相似性舉例，觀察兩個人是否相似並非觀察兩人的背景、興趣等是否相似，而是觀察兩人在社交圈的聯集中（或是結構類似的社交圈子集中），在社交圈中所處的位置是否相似。常用在尋找圖型中相同社會角色 (social role) 中。常用的相同定義有 structure equivalence、automorphic equivalence、regular equivalence。

Structure equivalence 需為一節點關係上完全等同其他節點，在位置順序上必須可以相互取代。automorphic equivalence 則是一群 (set) 擁有相同連結 (ties) 的節點被區域結構包圍 (local structure)，亦即此兩群節點是平行結構。Regularly equivalent 是兩節點擁有相同的連結 (ties)，而同一群中的節點亦是 regularly equivalent。(Hanneman & Riddle, 2005) 上述之分析方式可用來尋找不同書櫃的巨大元件 (giant component) 中具有相同 (equivalence) 性質的節點 (書籍) 的特性。

而針對多種數值性質形成的組合，有多種常用相似性計算方式。首先最常見者為 Cosine Similarity, (Spertus, Sahami, & Buyukkokten, 2005) 及比較了多種相似性測量方式來分析 Orkut Social Network 內社群與使用者間相似性作為推薦社群之依據，而以 Cosine Similarity 的方式最好。Cosine Similarity 為將多種性質數值排列成一向量 (vector)。每一性質為一維度，而一個具有 N 個性質的性質組合將是一組 N 維向量。Cosine Similarity 的計算方式即為計算兩向量在此 N 維向量空間的角度差異。著重的是差異的累積大小，但避免在計算 Euler Distance 時須考慮正負數值，以及會易受到過大離群值影響的問題。

$$\text{Similarity}(A, B) = \cos \theta = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} = \frac{\sum_{i=1}^n \vec{A}_i \times \vec{B}_i}{\sqrt{\sum_{i=1}^n (\vec{A}_i)^2} \times \sqrt{\sum_{i=1}^n (\vec{B}_i)^2}}$$

其中 A, B 為行向量，且兩者維度須相同，n 為兩向量之維度。Similarity(A, B) 即為向量 A 與向量 B 之相似度。(Spertus et al., 2005) 比較不同的相似性指標在分析 Orkut 社會網絡的使用者社群相似性時的表現，就以 Cosine Similarity 為最佳。

另一常見之計算方式即為計算兩組數值間的相關性 (Correlation)，其計算方式等同於離散數值 (discrete) 的相關性係數計算。著重於兩數值組在變動時共同變化情形的大小。在計算兩兩相似性後，可以得知所有節點之間的根據數值距離計算結果得知的距離關係。此一距離關係可供偏好結構分析用。

$$\text{Correlation}(A, B) = \frac{\text{cov}(A, B)}{\sigma_A \sigma_B} = \frac{E[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B}$$

在評估相關性公式的計算效果時，(Klavans & Boyack, 2006)，提出了幾個可以觀察的面向：準確率(Accuracy)、包含率(Coverage)、規模可變性(Scalability)與強固性(Robustness)。準確率是指在一固定的相似性閾值或取固定數量的結果時，有多少結果是正確結果。而包含率則是同樣在一固定的相似性閾值或取固定數量的結果時，包含了正確結果的多少比率。上述兩概念相近於資訊檢索(Information Retrieval)的準確率(Precision)與召回率(Recall)。提升準確率與提升包含率有相互取捨之關係(trade-off)，若閾值取越高或數量取越少，只取前幾名則準確率可能很高，但包含率可能很少；反之，若取大量資料或降低閾值，則包含率可能提高但準確率可能會下降。規模可變性是指公式是否能同時套用在少量與大量的資料中。強固性指的是面對經過不同程度的維度降低轉換(Dimension Reduction)後是否能維持結果的一致。

三、閾值(Threshold)選擇與分群(Clustering)

在建立相似性關係後，若以書即為節點，書籍間的相似性為連結之距離或權重，可利用書籍間之相似性關係建立一圖型。此圖型可展現書籍間的相互關係，並可以此為基礎將書籍進行分群(clustering)，在(Henry Small, 2009)即是使用分群方式將多個領域的論文建立連結關係。因建立之連結圖形往往複雜且相互連結，難以分析，為了能進行分群，避免過多不相關資料影響分群結果與速度，需要訂定一閾值(Threshold)作為輔助。(Henry Small, 2009)即是利用一適當閾值，論文間相似性在此閾值之下視為無連結，在此之上為具有連結。而其中相互連結的論文會形成數個連結子圖(connected sub-graph)，並將其中最大者稱為巨大元件(giant component)，前述閾值即決定於巨大元件尺寸的急遽改變點，若某一閾值之下會讓巨大元件急遽擴張或尺寸改變速度增加，此閾值即為適當閾值。(林

頌堅, 2010)在論文詞語高頻分析的研究中，亦使用閾值來保留重要的連結關係並移除不必要的干擾部分，而該研究採用的閾值選擇考慮條件有連結線數、網絡密度、平均連結線點數、成分數與最大成分所佔比例。



表2 各種不同閾值所產生的連結線數目以及其他的結構特性

閾值	連結線數	網絡密度	平均連結節點數	成分數	最大成分所占比例
0.2	5849	8.95%	32.31	1	100.00%
0.3	1955	2.99%	10.80	2	99.72%
0.4	1044	1.60%	5.77	23	93.37%
0.5	487	0.75%	2.69	85	74.59%

表 4 研究論文詞語高頻分析網絡分析閾值使用之結構特性(林頌堅, 2010)

針對單一書櫃中書籍進行分群需利用閾值。本研究預計採用 UCInet 軟體針對個別書櫃進行階層式分群 (Hierarchical Clustering), (Hanneman & Riddle, 2005) 說明可使用此種方式進行社會網絡分析。階層式分群是將最相似之群集兩兩合併，不斷重複此一過程，直至所有群集合併成唯一群集。可透過檢視合併的歷程了解群集的結構關係，或是在合併過程中需選擇適當停止點以了解群集的分布情形。在考慮群集間相似時可能有不同計算方式，如 UCInet 有四種相似距離定義，包括 Single Link、Complete Link、WTD Average 與 Simple Average。Single Average 為兩群集間之最短距離為群集間距離，Complete Link 為兩群集集合中最大距離為群集距離，WTD Average 為群集點平均距離除以群集尺寸，而 Simple Average 則直接使用群集平均距離。

四、多樣性分析

目前休閒閱讀書籍並無客觀且廣泛的區分書籍種類方式，主題標目 (subject heading) 與書籍寫作風格不一定有關，一般圖書分類法休閒閱讀的書籍又可能大量被分在同一類，如武俠小說類就包含的大量的不同作者與類型風格，讀者往

往不會接受全部分類中的書籍。根據(謝宜瑾, 2012)，休閒閱讀讀者最常利用作者來尋書，因此利用作者區分書籍種類可能是合理的嘗試。而書櫃書書籍的瀏覽列表與論文文獻中的引用文獻有些許類似之處，因此除社會網絡分析外，本研究將應用書目計量學中的多樣性指標（citations）來測量個別書櫃的多樣性，對於單一書櫃或可用多樣性（diversity）指標針對其組成做分析。多樣性是書目計量學針對跨領域主題常用的分析方式，論文的主題通常屬於某一領域，但引用可能來自不同領域的資料，因此發展多樣性指標計算該論文跨領域的程度。而一讀者書櫃的書櫃亦可能包含不同書籍種類，因此可嘗試用以分析休閒閱讀之多樣性。根據前述分析，書籍種類在本研究中將以不同作者視為不同種類的方式作為多樣性分析中研究領域的對應。多樣性(diversity)可由三個性質來考慮:種類(variety)、平衡(balance)與差距(disparity)。種類是指不同類別(distinctive categories)的數量，愈多不同種類領域多樣性愈高。平衡是指類別的分布或是平均程度，若種類大小分布不均，則領域則主要由種類數量最大的領域為代表，多樣性較低；若種類大小分布平均，則代表同時且平均地具備多種領域的特質，多樣性較高。差距是指類別與類別中的差別。若種類間差別大，譬如以大學系所做為領域種類依據，同樣是文學院的系所，與電資學院語文學院的系所，由於不同學院所學差距較同學院大，後者的種類差別較大。(Purvis & Hector, 2000; Andrew Stirling, 1998; Andy Stirling, 2007)因此，以上三個性質的增加皆會增加其多樣性。

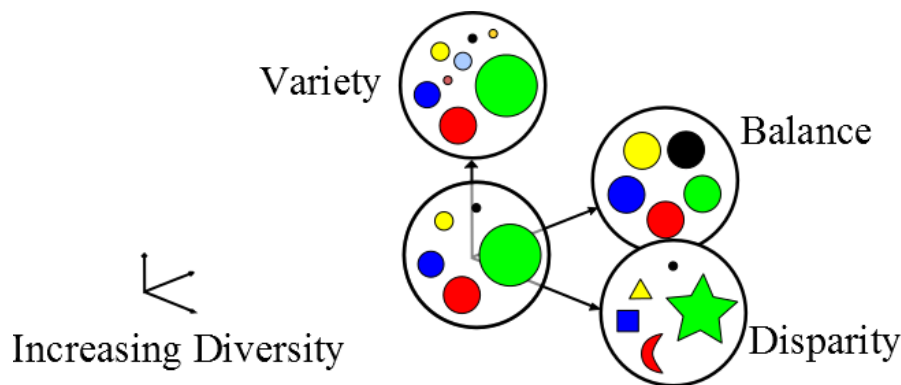


圖 2 Schematic representation of the attributes of diversity, based on (Andrew Stirling, 1998),p. 41)


(Rafols & Meyer, 2010)展示了常見的多樣性衡量指數，如 Simpson's 或 Simpson's 同時衡量了 variety 和 balance，但無法去衡量類別中的相似度或是距離，因此他提到(Andy Stirling, 2007)提出了更一般化的多樣性指標，本研究可以利用此一計算方式計算單一書櫃中作者的多樣性。常見之衡量多樣性之指標另有吉尼係數 (Gini) 與熵 (Entropy)。吉尼係數是義大利統計學家 Corrado Gini 所設計來計算數值散布程度 (Statistical dispersion) 的指標，吉尼係數通常被用來衡量統計數值分布不均的狀況，在社會學與經濟學皆有應用，最常見的應用為計算國家國民所得不均的情形。熵在資訊系統中為衡量系統資訊量的指標，若一系統包含大量資訊—如讀者閱讀的興趣廣泛含多種類型，則系統之資訊量就高。

本研究將以 Simpson's Diversity、Stirling's Revised diversity、Gini 與 Shannon's Shannon's Entropy 並評估該書櫃的多樣性，並比較上述指標的差異。例如不同讀者可能具有不同的偏好多樣性 (Preference diversity)，偏好多樣性是使用者偏好的廣泛程度，或說使用者的偏好是分散或是集中。偏好同質性與偏好穩定度都反應了使用者興趣的可能改變程度，但重要的不同之處在於使用者可能擁有廣泛而穩定的興趣，亦可能使用者有狹窄但不穩定的興趣。而在本研究中係指讀者閱讀類型的多元程度，閱讀偏好同質性高代表偏好特定書籍；閱讀偏好同質性低則代表閱讀的類型較廣泛。

Notation:	
Proportion of elements in category i :	p_i
Distance between categories i and j :	d_{ij}
Similarity between categories i and j :	$s_{ij} = 1 - d_{ij}$
Indices:	
$N =$ Variety	N
$H =$ Shannon	$-\sum_i p_i \ln p_i$
$I =$ Simpson diversity ³	$\sum_{i,j(i \neq j)} p_i p_j = 1 - \sum_i p_i^2$
$\Delta =$ Stirling ($\alpha=1, \beta=1$)	$\sum_{i,j} d_{ij} p_i p_j = 1 - \sum_{i,j} s_{ij} p_i p_j$
Generalised Stirling	$\sum_{i,j} d_{ij}^\alpha (p_i p_j)^\beta$

表 5 selected measures of diversity (Rafols & Meyer, 2010)

五、小結



據前述有關偏好結構之文獻回顧一節，可得知休閒閱讀使用者應有其休閒閱讀偏好結構，而偏好結構皆有其特性如多元或單一、穩定或不穩定。若個別書櫃可代表使用者之閱讀偏好結構，那麼就可以假設在整體 aNobii 網路書櫃中共現模式（pattern）相同之書籍應具有較高相似性。建立書籍與書櫃擁有者間的非對稱矩陣（Asymmetric matrix），進行相似性計算，在取用適當閾值後，可從相似關係圖型中取得巨大組件，可以巨大組件為代表，基於等同（equivalence）的概念針對全體書櫃與個別書櫃的書籍結構作為分析，並將多樣性計算結果與讀者自我認知鄉比較，可探究在社會媒體中讀者偏好與社會推薦發展的可能。


第三章 研究方法與步驟

第一節 研究方法與架構



在(Buldú et al., 2007)利用歌曲網絡分析音樂品味的研究中，對於音樂品味的概念是共同出現在同一播放清單之歌曲可作為清單擁有者的音樂品味；同樣地，在休閒閱讀中，出現在同一書櫃中之書籍集合也可視為該讀者的選書品味與閱讀偏好。此外，於書目計量學的領域中，曾使用多樣性指標分析引用文獻，探究學術研究的跨領域特性；於本研究中，使用者將閱讀的書籍加入網路書櫃的行為可類比至書目計量學中使用多樣性指標的情境。因此，本研究即在嘗試透過相同的概念，利用特定書籍共同出現在特定閱讀品味書櫃，與利用多樣性指標分析個別書櫃中書籍的作者，進行非干擾式（non-obtrusive）方法分析讀者的偏好結構。

本研究將嘗試兩種方式：首先利用(Buldú et al., 2007)的研究概念，該作者利用共同出現在同一播放清單作為歌曲間的連結，透過網絡分析播放清單間的關係，了解歌曲間的相互連結與相關性。本研究亦嘗試探索採用相同的方式，希望透過觀察書籍在書櫃間共同出現的網絡結構，建立書與書之間的連結關係，並依此關係定義書籍的相似性，呈現單一使用者書櫃的偏好結構。然而，此研究仍於多處與本研究有所差異。首先，(Buldú et al., 2007)於研究資料處理上，由於其研究希望建立歌曲間關係，故從播放清單建立歌曲連結後，即著重在歌曲間的連結關係；而本研究欲分析單一讀者之閱讀偏好結構，因此從整體書籍網路出發，利用共現矩陣建立書籍間的相似關係，再依此書籍相似性來呈現個別書櫃的閱讀偏好多樣性。在網絡分析中，首先應確立書櫃中書籍間的連結依據：利用共同出現在同一書櫃（以下簡稱共現）的次數來代表書籍間關連的強度，並以共現矩陣表示，如(林頌堅, 2010)之研究即是利用字詞間共同出現的頻率來建立共現矩陣，代表字詞間的關聯。在建立連結關係後，採用 5 種相似性測量(Spertus et al., 2005)包含 Cosine Similarity 與常見的 Correlation Coefficient 來分析書籍間的相似性。依據連



結關係與書籍間相似性將可建立書籍間的連結網路。在連結網路中，本研究參考 (Henry Small, 2009) 研究中依照巨大元件 (giant component) 變化與 (林頌堅, 2010) 採用的網絡特性包含連結線數與所占成分比率做為選擇閾值的方式，協助釐清分群 (clustering) 的適當位置。接著將於整體書櫃網路中取出之閾值套用至個別書櫃進行分群，即可找出個別書櫃之組成，群集數量多寡即為書櫃多樣性高低。

另一方式，考慮讀者書櫃的書目可能包含多種類型，如推理小說、商業報導等等，因此適合採用書目計量學的中計算研究跨領域的多樣性指標來分析讀者書櫃的多樣性。然而，書目計量學中每篇研究的領域為事先指定；本研究的研究場域—aNobii 中使用 41 種不同類型的標記 (Tag)，且為多重標記 (multi-tagging)，一本書可被一個使用者標記最多三種標籤，使用者更可自訂其書櫃內的標記，因此不僅一本書可能被不同讀者標記為不同類型，亦可能有只有某一使用者使用的個人標記，而且 aNobii 預設之標籤可能屬階層關係如小說與文學、旅行、同性戀等，因此不適合在此採用作為分類依據。



圖 3 anobii 讓使用者建議該書分類之畫面

休閒閱讀讀者選書最重要的參考因素為作者(Mikkonen & Vakkari, 2012; Tang et al., 2012), 因此本研究在採用多樣性指標分類時將類型的定義回歸至最重要的參考因素—作者, 將同一作者視為同一類, 不同作者視為不同類。採用與(Rafols & Meyer, 2010)研究的相同指標分析書櫃的多樣性, 將會計算 Simpson's Diversity、Stirling's Revised Diversity、Gini 與 Shannon's Entropy。而該研究中 Stirling's Revised Diversity 考慮到類別間的相似性, 在此即為作者之間是否相似。因此我們嘗試結合共現矩陣的概念, 利用書籍所屬作者為單位, 建立書籍作者的共現矩陣, 並依此建立作者相似性。並以作為 Stirling's Revised Diversity 公式中計算類別相似性程度的依據。統整上述兩項方式將分群數量與多樣性指標兩種結果與使用者自我評量之閱讀多樣性做比較, 分析何種方式最能夠代表使用者的偏好結構。

在本研究之前並無利用非干擾式分析網路書櫃書並進行相似性測量與使用者分析, 因此本研究著重於探索此一方式的可能性。整體而言, 在架構上系統面需有社會網絡資料收集、共現網路建立、相似性計算、閾值選擇與分群。在使用

者面，量化分析包含完成問卷訪談、自我認知與多樣性比較。



系統面	使用者面
<ul style="list-style-type: none"> ● 社會網絡分析 <ul style="list-style-type: none"> ➢ 共現網路建立 ➢ 相似性計算 ➢ 閾值選擇 ➢ 分群 ● 多樣性分析 <ul style="list-style-type: none"> ➢ 作者共現矩陣 ➢ 多樣性計算 	<ul style="list-style-type: none"> ● 量化 <ul style="list-style-type: none"> ➢ 問卷訪談 ➢ 自我認知與多樣性比較

表 6 資料收集與研究結果比較架構

本研究希望能探索利用社群網絡之尋書工具資料分析呈現休閒閱讀的偏好結構與多樣性，研究資料來源可分為兩部分，系統面與使用者面。系統面可直接從 aNobii 網站取得，收集書籍共同出現在特定書櫃之連結關係，期望進一步分析資料探索網路書櫃使用者之閱讀偏好結構。因此，系統面之資料收集上若採用非干擾性之量化研究方式進行，較能保持研究的客觀性。使用者面資料收集，為了在 50 個抽樣書櫃中深入了解使用者，採用便利抽樣問卷訪談。便利抽樣 (Convenience Sampling)，又稱為「任意抽樣、偶遇抽樣法」，為以方便為主要考量的抽樣方法，樣本的選擇只考慮方便取得或觀察。此種抽樣方法容易進行，但必須了解樣本的特質才適合進行。將會調查使用者對偏好結構自我評量作為評量系統面資料之標竿。並加以分析探討，因此質性與量化之研究方法均有採用。

第二節 研究實施



一、研究對象

本研究關注的對象為利用社群網絡尋書工具之休閒閱讀讀者的偏好結構，因此研究對象需要同時兼具休閒閱讀習慣與使用社群網絡尋書工具兩種條件。由於社群網絡尋書工具中 aNobii 為台灣地區使用率最高之書籍社群網絡且正體中文書籍資料豐富，因此本研究將以 aNobii 為場域。研究對象條件設定為：具備 aNobii 網路書櫃帳號、書櫃需有 80 本書及 5 個以上朋友、願意接受問卷調查，且提供 aNobii 書櫃供系統收集資料。滿足上述條件才得以確保研究對象有使用 aNobii 為其管理書籍平台，並在平時就記錄其擁有或看過之圖書。

針對上述需要本研究採取便利抽樣，徵求消息於 aNobii 中人數與討論較多的熱門群組與台灣最大的電子佈告欄系統（BBS）批踢踢（ptt.cc）中 book 板中張貼。參與實驗者需附上 aNobii 個人頁面網址，並填寫實驗申請表，研究人員確認受試者符合條件後即進行問卷訪談，完成後受試者可獲得車馬費。本研究基於量化與質性分析之需要，收集 50 位受試者之資料，共擁有約有兩萬本不同之書籍。



二、研究步驟

(一) 社會網絡資料收集

如前段敘述，本研究開始時以便利抽樣選擇 50 位受試者接受關於書櫃多樣性之 7 李氏量表問題，問題如下。問卷填答結果利用 SPSS 做因素分析，排除不合適的題目，產生的多樣性分數即為使用者自評之多樣性分析。

問卷題項

偏好同異性

我有固定追隨的作者或作品類型，不熟悉的領域我不會想主動找來看

我的閱讀興趣十分廣泛，不限於某種類型(*)

我的閱讀興趣很固定，變化不大

除了我喜歡的作品類型及作者外，我也會主動尋求不熟悉的類型或作者的作品(*)

(*) 題目為反向題

表 7 偏好結構問卷之平均數與標準差

受試者並提供其 aNobii 網路書櫃網址，研究人員使用程式抓取所有受試者擁有之書籍，建立受試者擁有之所有書籍清單，根據此一清單從 aNobii 網頁中抓取清單中書籍在全體 aNobii 網路書櫃中共同出現之情形，並記錄存入研究人員之資料庫中。



(二) 社會網絡分析與多樣性指標分析

在資料蒐集完成後，資料以下列兩種方式分析，並在(三)將兩種分析方法的結果加以比較。

1. 社會網絡分析：

(1) book-owner 矩陣建立、建立本研究使用之共現矩陣

為了讓書籍在多個書櫃的共現情形能夠表達呈現與進一步計算，在抽樣 50 個書櫃後，我們採用共現矩陣 (Co-occurrence Matrix) 的方式來表現書籍間的連結關係與共現網絡，本研究採用(Leydesdorff & Vaughan, 2006)與(林頌堅, 2010)中建立共現矩陣 (Co-occurrence Matrix) 的方式。

本研究共現矩陣中書籍共同出現之次數計算方式為從抽樣 50 個書櫃中所有的書籍清單出發，計算任意兩本書於 aNobii 整體網絡中共同被書櫃擁有的次數總和，因此最大可能共現次數並非 50 次。亦即若書籍 A (如圖 4)「Whoops!! 大債時代」在所有 aNobii 書櫃中出現 1195 次，且其中有 1000 個書櫃亦有出現書籍 B，則書籍 A 與書籍 B 的共現次數為 1000 次，並不侷限在本研究抽取的 50 個書櫃。研究中書櫃的抽樣僅決定哪些書籍會被列入共現次數計算。例如，假設書籍 C 皆不在此 50 個書櫃中，則書籍 C 與書籍 A 或書籍 B 的共現情形將不會被考慮。

此外，由於本研究抽樣的書櫃以中文書為最大宗，由於中文編碼的問題，建立共現矩陣時中文書名顯示常會有問題，於矩陣中以書名作為依據則不容易辨識。因此，擷取 aNobii 個別書籍的頁面後的編碼 (即圖 4 之網址列最後 18 碼英文數字「0167de97a63ffd09d1」)，作為本研究中抽樣書籍的辨識依據，解決個別書櫃中因各種因素產生的書籍重複問題。

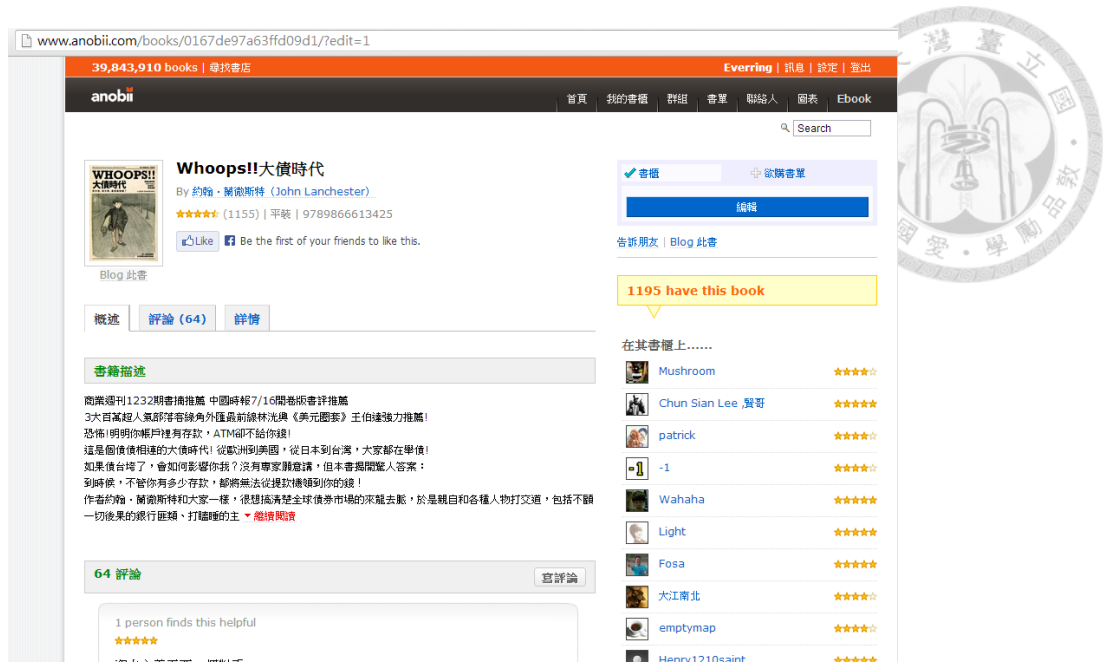


圖 4 擷取抽樣書櫃中每一本書籍的所有書櫃擁有者，以「Whoops!!大債時代」一書為例，這本書有 1195 個書櫃擁有者將其加入書櫃

共現矩陣是計量所有書籍在個別書櫃中兩本書共同出現情形的矩陣。在建立共現矩陣之前，首先必須建立由所有個別書櫃中出現的書籍建立整體書櫃的書目，此一書目亦為共現矩陣的維度，接著收集擁有 50 個書櫃內所有書籍 (book, 19776 本) 的所有書櫃擁有者 (後簡稱 owner)，以建立 book-owner 矩陣。書目中每本書籍與擁有 50 個書櫃內書籍集合之任一書櫃擁有者各由一維度代表。統計所有個人書櫃，若 book i 代表維度為第 i th 維，owner 代表維度為第 j th 維。此一共現矩陣為非對稱方矩陣 (Asymmetric Square Matrix)，利用此 book-owner 矩陣可表達出各個書籍間被個別書櫃共同持有之情形。(如圖 5)

	Owner 1	Owner 2	Owner 3
Book a	1	1	1
Book b	0	1	1
Book c	0	1	0

圖 5 book-owner 矩陣

(2) 相似性計算、利用共現矩陣定義相似性

在 book-owner 矩陣中，每一列 (Row) 皆為一向量，而此向量的每一維度皆為該維度代表之書籍於該列代表之書櫃擁有者之出現情形。我們可以從書籍間於特定書櫃出現的情形異同，分析書籍間的相似性。

在區域共現觀點，我們可以透過正規化共現矩陣的共現情形了解書籍間共現相似情形，如文獻回顧一節中(林頌堅, 2010)分析詞語頻率即是偏向此一觀點，我們採用三種計算方式，方式一為 Jaccard Index，若兩本書為 A、B，其計算方式為 $\frac{A \cap B}{A \cup B}$ ， $A \cap B$ 為書籍 A 與書籍 B 之共現次數， $A \cup B$ 為書籍 A 在所有書櫃中總數加上書籍 B 在所有書櫃中總數扣掉共現部分。方法二 Inter 為 $\frac{A \cap B}{\text{Min}(A,B)}$ ， $A \cap B$ 為書籍 A 與書籍 B 之共現次數， $\text{Min}(A,B)$ 為書籍 A 與書籍 B 之整體書櫃出現次數 (被持有數) 較小者，與(林頌堅, 2010)計算詞語間共現比重公式相同。但我們發現方法二在某兩本書在全部書櫃中共現次數僅一次時， $A \cap B$ 與 $\text{Min}(A,B)$ 均為 1，而結果值亦達到極大值 1，但此結果並不合理，因發生次數僅 1 次卻造成相似結果造成極大值。以實際資料來計算，根據方法二，旅遊書「香港」與「小王子」在全體書櫃中共現次數僅 1 次，根據方法二計算此兩本書之相似性為相似性之最大值 1，是相互間最相似的書籍之一，結果並不準確，因此我們嘗試修改提出方法三 Interminus 為 $\frac{A \cap B - 1}{\text{Min}(A,B) - 1}$ ，利用將分子分母次數同減一的方式排除方法二在邊界值極端之不合理情形。根據方法三，前述偶然出現的情況將不被考慮，最相似之書籍為其他本書，較為合理。此方法之表示符號意義同方法二。考慮整體結構，透過計算兩向量間相似度的方式來了解兩書籍間的在圖形結構相似度。在計算向量相似度測量中最普遍且常見之兩方法為餘弦相似度 (Cosine Similarity) 與相關性 (Correlation)。

定義書籍 A，書籍 B 之餘弦相似度 (Cosine Similarity) 公式如下：

$$\text{Similarity}(A, B) = \cos \theta = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} = \frac{\sum_{i=1}^n \vec{A}_i \times \vec{B}_i}{\sqrt{\sum_{i=1}^n (\vec{A}_i)^2} \times \sqrt{\sum_{i=1}^n (\vec{B}_i)^2}}$$

其中 \vec{A} 為 book-owner 矩陣中代表書籍 A 之行向量(於所有書櫃的出現情形)，其中 \vec{B} 為共現矩陣中代表書籍 B 之行向量，n 為共現矩陣之維度。Similarity(A,B) 即為書籍 A 與書籍 B 之相似度。

定義兩向量之相關性 (Correlation) 公式如下:

$$\text{Correlation}(A, B) = \frac{(\vec{A} - \bar{A}) \cdot (\vec{B} - \bar{B})}{\|\vec{A} - \bar{A}\| \|\vec{B} - \bar{B}\|}$$

其中 \vec{A} 為 book-owner 矩陣中代表書籍 A 之行向量(於所有書櫃的出現情形)，其中 \vec{B} 為共現矩陣中代表書籍 B 之行向量，n 為共現矩陣之維度， \bar{A} 中任一元素為 \vec{A} 所有元素的平均值， \bar{B} 中任一元素為 \vec{B} 所有元素的平均值。Correlation(A,B) 即為書籍 A 與書籍 B 之相似度。

透過上述區域觀點的三種方式與整體結構的兩種方式，即可得知兩兩書籍間之相似關係。本研究將計算上述五種方法並比較其結果。

(3) 閾值選擇並對個別書櫃進行分群

經由 (2) 可以建立所有書籍間兩兩之相似性之值。我們可以由任一本書為起始點，建立書籍與書籍間的相似圖形之網絡。該網絡之連結強度代表兩本書籍間之相似性。採用前述(Henry Small, 2009)與(林頌堅, 2010)的作法，透過將閾值之下相似性較低之弱連結視為無連結，相互連結之節點數最大之子圖為巨大元件，在網絡中逐步改變門檻值 (threshold)，觀察尋找巨大元件尺寸與書籍間連結數量改變量，最大之改變點即為最適當之分群點。

上述之閾值尋找過程是使用在所有調查書籍所建立的全部抽樣書籍網絡中，並將此閾值套用在 50 個個別書櫃中，觀察套用閾值後個別書櫃之書籍分群情形。

此外，採用全部抽樣書籍網絡而非個別書櫃各自計算的原因在於個別書櫃書籍數量不同、使用者使用書櫃時間長短不同、對於網路書櫃的使用習慣可能亦不同，難以針對個別書櫃尋求適當閾值，而全部抽樣書籍網絡已抽取網絡中所有書籍在所有 aNobii 書櫃中之情形，可呈現調查書籍在 aNobii 的全體情形，因此採用全部抽樣書籍網絡尋找閾值較適合。

本研究對個別書櫃利用 UCINet 之 Region、Block 與 Girvan Newman 進行個別書櫃的分群，個別書櫃之分群數量即代表其閱讀偏好多樣性。以下分述 3 種分群方式的演算法。

a. Region (Components) 分群

將網絡中的 Component 視為群，Component 定義為 Component 中任意兩點可以經由路徑連結，以圖 6 為例，圖中網絡有四個 Component。

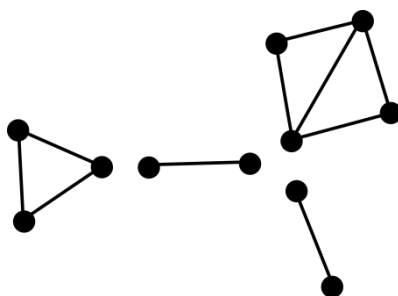


圖 6 Region 分群演算法

參考來源：(Hanneman & Riddle, 2005)

b. Block (Biconnected Components) 分群

首先，找出網絡中的橋 (Bridge)，將網絡的所有橋去除，而網絡中的每個 Component 與每個被移除的橋包含其對應的點為 Biconnected Component，本研究以 Biconnected Component 作為分群。其中，橋是一種邊，當它移除後會使得網絡的 Component 數目增加。以圖 7 為例，網絡中藍色的邊是橋，移除橋之

後，成為兩個 Biconnected Component，而橋有兩個也是 Biconnected Component，總共有四個 Biconnected Component，在本研究中將網絡視為四個群。

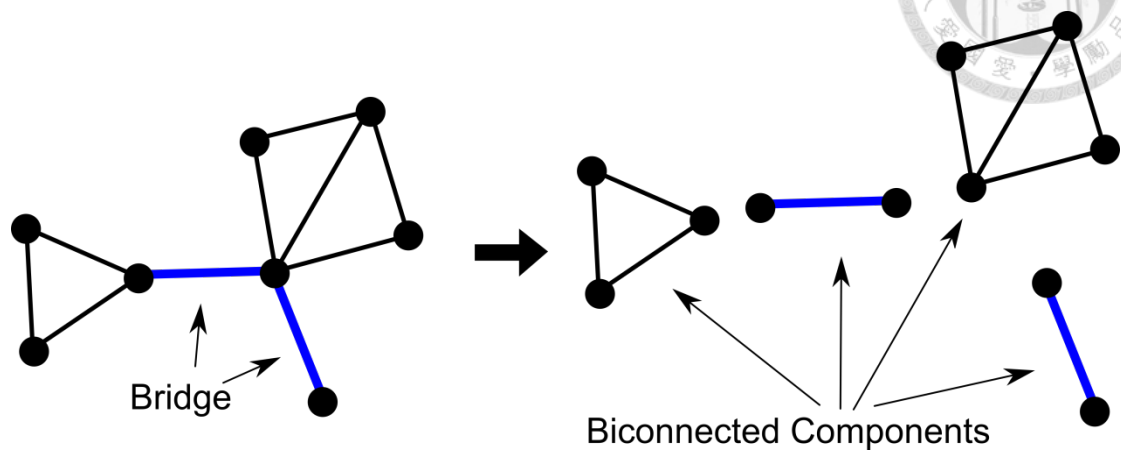


圖 7 Block 分群演算法

參考來源：(Cormen, Leiserson, Rivest & Stein, 2009)

c. Girvan-Newman 分群

將網絡以 Girvan-Newman 分群過程，會不斷的移除邊，由於連結斷掉，使得 Component 數目增加，此時將記錄當下網絡並且網絡中每個 Component 視為群。在 UCINET6 中，每次移除邊之後，將計算網絡的品質 (Quality, 即為 Q 值)，最後移除完所有的邊後，選出品質最好的當下網絡狀態。

移除邊的演算法為：

1. 對於網絡中的每個邊，計算其 Edge Betweenness。
2. 移除 Edge Betweenness 最高的邊。
3. 重新對圖中剩餘的邊，計算其 Edge Betweenness。
4. 從 2 重複循環，直到沒有任何剩餘邊。

而任意邊 e 的 Edge Betweenness 為



$$EB(e) = \sum_{s \neq t} \frac{|p_{st}(e)|}{|p_{st}|}$$

其中，s 與 t 為網絡中任意相異點，而 p_{st} 為以點 s 為起點，點 t 為終點的最短路徑集合； $p_{st}(e)$ 路徑通過邊 e，並且為 p_{st} 的子集。以圖 8 為例，選定網絡中節點 s、t 與邊 e， p_{st} 為兩條藍色路徑， $p_{st}(e)$ 為通過 e 的一條紅色路徑， $\frac{|p_{st}(e)|}{|p_{st}|} = \frac{1}{2}$ 。

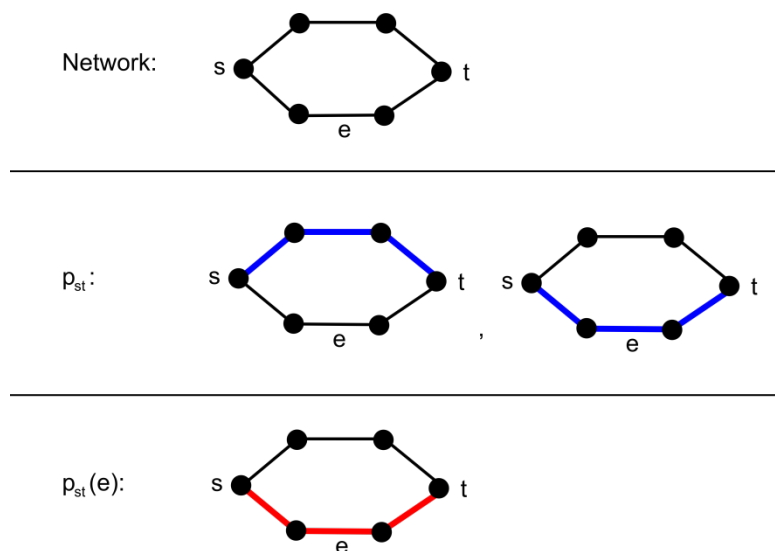


圖 8 Girvan-Newman 分群演算法

參考來源：(Pinney & Westhead, 2006)

網絡的品質定義為

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j)$$

其中 m 為網絡中邊總數，i 與 j 為網絡中節點，而 k_i 與 k_j 分別為 i 與 j 擁有邊的數量。 C_i 與 C_j 是 i 與 j 所屬的群，而 $\delta(C_i, C_j)$ 為一函數，定義如下：

$$\delta(C_i, C_j) = \begin{cases} 1, & \text{若 } C_i = C_j \\ 0, & \text{若 } C_i \neq C_j \end{cases}$$

參考來源：(Pinney & Westhead, 2006；Fortunato, S., 2010)



2. 利用作者分類與作者相似矩陣

(1) 計算多樣性指標

本研究另一探索方向為將讀者閱讀書籍與論文引用文章類比為相似概念形式，採用書目計量學之指標分析其多樣性。因休閒閱讀並無普遍適切之區分書籍種類方式，採用主題標目（subject heading）或是圖書分類法皆難以清楚區分書籍是否為同一類型，因此本研究根據休閒閱讀常用之尋書方式(Mikkonen & Vakkari, 2012; Tang et al., 2012; 謝宜瑾, 2012)，以不同作者為不同書籍種類的方式計算多樣性。根據(Rafols & Meyer, 2010)指出多種常見多樣性計算指標，本研究將使用 Simpson's Diversity、Stirling's Revised Diversity、Gini 與 Shannon's Entropy 等書目計量學中常用之多樣性指標。而 Stirling's Revised Diversity 需考慮作者間的相似性，將於下一步驟建立。

可直接使用之公式如下列:

a. Simpson's Diversity 公式

$$\text{Diversity} = \sum_{i,j(i \neq j)} P_i P_j = 1 - \sum_i P_i^2$$

其中 P_i 、 P_j 分別代表第 i 類型、第 j 類型的比率

b. Gini 公式

$$\text{Gini} = 1 - \sum_{i=1}^n (X_i - X_{i-1})(Y_i + Y_{i+1})$$

其中 n 為書櫃中作者總數，將該書櫃中所有作者依該書櫃中包含其著作數由少至大排序，第 i 作者為書籍排名第 i 少（或是第 $n-i+1$ 多）之作者。 X 為書櫃中作者函數，在此每一作者之函數值皆為 $\frac{1}{\text{書櫃中作者總數}}$ ， Y 為書櫃中書籍比率累積函數（cumulative function）， Y_i 為作者 1 至作者 i 之所有書籍加總佔書櫃中書籍總數所佔比率。



c. Shannon's Entropy 公式

$$\text{Entropy} = -\sum P_i * \log_2 P_i$$

其中 P_i 代表第*i*類型的比率。

(2) 作者相似矩陣

指出了 Simpson Diversity 無法衡量種類間相似性的問題，而在休閒閱讀中，也因不同作者間風格可能迥異亦可能較為類似，亦無法用 Simpson Diversity 表達此一情形。因此本研究亦嘗試使用 Stirling 對 Simpson's Diversity 所提出的類型間的相似性修正，同時採用考慮作者相似的 Stirling Revised Diversity 公式。而該公式需要類型間的相似度，在此即為作者間的相似性。因此本研究亦需要建立作者相似性。為達到非干擾式且資料量大亦能分析的目標，本研究採用與網絡分析相同的方式建立作者相似性，先建立作者共現矩陣；將前述以書籍共現方式建立的共現矩陣，以作者共現的方式取代，即只要書櫃中出現有作者 A 寫的任一本書與作者 B 寫的任一本書，就是作者 A 與 B 共現，再以與網絡分析步驟 (2) 相同之相似性計算方式建立作者相似性，並將作者相似性帶入 Stirling's Revised Diversity 公式中。

Stirling's Revised Diversity 公式:

$$\sum_{i,j} d_{ij} p_i p_j = 1 - \sum_{i,j} s_{ij} p_i p_j$$

其中 p_i 、 p_j 為第*i*類型、第*j*類型的比率， d_{ij} 為作者*i*與作者*j*之不相似性 (dissimilarity) 之值， s_{ij} 為作者*i*與作者*j*之相似性之值。 $0 \leq d_{ij} \leq 1, 0 \leq s_{ij} \leq 1$ 。

(三) 以關聯性分析 aNobii 使用者閱讀偏好結構

在計算出書櫃的各種多樣性指標之後，將計算結果與該讀者李氏量表問卷的結果作相關性分析，以讀者的認知作為標竿，利用分析結果回答本研究提出的各

項問題。

本研究預計進行作者分析與社會網絡分析，而社會網絡分析中相似性測量方式有正規化、Cosine Similarity 與 Correlation 三種，而分析結束後多樣性指標共有 Diversity（社會網絡分析採用 Simpson's Diversity，作者分類可考慮作者相似而可採用 Stirling Revised Diversity）、Gini 與 Shannon's Entropy 三種，因此每個書櫃兩種分析會至少有 12 種組合，社會網絡分析採用不同閾值會更增加結果的數量，因此本研究將會列出數個閾值與其各自組合結果，並加以分析，期能探索出適合之社群網絡尋書資料之非干擾性測量分析方式。

第三節 研究範圍與限制



一、以 aNobii 網路書櫃作為研究場域

aNobii 為台灣讀者使用率最高的網路書櫃，且正體中文書籍豐富。本研究之研究目標為使用量化、非主觀、非干擾性的方式，利用網路書櫃取得之資料分析呈現個別讀者的閱讀偏好結構，考量抽樣書櫃內容可分析性與使用者可接洽性，選擇最多華人使用的 aNobii 網路書櫃作為研究場域。但 aNobii 的使用者多為喜好休閒閱讀之使用者，閱讀行為之涉入性較高，其結果未必能類推至一般社會大眾。

二、抽樣書櫃的數目

由於本研究之進行，必須仰賴受試者的判斷，受限於時間與地理因素，研究對象為便利抽樣之 50 個網路書櫃，以取得受試者的判斷作為抽樣的基本條件。承上述，本研究藉由問卷訪談蒐集讀者自我評量之偏好結構多樣性，並以此作為多種相似性測量呈現偏好結構的判準標竿。

三、受試者個人書櫃的完整度

本研究在受試者的同意下，蒐集書櫃清單以建立書櫃網路。若受試者將書籍設為不公開，為尊重受試者隱私，不會將書籍列入書籍清單。因此在特殊情形如使用者隱藏大量書籍，或未將閱讀的書籍加入清單等因素皆會影響書櫃的完整度，因此可能無法完整反映讀者的閱讀偏好結構。

四、使用問卷作為使用者自我評量標竿

使用者自我評量之資料最終將於資料分析階段，與網路書櫃資料分析呈現之偏好結構做關聯式分析，因此採用 0-6 分七點量表問卷作為使用者自我評量之研究工具。惟應留意在推論與解釋時，應考量問卷調查的限制，須注意不同使用者自我認知與實際情況的主客落差。

第四章 結果與討論

第一節 研究結果



本研究於 aNobii 網路書櫃服務中抽樣 50 個使用者，建立書籍共現矩陣作為建構使用者閱讀偏好多樣性指標的依據。由於 aNobii 平台的限制，無法單次下載全部抽樣使用者書櫃的資料，資料收集時間為民國 2010 年 8 月 20 日至 2011 年 8 月 8 日。研究所需書櫃資料，由於時間差可能會有資料不一致的情形，因此在收集完畢 50 個抽樣書櫃的書籍並建立書籍清單後，重新比對資料做書籍清單的增加或刪減，並重新檢驗個別書籍被書櫃擁有的次數建立 book-owner 矩陣而後利用書籍共現矩陣進行書籍間相似性的計算。

本研究進行資料收集前，使用問卷訪談對 50 個抽樣書櫃的使用者，針對其基本資料與 aNobii 網路書櫃的使用習慣做調查。基本資料顯示（如表 8），男性共 21 名，占整體受測者 42%，女性共 29 名，佔 58%，性別比例大致相近，女性受測者較多；年齡分布從 20 到 34 歲，平均年齡為 23.98 ($SD = 3.485$)；書櫃書籍總數最小值為 80 本，最大值為 3780 本，平均數為 563.28 ($SD = 703.19359$)，顯示個書櫃間書籍總數差異大；閱讀偏好多樣性，於此為使用者自評之閱讀偏好多樣性的值，此統計量為利用 SPSS 統計軟體轉換李氏 0~6 分七點量表的值，最小值為 -2.45，最大值為 1.75，平均數為 0 ($SD = .98974$)。其中，書櫃編號 109 因未取得使用者自評閱讀偏好多樣性資料，後續作關聯性分析時去除此筆資料。

	最小值	最大值	平均數	標準差
年齡	20	34	23.98	3.485
閱讀偏好多樣性	-2.45	1.75	.0000	.98974
書籍總數	80.00	3780.00	563.2800	703.19359

N = 50

表 8 抽樣書櫃統計資料



一、書籍共現情形之多樣性指標

(一) 最適閾值

研究中採用3種正規化方式與2種相似性計算方式建立抽樣書櫃中所有書籍的相似性矩陣，接著以巨大組件大小減少比例為最適閾值判別依據，並輔以社會網絡分析中的網絡密度 (density)、平均連結線數 (average link) 作為輔助判斷指標。本研究利用整體書籍網絡矩陣之閾值結果作為評斷最適閾值之依據，以下分述5種矩陣計算方式之最適閾值結果分析，最適閾值將依據不同正規化與相似性計算方式套入個別書櫃中作為分群前的準備。

1. Jaccard

閾值	最大組件大小	連結線數	平均連結線數	網絡密度	最大成分所占比例	最大組件節點減少數	最大組件減少比例
0.0	21202	224751801	21201.00	1.00000	100.00%	0	0.00%
0.1	13679	2036410	192.10	0.00906	65.00%	7523	35.00%
0.2	6151	710868	67.06	0.00316	29.00%	7528	55.00%
0.3	3485	468378	44.18	0.00208	16.00%	2666	43.00%
0.4	1996	263806	24.89	0.00117	9.00%	1489	43.00%
0.5	1867	258435	24.38	0.00115	9.00%	129	6.00%
0.6	369	103330	9.75	0.00046	2.00%	1498	80.00%
0.7	369	98440	9.29	0.00044	2.00%	0	0.00%
0.8	369	94851	8.95	0.00042	2.00%	0	0.00%
0.9	369	91187	8.60	0.00041	2.00%	0	0.00%
1.0	369	89288	8.42	0.00040	2.00%	0	0.00%

表 9 Jaccard 最適閾值

如表 9，最大組件隨著閾值的增加以減少一半的速度逐步減少，然而在閾值上升至 0.6 時，最大組件的節點數突然少了 80% 左右，此時最大組件的節點數產生巨

大的變化；若從 0.4~0.5 閾值變化的數值來看，閾值在上升至 0.5 時，最大組件節點數僅減少 6% 左右，代表 0.5 是個最大組件變化的停滯點，適合用來作為閾值。此外，從 0.6 開始後增加閾值大小不會改變巨大組件的大小，代表在 Jaccard 計算下的數值閾值取在 0.1~0.5 之間較能夠看出書櫃中分群的變化。因此取 0.5 作為 Jaccard 計算下的最適閾值。

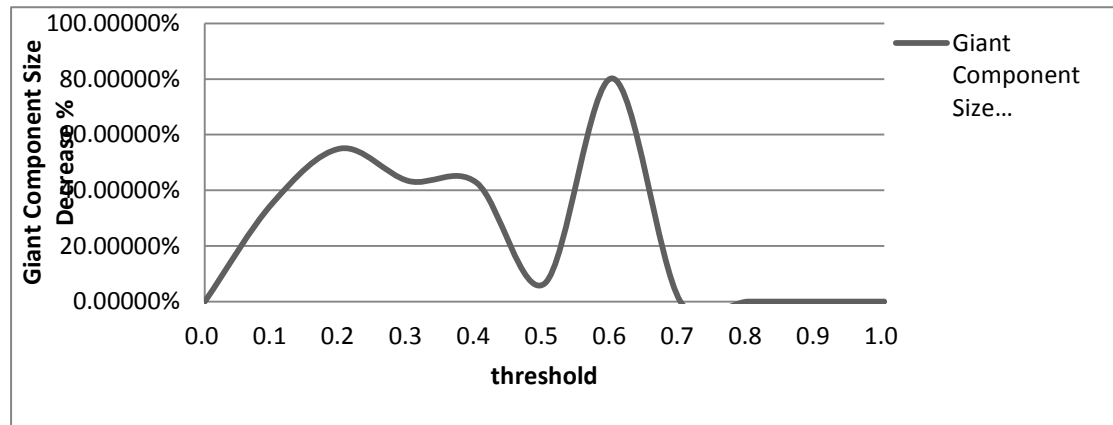


圖 9 Jaccard 最適閾值

2. Inter (Intersection (A, B) / min (A, B))

閾值	最大組件大小	連結線數	平均連結線數	網絡密度	最大成分所占比例	最大組件節點減少數	最大組件減少比例
0.0	21202	224751801	21201.00	1.00000	100.00%	0	0.00%
0.1	21201	12845507	1211.73	0.05715	100.00%	1	0.01%
0.2	21197	7930273	748.07	0.03529	99.98%	4	0.02%
0.3	21162	6005832	566.53	0.02672	99.81%	35	0.17%
0.4	21033	4234441	399.44	0.01884	99.20%	129	0.61%
0.5	20877	4090554	385.86	0.01820	98.47%	156	0.74%
0.6	20496	2237524	211.07	0.00996	96.67%	381	1.83%
0.7	20324	2075632	195.80	0.00924	95.86%	172	0.84%
0.8	20224	2041987	192.62	0.00909	95.39%	100	0.49%
0.9	20194	2034054	191.87	0.00905	95.25%	30	0.15%
1.0	20182	2025912	191.11	0.00901	99.79%	12	0.06%

表 10 Inter 最適閾值

由於分母為取兩本書之間被書櫃擁有次數的最小值，在多數常出現在多個書櫃（譬如暢銷書）的書籍與其他書籍的相似性，會因為此種正規化方式分母的處理方式，使得書籍間的相似值分布最平均的一種方式。因此，在閾值設定的逐步增加下，最大組件大小也僅是以非常小的幅度逐漸減少（如表 10）。儘管如此，相較之下閾值上升至 0.6 時，巨大組件的大小減少幅度最大，代表巨大組件的崩解，此時大部分多餘的連結都經由閾值的篩選，若更進一步增加閾值，則可能進入下一個分群的階層。因此，在巨大組件崩解前的網絡狀況最適合做分群，選擇 0.5 作為最適閾值。

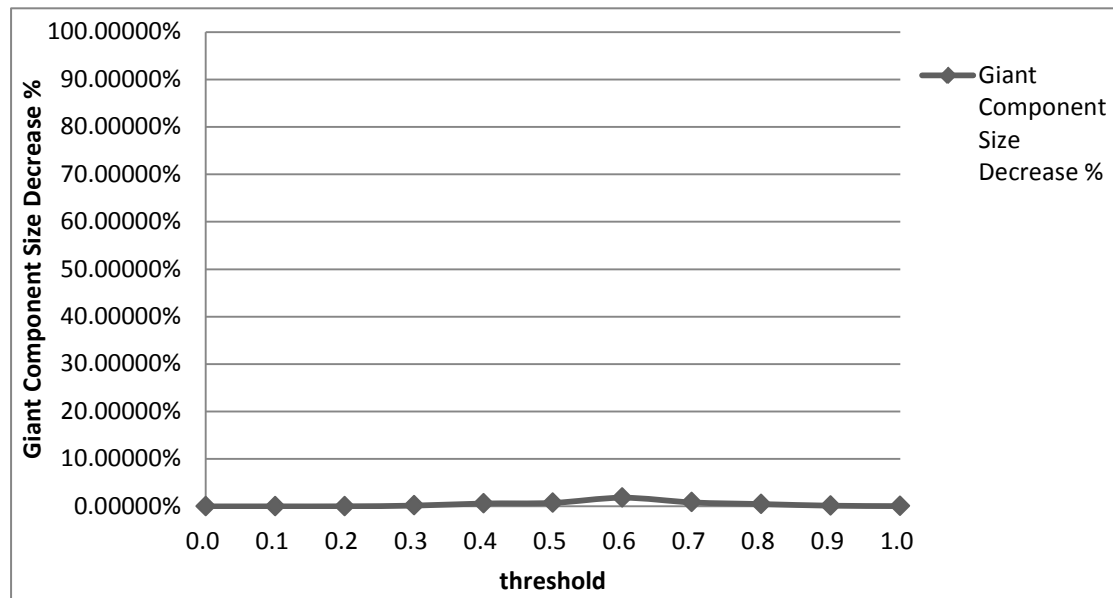


圖 10 Inter 最適閾值

3. Interminus (Intersection (A, B)-1 / min(A, B)-1)

閾值	最大組 件大小	連結線數	平均連結 線數	網絡密度	最大成 分所占 比例	最大組 件節點 減少數	最大組 件減少 比例
0.0	21202	37780694	3563.88	0.16810	100.00%	0	0.00%
0.1	20123	3766215	355.27	0.01676	94.91%	1079	5.09%
0.2	19977	1639819	154.69	0.00730	94.22%	146	0.73%
0.3	19602	892808	84.22	0.00397	92.45%	375	1.88%
0.4	18480	473968	44.71	0.00211	87.16%	1122	5.72%
0.5	17432	383326	36.16	0.00171	82.22%	1048	5.67%
0.6	13803	162174	15.30	0.00072	65.10%	3629	20.82%
0.7	12187	126976	11.98	0.00057	57.48%	1616	11.71%
0.8	11202	111730	10.54	0.00050	52.84%	985	8.08%
0.9	10815	105377	9.94	0.00047	51.01%	387	3.46%
1.0	10582	97459	9.19	0.00043	49.91%	233	2.15%

表 11 Interminus 最適閾值

在此種計算方式下，讓最大組件的大小在閾值上升到 1，仍舊有 10582 個節點。而隨著閾值的上升，最大組件的減少比例依據閾值呈不規則分布。然而如表 11，閾值由 0.5 上升至 0.6 時，巨大組件的大小減少幅度明顯最大，至少是其他閾值區間的 2 倍。此外，density 於閾值 0.5 時，停頓於 0.001706，若進行至 0.6，density 將驟減至 0.000722。同時考量最大組件大小的變化與 density 變化的停頓點，選擇 0.5 作為最適閾值。

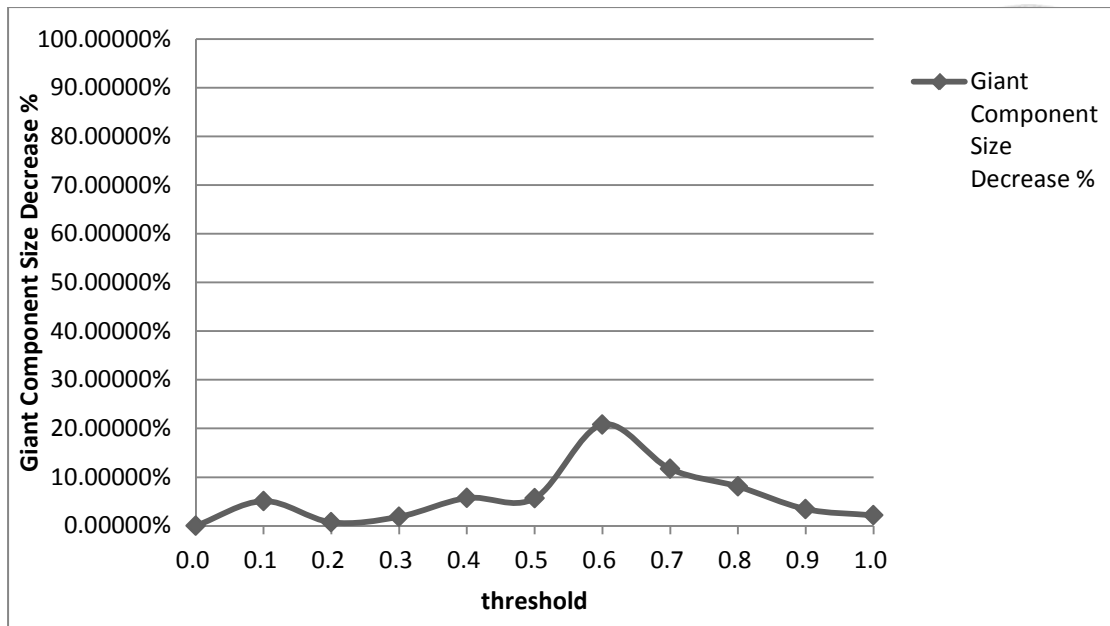


圖 11 Interminus 最適閾值



4. Cosine

閾值	最大組件大小	連結線數	平均連結線數	網絡密度	最大成分所占比例	最大組件節點減少數	最大組件減少比例
0.0	21202	224751801	21201.00	1.00000	100.00%	0	0.00%
0.1	21122	7299165	688.54	0.03248	99.62%	80	0.38%
0.2	16901	2797888	263.93	0.01245	79.71%	4221	19.98%
0.3	10862	1402086	132.26	0.00624	51.23%	6039	35.73%
0.4	7351	854556	80.61	0.00380	34.67%	3511	32.32%
0.5	4851	560315	52.85	0.00249	22.88%	2500	34.01%
0.6	2206	266740	25.16	0.00119	10.41%	2645	54.53%
0.7	1761	256696	24.21	0.00114	8.31%	445	20.17%
0.8	369	102958	9.71	0.00046	1.74%	1392	79.05%
0.9	369	95265	8.99	0.00042	1.74%	0	0.00%
1.0	369	89512	8.44	0.00040	1.74%	0	0.00%

表 12 Cosine 最適閾值

Cosine 的相似性計算方式為把兩本書視為兩個向量，計算兩本書的餘弦相似度。由於計算方式與前面所列 3 種方式不同，在設定閾值時最大組件的大小分布與減少情形幅度相較之下較大（如表 12）。閾值 0.8 是 Cosine 計算方式下，最大組件大小變化最大的閾值，0.7 亦是網絡密度減少至某一程度的停頓點。此外，觀察最大組件在閾值 0.7 的表現可以發現，若繼續將閾值提升，最大組件的大小則在 369 個節點作停頓。這代表若選擇 0.8 作為閾值，則會留下 369 個相似性為 1 的節點，若一個網絡只剩下相似性等於 1 的節點，在書籍網絡的下可能代表的是同屬一系列的書，而不是彼此間「類似」的書籍。這種情形不符合最初設定閾值以輔助分群的目的，因此選擇 0.7 為最適閾值。

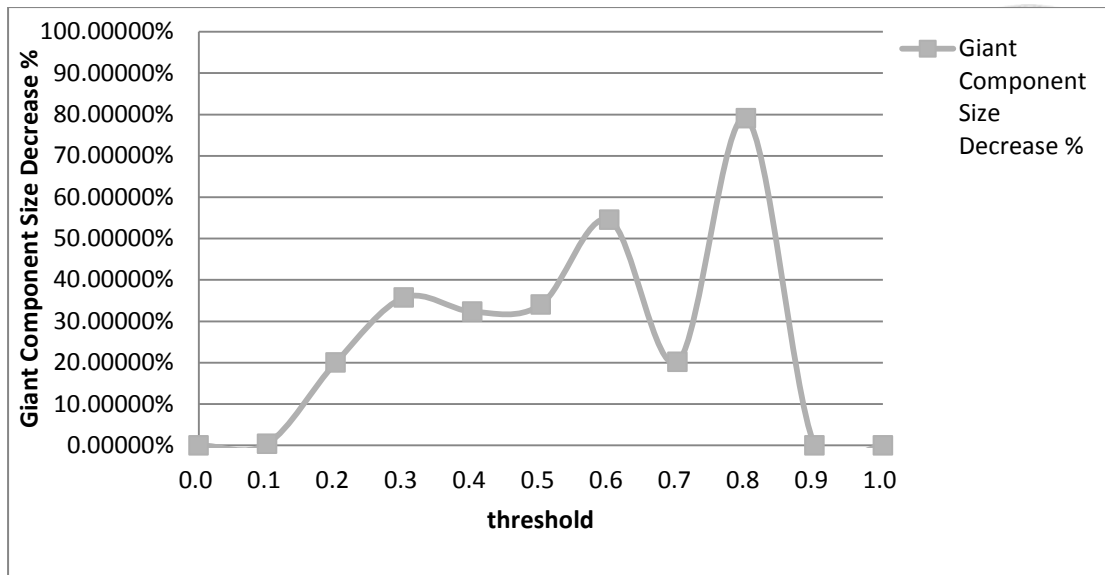


圖 12 Cosine 最適閾值

5. Correlation

閾值	最大組 件大小	連結線數	平均連結 線數	網絡密度	最大成 分所占 比例	最大組 件節點 減少數	最大組 件減少 比例
0.0	21202	224751801	21201.00	1.00000	100.00%	0	0.00%
0.1	21201	13624300	1285.19	0.06062	100.00%	1	0.01%
0.2	19033	3869104	364.98	0.01722	89.77%	2168	10.23%
0.3	13393	1640063	154.71	0.00730	63.17%	5640	29.63%
0.4	7447	918933	86.68	0.00409	35.12%	5946	44.40%
0.5	4852	577396	54.47	0.00257	22.89%	2595	34.85%
0.6	2208	270860	25.55	0.00121	10.41%	2644	54.49%
0.7	1761	257239	24.27	0.00115	8.31%	447	20.25%
0.8	369	103206	9.74	0.00046	1.74%	1392	79.05%
0.9	369	95497	9.01	0.00043	1.74%	0	0.00%
1.0	369	89512	8.44	0.00040	1.74%	0	0.00%

表 13 Correlation 最適閾值

Correlation 的計算方式與 Cosine 較相似，因此在設定閾值與組件大小變化的情況也很類似。閾值大小在 0.8 時，組件大小的減少比例為 79.05%（如表 13），為閾值區間中變化最大。若考量組件大小與網絡密度觀察網絡崩解的停頓點，則 0.7 為 Correlation 計算方式下的最適閾值。

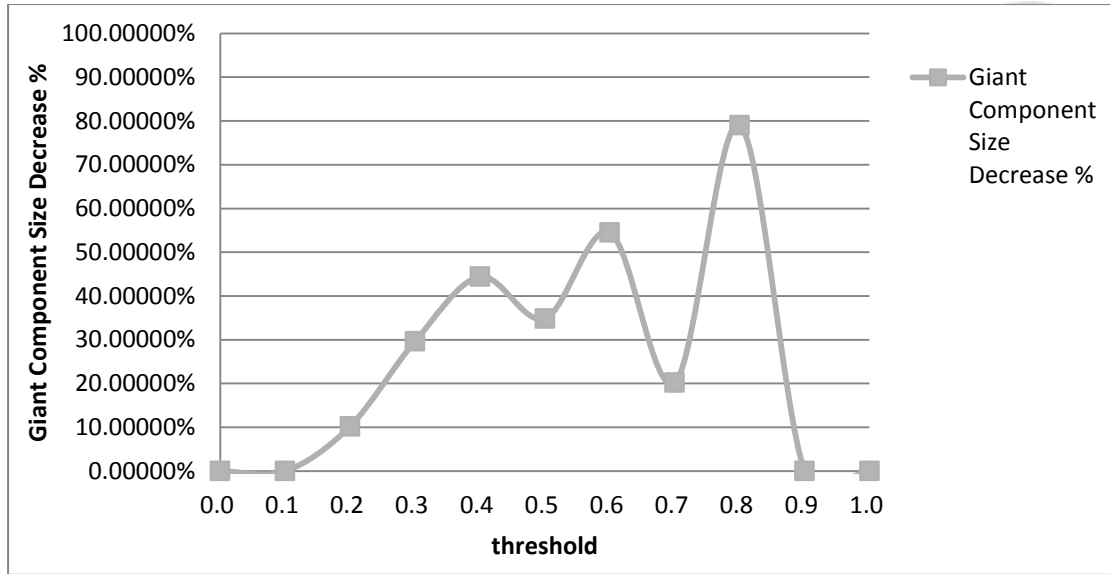


圖 13 Correlation 最適閾值

(二) 書籍分群多樣性與使用者自評多樣性之關聯

將 5 種書籍相似性計算方式，依據前述研究結果套入最適閾值後，採用 UCInet 軟體內建的 3 種分群方式 Region、Block、Girvan Newman，3 種分群方式在閾值為 0 時分群結果在 5 種相似性計算方式的任一書櫃分群數量皆為 1，表示在未設定閾值的情形下，如果沒有自行設定個別書櫃的分群目標數，皆無法對個別書櫃進行分群。以下依據不同分群方式，將套入最適閾值的 5 種相似性計算方式分群結果與使用者自評之閱讀多樣性分數作雙變數相關性分析。

1. Region-binary graph

UCInet 執行程序：Network>Regions>Components>binary graphs>weak component

UCInet 中用來區分 subgroup 的 region 若選擇 binary graph，則是單純計算一個網絡中的組件數，weak component 為不考慮連結方向 (undirected) 計算組件。本研究於選擇最適閾值後，個別書櫃的檔案則使用二分法 (dichotomize) 將書的共現矩陣轉換為 0 與 1，意即網絡中的連結沒有程度上的差別。個別書櫃在套入最適閾值後，網絡會去除一些連結，剩下的組件數則可視為個別書櫃擁有者的閱讀偏好多樣性。

在排除缺乏使用者自評閱讀偏好多樣性資料的 109 號書櫃後，共計分析 49 個書櫃。描述性統計結果顯示 (如表 14)，Jaccard_0.5 (意指在 Jaccard 的相似性測量下，採用 0.5 作為網絡的最適閾值)、Cosine_0.7、Correlation_0.7 在敘述性統計下的數值型態相似，組件數最小值在 60 左右，平均數則在 410 左右 (標準差約等於 510)，顯示個別書櫃在以上相似性計算之下，組件數較多，同時差異也大；Inter_0.5 與 Interminus_0.5 數值型態則是組件數較少，差異相對不大。

SPSS 執行程序：分析>敘述統計>描述性統計量

	最小值	最大值	平均數	標準差
Jaccard_0.5	58.00	2873.00	413.9388	504.18409
Inter_0.5	1.00	349.00	27.2857	66.19164
Interminus_0.5	1.00	450.00	92.6122	84.06590
Cosine_0.7	66.00	2934.00	426.3469	520.43398
Correlation_0.7	61.00	2916.00	419.6327	516.18293

N = 49

表 14 Region 分群敘述性統計

下表(如表 15)為 5 種相似性測量，單純考量最適閾值之個別書櫃組件數，與使用者自我評量多樣性間相關的表現，依據與使用者自評之閱讀偏好多樣性之顯著性升冪排序。以使用者自評之閱讀偏好多樣性為標準，Interminus 在閾值 0.5 之下呈顯著正相關 ($r=.287, p=.046$)。

然而，與書籍總數最顯著正相關的 3 種測量方式，與使用者自評的相關性表現最差。因此，可推測不適合單純以書籍總數來預測使用者的閱讀偏好多樣性。

SPSS 執行程序：分析>相關>雙變數

	使用者自評 r	使用者自評 p	書籍總數 r	書籍總數 p
Interminus_0.5	.287*	0.046	.562**	0
Inter_0.5	-0.19	0.191	-0.135	0.355
Cosine_0.7	0.052	0.723	.984**	0
Correlation_0.7	0.051	0.729	.985**	0
Jaccard_0.5	0.037	0.801	.978**	0

* $P<.05$, ** $P<.01$

表 15 Region 分群相關性分析

2. Block (bi-component)

UCInet 執行程序：Network>Regions>Bi-component(Block)

Block 在移除網絡中的橋後，計算網絡內的組件數。其中，被移除的橋(bridge)

與其連結的點另外算為一個組件。此外，此種方法計算組件時，會排除沒有任何連線的點（isolate），可能造成套入最適閾值之後產生多個獨立點的書櫃組件數急遽減少，考量獨立點在個別書櫃中存在的比例，可能代表閱讀偏好較多樣。

由於排除獨立點的特性，於此分析方法中除了排除 109 號書櫃外，亦排除加入最適閾值之後，最後只剩下獨立點而無任何組件數的 114、134 書櫃，共計分析 47 個書櫃。敘述性統計結果顯示（如表 16），各種相似性測量的最大值相較 Region 計算組件少一半，平均數與標準差也一併減少。

SPSS 執行程序：分析>敘述統計>描述性統計量

	最小值	最大值	總和	平均數	標準差
Jaccard_0.5	1	240	2016	42.89	52.660
Inter_0.5	1	93	627	13.34	24.262
Interminus_0.5	1	251	2687	57.17	45.421
Cosine_0.7	1	229	1985	42.23	51.369
Correlation_0.7	4	235	2122	45.15	52.584

N = 47

表 16 Block 分群敘述性統計

相關性分析結果顯示（如表 17），沒有任何一種方法接近顯著。

SPSS 執行程序：分析>相關>雙變數

	使用者自評 r	使用者自評 p	書籍總數 r	書籍總數 p
Inter_0.5	0.12	0.423	-0.205	0.167
Interminus_0.5	0.045	0.764	.331*	0.023
Correlation_0.5	0.03	0.839	.898**	0
Cosine_0.7	0.014	0.926	.892**	0
Jaccard_0.7	0.013	0.929	.896**	0

*P<.05, **P<.01

表 17 Block 分群相關性分析

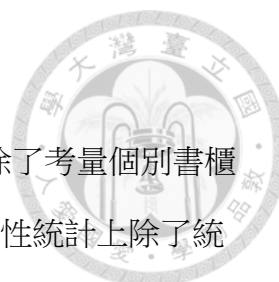
3. Girvan-Newman

UCInet_Netdraw 執行程序：Analysis>Subgroups>Girvan-Newman



在 Girvan-Newman 中,其概念為在不斷移除 edge betweenness 值最高邊(edge)過程中,計算網絡的品質(過程詳第三章 Girvan-Newman 分群),而本研究即在求網路品質最佳化下計算個別書櫃的組件數,作為個別書櫃的閱讀偏好多樣性依據。分析結果為成對的組件數與 Q 值(亦即網絡品質指標),研究中紀錄最佳 Q 值的組件數,然而分群數量在 Q 值極大化下不一定是特定的組件數,有時會有多個或是一定範圍內的組件數(數值不一定連續)。然而,Q 值極大化的多個分群數中,由於最佳組件數的數值不一定連續,不能採用平均數的方式計算最佳分群數。因此,本研究在處理非單一最佳組件數的書櫃時,採用 Q 值最大化之最大與最小的最佳組件數,下列結果亦分成最佳組件最小值與最大組件最大值作分析。

在分析取樣上,去除缺乏使用者自評之閱讀偏好多樣性之 109 號書櫃。由於此種組件計算方式必須計算 edge betweenness,計算較前兩種方式繁複,在本研究中個別書櫃書籍總數超過 3000 的書櫃計算上無法負荷,必須去除 48、129 號 2 個書櫃;而在書籍總數過少的情形下,cosine、correlation 兩種相似性計算因其最適閾值為 0.7,造成書櫃中所有書籍全部分離,網絡中皆為獨立點而無任何連結的書櫃無法計算 edge betweenness,去除 114、134 號 2 個書櫃。此外,在 Inter_0.5 與 Interminus_0.5 兩種相似性計算下,個別書櫃書籍總數在 1000 以上亦無法正常計算,去除 55、72、119、122、131 號 5 個書櫃。最終在去除 10 個書櫃的情況下,不論是取組件數最大或最小值,取樣 40 個書櫃作分析。



(1) 最佳組件最小值

在此種組件計算之下，在分析個別書櫃的閱讀偏好多樣性時，除了考量個別書櫃的組件數外，也應該考量代表網絡品質的 Q 值。因此，在描述性統計上除了統計個別相似性計算下的組件數值外，亦將各相似性計算相對的 Q 值納入統計判別考量。網絡品質最佳化下，取得最佳組件數範圍內的最小值之敘述性統計結果顯示（如表 19）Interminus_0.5 相對應之 Q 值分布最小值沒有負數且標準差小，這代表在此相似性計算之下，取樣的書櫃分群網絡品質穩定。此外，觀察 Inter_0.5 與 Interminus_0.5 在 Q 值的總和、平均數與標準差，可看出 Interminus 相似性計算在本研究的案例中，的確可有效排除於研究方法所述之書籍共現特殊情形，改善書籍共現的相似性計算。

SPSS 執行程序：分析>敘述統計>描述性統計量

	最小值	最大值	總和	平均數	標準差
Jaccard_0.5	2.000	104.000	1048.000	26.20000	21.073984
Inter_0.5	1.000	646.000	3409.000	85.22500	112.229958
Interminus_0.5	1.000	259.000	1518.000	37.95000	48.115220
Cosine_0.7	2.000	104.000	1026.000	25.65000	20.850567
Correlation_0.7	4.000	103.000	1011.000	25.27500	20.566325

N = 40

表 18 Girvan-Newman 最佳分群最小值分群數敘述性統計

	最小值	最大值	總和	平均數	標準差
Jaccard_0.5	-.500	.948	18.631	.46577	.384341
Inter_0.5	-.095	.839	6.674	.16685	.245731
Interminus_0.5	.000	.874	19.440	.48600	.215311
Cosine_0.7	-.500	.946	19.814	.49535	.370230
Correlation_0.7	-.088	.947	23.732	.59330	.296542

N = 40

表 19 Girvan-Newman 最佳分群最小值 Q 值敘述性統計

相關性分析結果顯示(如表 20),除 Inter_0.5 為不顯著正相關($r = 0.11, p = .5$)外,其餘相似性計算皆為不顯著負相關。此外,依據顯著性做排序,Cosine、Correlation、Jaccard 殿後的結果與前述兩種分群方式結果相似,相似性計算與使用者自評相關顯著 p 值愈小,與書籍總數之相關顯著 p 值則愈大。

SPSS 執行程序：分析>相關>雙變數

	使用者自評 r	使用者自評 p	書籍總數 r	書籍總數 p
Interminus_0.5	-0.188	0.245	.483**	0.002
Inter_0.5	0.11	0.5	.420**	0.007
Cosine_0.7	-0.022	0.892	.776**	0
Jaccard_0.5	-0.013	0.938	.784**	0
Correlation_0.7	-0.013	0.938	.827**	0

*P<.05, **P<.01

表 20 Girvan-Newman 分群最小值相關性分析

(2) 最佳組件最大值

SPSS 執行程序：分析>敘述統計>描述性統計量

不論最佳組件選擇最小或最大值分析,皆取 Q 值最大化的結果,因此 Q 值在敘述性統計下的結果一致(如表 22)。然而在不同相似性組件數值統計結果顯示,在最佳組件最大值的情況下,各個相似性計算下的組件數標準差皆上升。

	最小值	最大值	總和	平均數	標準差
Jaccard_0.5	2.000	116.000	1068.000	26.70000	22.379249
Inter_0.5	1.000	646.000	3460.000	86.50000	112.468799
Interminus_0.5	1.000	259.000	1583.000	39.57500	50.347962
Cosine_0.7	2.000	117.000	1043.000	26.07500	22.203473
Correlation_0.7	4.000	103.000	1015.000	25.37500	20.614518

N = 40

表 21 Girvan-Newman 最佳分群最大值分群數敘述性統計

	最小值	最大值	總和	平均數	標準差
Jaccard_0.5	-.500	.948	18.631	.46577	.384341
Inter_0.5	-.095	.839	6.674	.16685	.245731
Interminus_0.5	.000	.874	19.440	.48600	.215311
Cosine_0.7	-.500	.946	19.814	.49535	.370230
Correlation_0.7	-.088	.947	23.732	.59330	.296542
N = 40					

表 22 Girvan-Newman 最佳分群最大值 Q 值敘述性統計

相關性分析結果顯示(如表 23),除 Inter_0.5 為不顯著正相關($r = 0.112$, $p = .492$)外,其餘相似性計算皆為不顯著負相關。

SPSS 執行程序：分析>相關>雙變數

	使用者自評 r	使用者自評 p	書籍總數 r	書籍總數 p
Interminus_0.5	-0.177	0.273	.496**	0.001
Inter_0.5	0.112	0.492	.428**	0.006
Cosine_0.7	-0.016	0.92	.760**	0
Correlation_0.7	-0.008	0.959	.828**	0
Jaccard_0.5	-0.008	0.962	.779**	0

* $P < .05$, ** $P < .01$

表 23 Girvan-Newman 分群最大值相關性分析

二、個別書櫃作者類型之多樣性指標



此部分研究結果與前述不同，僅觀察個別書櫃中所含作者的比例與類型，利用多樣性指標計算個別書櫃的閱讀偏好多樣性，研究結果仍需與使用者自評之閱讀偏好多樣性結果比較，因此在取樣上仍需排除缺少使用者自評多樣性資料的 109 號書櫃，共計分析 49 個書櫃。

敘述性統計結果顯示，Simpson's Diversity 與 Stirling's Diversity_cosine 的標準差明顯較小，而 Gini 的值在各項統計數值與標準差都偏大，平均數超過 1 (mean = 6.962531, S.D = 1.1798836)。

SPSS 執行程序：分析>敘述統計>描述性統計量

	最小值	最大值	總和	平均數	標準差
author_diversity	.8394	.9993	48.1134	.981906	.0244141
author_Gini	.0899	.6026	16.7441	.341716	.1443659
author_Entropy	4.0692	10.1405	341.1640	6.962531	1.1798836
author_diversity_cosine	.6637	.9510	36.1558	.737874	.0460629

N = 49

與使用者自評多樣性之相關性結果顯示，除了 Stirling's Revised Diversity_cosine 為不顯著正相關外，其餘多樣性指標皆顯示為不顯著負相關。

SPSS 執行程序：分析>相關>雙變數

	使用者自評 r	使用者自評 p	書籍總數 r	書籍總數 p
author_diversity	0.143	0.328	0.199	0.171
author_Gini	-0.024	0.87	.283*	0.049
author_Entropy	-0.015	0.921	.764**	0
author_diversity_cosine	-0.012	0.936	0.164	0.26

*P<.05, **P<.01

過去研究 (Mikkonen & Vakkari, 2012; Tang et al., 2012) 認為作者是重要的選書因素，然而利用作者作為多樣性指標計算依據時，或許應考慮作者間的相似性。

同樣包含 20 種作者的兩個書櫃，若其中一個書櫃的作者彼此間相似度較高，都是推理小說的作者；另一書櫃的作者彼此間相似度較低，包含商管、小說、旅遊等類別，則後者會是閱讀偏好多樣性較高的書櫃。



第二節 綜合討論

一、最適閾值的選擇策略

本研究選擇最適閾值的優先參考依據為最大組件的變化。在不同的相似性計算下，整體書籍共現網絡的連結會在閾值的逐步上升中逐漸減少。在網絡連結減少的過程中，節點之間會逐漸依據網絡特性分解成不同的組件。而最大組件在本研究的主題下代表的則是個別書櫃擁有者的主要閱讀偏好，因此最大組件的數量、在閾值增加時組件大小減少的模式，可代表網絡在特定閾值下的分群情形，本研究即依據此種概念與過去最適閾值相關研究作法結合，以最大組件減少的情形做為判斷最適閾值的主要因素。

然而，判斷的模式與 Small (2009) 的研究不同。在 Small 的研究中，從最大閾值出發，選擇最大組件在閾值的減少下增加的變化量最大的前一刻作為最適閾值；在本研究中，由於休閒閱讀與學術研究的場域不同，將引用學術文章的行為一般認為比書籍加入網路書櫃來的謹慎。此外，閾值的採用最終目標是讓網絡變得適合分群，套入最適閾值的書籍共現網絡猶如修剪過的樹木，會變得整齊且層次分明，書櫃則會變得分群清晰。因此，本研究從最小閾值出發，選擇最大組件在閾值的上升下減少的變化量最大的前一刻作為最適閾值，假定那時候的網絡是最適合分群的。

由於兼併參考林頌堅(2010)研究中選擇最適閾值的作法，除採用上述策略，也一併將網絡密度、最大組件所占比例、平均連結線數作為選擇最適閾值的參考因素。在實際處理最大組件與閾值資料的過程中，我們發現選擇最大組件減少變

化量最大的前一刻時，若進一步增加閾值，在不同的相似性計算下，最大組件減少的情形幾乎呈現停滯的狀況，這代表書籍共現網絡的相似值幾乎在下一個增加的閾值以上，因此閾值繼續上升幾乎不具意義。

綜上述，在 Small (2009) 與本研究於最大組件與閾值的研究結果可發現，兩種利用最大組件判斷最適閾值的方法，推測可做為有效閾值的界線。在本研究中，最大組件減少最多的前一刻為閾值的上界，閾值繼續上升的結果對最大組件的大小幾乎不產生影響。而對於判斷閾值的最佳方法，考慮組件的變化量與其他網絡特性，應可大略推估最適閾值的範圍。

二、相似性測量於書籍相似性計算的表現

我們藉由相似性測量在個別書櫃的分群表現判別 5 種相似性測量的優劣，其中 *Interminus* 為表現最好的相似性測量。書籍共現網絡是否能成功地呈現使用者的閱讀偏好多樣性，其中包含兩個關鍵，一是相似性測量的選擇，二是閾值的選擇。相似性測量與閾值在呈現多樣性的過程中，廣義而言皆扮演對書籍網絡連結進行去蕪存菁的角色。適合書籍共現網絡的相似性測量，必須考量書籍共現的實際情形，修正不同相似性測量計算方式所造成的誤差。而閾值的選擇，除了需要考量書籍共現的實際情形外，尚須權衡不同相似性測量的數值特性。

在 5 個相似性測量中，*Jaccard*、*Inter*、*Interminus* 為較相近的 3 種相似性測量，其原理皆為以兩本書的共同出現作為基礎，共現次數愈多代表愈相似，接著利用兩本書在個別書櫃的出現次數正規化。其中，*Jaccard* 的分母為任意兩本書於個別書櫃出現次數的聯集；*Inter* 與 *Interminus* 則皆採任意兩本書於個別書櫃出現次數的最小值，不同點在於 *Interminus* 將分子分母個別減一。*Cosine* 則把任意兩本書於個別書櫃的出現情形視為向量，夾角愈小則兩本書愈相似，*Correlation* 比較兩本書於個別書櫃的出現情形，於不同書櫃的出現情形愈有相關則愈相似。

深究 Interminus 相似性測量的表現，採用最小值作為正規化，能夠提升暢銷書與其他書籍相似性低落的表現，而減一的動作可避免兩本冷門的書因巧合或隨機出現所造成的計算偏差。綜合兩種優點，是 Interminus 於書籍相似性表現較好的原因。在實際統計上，在抽樣 21199 本書中，只被一個書櫃擁有的書籍有 970 本，約為整體抽樣書籍的 4%。上述書籍容易被少部分讀者擁有，如特定領域的作品集，而它們出現在少數的書櫃中，這些書籍在 Interminus 的計算下，於個別書櫃顯示為獨立的點，視為特定讀者的多元閱讀偏好之一

綜上述，在多數相似性測量之下，兩本書冷門，但恰好共同出現在使用者書櫃時，就容易產生指標的偏差。這種因巧合或隨機共同出現，並不代表兩本書真的相似。而 Interminus 即是控制了這種偏差，因此產生較接近真實的相似性。

三、最適閾值下個別書櫃的分群表現

綜合各種相似性測量配合最適閾值，與使用者自評多樣性之相關性分析結果顯示，Interminus_0.5 為本研究採用之所有相似性測量與閾值配對中表現最好的。在休閒閱讀的情境下，一般認為閱讀書籍的多寡與閱讀多樣性相關，因此在相關性分析時將書籍數量作為相關性分析的變數之一。結果顯示，與使用者自評多樣性相關分析結果愈好的相似性測量，與書籍數量相關性分析結果愈差。為檢測書籍數量在預測閱讀偏好多樣性時是否具有解釋力，並進一步分析使用者自評多異性、書籍數量與相似性測量間的關係，研究中最後嘗試將閱讀偏好多樣性、研究中所計算閱讀偏好多樣性、網路書櫃書籍數量等因素放入簡單線性回歸模型。

若以使用者自評之多樣性作為預測的依變數，利用逐步線性回歸檢視書籍相似性之多樣性，以及個別書櫃書籍總數兩個變數的解釋力。結果顯示書籍總數為被排除的變數，兩變數（書籍總數與書籍相似性之多樣性）呈顯著正相關（ $r=.559$, $p=.000$ ），書籍相似性之多樣性與使用者自評之多樣性呈顯著正相關（ $r=.286$,

p=.022)，書籍總數為被排除的變數。以上結果顯示雖然個別書櫃總數對多樣性計算有影響，但不足以直接解釋使用者自評之多樣性。因此，若須得知使用者之閱讀偏好多樣性，無法單純觀察使用者書櫃的書籍總數而推斷其閱讀偏好多樣性，而研究中採用的書籍相似性則可有效的利用於分析使用者閱讀偏好多樣性。

三、利用個別書櫃作者類型計算之多樣性指標

在本研究中，採用多樣性指標的目的在於兼顧 Rafols & Meyer (2010) 中的 3 個多樣性組成元素：類別種類數量、個別類別所佔比例、類別相似性，最終採用 4 種多樣性指標，Simpson's Diversity 將種類數量、比例做為考量；Gini 與 Shannon's Entropy 則著重種類間分布不均的情形；Stirling's Revised Diversity 則兼併 3 個多樣性組成元素，同時也是本研究中 4 個指標中唯一考量類別相似性的多樣性指標。

在多樣性指標的良莠比較上，由於過去研究結果 (Mikkonen & Vakkari, 2012; Tang et al., 2012) 顯示作者為重要的選書因素，又在本研究中之研究場域 aNobii 中，書籍類別為普遍認為不有效的選書因素，因此在多樣性指標的類別（即多樣性的「樣」）選擇作者為多樣性中的類別。選擇作者作為類別，亦可利用本研究中的書籍共現網絡，在進一步辨識書籍的作者的同時，在取得作者所有的作品後建立作者的共現網絡，如此一來便能建立作者的相似性網絡。

將 4 種多樣性指標的個別書櫃結果與使用者自評之閱讀偏好多樣性做相關性分析，結果顯示兼併 3 個多樣性組成元素的 Stirling's Revised Diversity 表現最好，然而此研究結果亦有精進的空間，本研究採用 Cosine 作為計算作者相似性的相似性測量方法，作為作者相似性在多樣性指標有效性的初探，研究結果證明相似性測量能改善多樣性指標計算的準確度。

第五章 結論與建議

第一節 結論



本研究之目的在探索利用非干擾式書櫃資料收集，呈現使用者閱讀偏好的方法。依據研究流程與研究問題，本研究結果大致上可分為書籍共現情形之多樣性指標，以及以作者作為類別之多樣性指標，結論依據上述兩種方式分述。

一、書籍共現網絡分群之閱讀偏好多樣性

(一) 書籍共現網絡的最適閾值選取

於最佳閾值的選取方式，本研究參考 Small (2009) 與林頌堅 (2010) 的研究，採用最大組件變化量最大的前一個閾值點作為最佳閾值。研究結果顯示，於本研究採用的 5 種相似性測量中，採用最大組件減少最多的前一個閾值點作為最適閾值，可取得閾值的停頓點。此外，若在選取閾值參考其他網絡特性，可進一步藉由平均連結線數與網絡密度，判斷套入閾值後的網絡是否過度稀疏，造成多數書櫃是否有過多的單獨點產生 (isolate)。

(二) 書籍分群多樣性與使用者自評多樣性之關聯

本研究中總共採用 5 種相似性測量，其中 Jaccard、Cosine 與 Correlation 都是常見的相似性測量方式。然而，研究結果顯示，在本研究的場域內，Interminus_0.5 卻是表現最好的相似性測量演算法。Interminus 是研究中觀察書籍共現情形後，為修正特定兩本書櫃僅共同出現在同一個書櫃 1 次，使用 Inter 相似性測量所計算出的相似性卻是 1 的這種情形。上述這種偶然共同出現在同一個書櫃的情形其實很常見，但是這種書籍共現的情形不代表這兩本書非常相似，而是因為若以出

現的書櫃作為這兩本書的特徵，他們都正好只有出現在一個書櫃的特徵，造成相似性測量中分母分子皆為一的情形發生。因此，進行相似性測量相關的研究時，應選擇適合研究場域的相似性測量作為研究工具。



二、個別書櫃作者類型之多樣性指標

本研究採用多樣性指標的方法，同時考量三個多樣性的特性，即種類數量愈多、種類比例愈不平均、種類相似度愈低，則多樣性程度愈高。然而，本研究選用的 4 個指標中，唯有 Stirling's Revised Diversity 同時考量多樣性的三個特性，計算種類相似性時，研究中僅採用 Cosine 作為相似性測量的方法。過去休閒閱讀的研究(Mikkonen & Vakkari, 2012; Tang et al., 2012)顯示作者是讀者重要的選書因素，然而最終研究結果顯示所有的多樣性指標皆與使用者自評之多樣性無顯著相關，依據顯著性排序 Stirling's Revised Diversity 與使用者自評的閱讀偏好多樣性相關性最為顯著。因此，若未來採用更多適合計算作者相似性的測量方法，應能有效改善以作者類型計算之多樣性指標。

第二節 建議

本研究於選定閾值時，主要依據最大組件的大小變化。然而，若要更精細的選擇最適閾值，可利用二分逼近法增加閾值的精度，逐步逼近最適閾值。研究之相關性分析結果為特定相似性測量搭配研究中選擇的最適閾值，若要進一步測試閾值與相似性測量之間的交互關係，可固定閾值或相似性測量，反覆驗證不同閾值與相似性測量與使用者自評之閱讀偏好多樣性之間的相關性，並可藉由不同相似性測量與閾值之修正逼近使用者真實的閱讀偏好多樣性。

此外，相似性測量為本研究中呈現使用者閱讀偏好的根本。於網路書櫃的場域，研究中僅採用 5 種相似性測量，與 3 種分群方式，作為書籍共現網絡呈現使用者偏好的初探。在分析個別書櫃的分群結果時，為確認不同分群方法的特性，

過程中選擇多種特性的書櫃，如獨立點多的書櫃、分群清晰的書櫃做分群方法特性的確認。由於分群結果為個別書櫃多樣性之依據，因此確認分群方式的特性在本研究中尤其重要。

綜言之，利用書籍共現呈現使用者閱讀偏好建立在相似性測量的準確性、閾值與分群的適切性，因此於三個流程中都需要仔細的檢查抽樣書櫃資料是否有異常。研究結果顯示此種非干擾式的方法對於使用者自評之閱讀偏好多樣性有解釋力，而作者類型之多樣性指標在同時考量類別數量、類別比例與類別相似性能讓相似性測量更準確。

第三節 進一步研究之建議

本研究僅專注於如何藉由相似性測量，以及將類別相似性加入多樣性指標，呈現使用者的閱讀偏好多樣性，對於其他影響閱讀多樣性之因素，並未深入探究。然而，應有閱讀背景因素如閱讀書籍總量、讀者面對閱讀的態度等因素皆可能對閱讀偏好多樣性產生影響。例如，過去一般認為書櫃擁有者之書籍總數，與閱讀偏好多樣性有關。

研究資料處理過程中，初步觀察個別書櫃書籍總數與分群數之數值分布，發現書籍總數愈多的書籍，其閱讀偏好多樣性可能愈高。在進一步針對本研究計算之相似性測量之分群結果、使用者自評之閱讀偏好多樣性與個別書櫃書籍總數做相關分析，結果卻發現與使用者自評多樣性相關性愈顯著之相似性測量，與書籍總數的相關性顯著性愈低。

此外，實際將研究中之分群數與個別書櫃書籍總數，作為預測使用者自評之多樣性的變數，簡單線性回歸結果顯示書籍總數違背排除的變數，而相似性測量產生的分群數與使用者自評之多樣性呈現顯著正相關。上述結果可發現，我們無法直接利用書籍總數作為判斷使用者閱讀偏好多樣性的依據，儘管一般而言書籍總數愈多，閱讀偏好多樣性也愈高。

關於使用者書櫃書籍總數的多寡，除了可能受使用者閱讀偏好多樣性的影響，在網路服務中，加入網路書櫃服務時間愈長，書籍總數也會愈多。因此，閱讀的背景因素暫時沒有直接與閱讀偏好多樣性呈現高度相關的現象，其與閱讀偏好多樣性之交互作用卻值得進一步探究。閱讀偏好多樣性與其他閱讀相關背景因素之研究，可進一步研究其他因素是否能加入增加解釋使用者自評多樣性的能力，例如閱讀對讀者的重要性、使用網路書櫃的頻率。

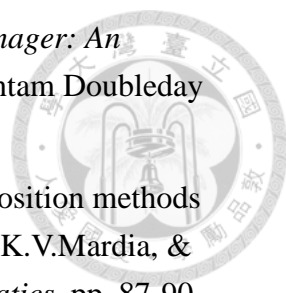
參考文獻



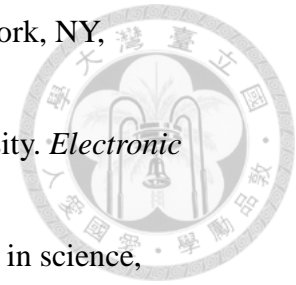
一. 西文部分

- Aiello, L. M., Barrat, A., Cattuto, C., Ruffo, G., & Schifanella, R. (2010, 20-22 Aug. 2010). *Link Creation and Profile Alignment in the aNobii Social Network*. Paper presented at the Social Computing (SocialCom), 2010 IEEE Second International Conference on.
- Armstrong, A., & Hagel Iii, J. (1996). The real value of on line communities. *Harvard Business Review*, 74(3), 134.
- Bell, Robert M., & Koren, Yehuda. (2007). Lessons from the Netflix prize challenge. *SIGKDD Explor. Newsl.*, 9(2), 75-79. doi: 10.1145/1345448.1345465
- Bennett, J., & Lanning, S. (2007). *The Netflix Prize*. Paper presented at the Proceedings of the KDD Cup Workshop 2007, New York. <http://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf>
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25(3), 187-217. doi: 10.1086/209535
- Buldú, Javier M, Cano, P, Koppenberger, M, Almendral, Juan A, & Boccaletti, S. (2007). The complex network of musical tastes. *New Journal of Physics*, 9(6), 172.
- Dieberger, A., Dourish, P., Höök, K., Resnick, P., & Wexelblat, A. (2000). Social navigation: techniques for building more usable systems. *interactions*, 7(6), 36-45. doi: 10.1145/352580.352587
- Dourish, P., & Chalmers, M. (1994). Running Out of Space: Models of information navigation *HCI, Glasgow*.
- Franke, Nikolaus, Keinz, Peter, & Steger, Christoph J. (2009). Testing the Value of Customization: When Do Customers Really Prefer Products Tailored to Their Preferences? *Journal of Marketing*, 73(5), 103-121. doi: 10.1509/jmkg.73.5.103
- Hanneman, R.A., & Riddle, M. (2005). *Introduction to Social Network Methods*: University of California.
- Haubl, G., & Murray, K. B. (2003). Preference construction and persistence in digital

- marketplaces: The role of electronic recommendation agents. *Journal of Consumer Psychology*, 13(1-2), 75-91. doi: 10.1207/153276603768344807
- Holbrook, Morris B., & Hirschman, Elizabeth C. (1982). The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun. *Journal of Consumer Research*, 9(2), 132-140. doi: 10.2307/2489122
- Hopper, Rosemary. (2005). What are teenagers reading? Adolescent fiction reading habits and reading choices. *Literacy*, 39(3), 113-120. doi: 10.1111/j.1467-9345.2005.00409.x
- Kari, Jarkko, & Hartel, Jenna. (2007). Information and higher things in life: Addressing the pleasurable and the profound in information science. *Journal of the American Society for Information Science and Technology*, 58(8), 1131-1147. doi: 10.1002/asi.20585
- Klavans, Richard, & Boyack, Kevin W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology*, 57(2), 251-263. doi: 10.1002/asi.20274
- Kramer, T. (2007). The effect of measurement task transparency on preference construction and evaluations of personalized recommendations. *Journal of Marketing Research*, 44(2), 224-233. doi: 10.1509/jmkr.44.2.224
- Kwon, Kwiseok, Cho, Jinhyung, & Park, Yongtae. (2009). Influences of customer preference development on the effectiveness of recommendation strategies. *Electronic Commerce Research and Applications*, 8(5), 263-275. doi: <http://dx.doi.org/10.1016/j.elerap.2009.04.004>
- Lancaster, F.W. (1998). *Indexing and abstracting in theory and practice*: Library Association Publishing.
- Leydesdorff, Loet, & Vaughan, Liwen. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, 57(12), 1616-1628. doi: 10.1002/asi.20335
- Mikkonen, Anna, & Vakkari, Pertti. (2012). Readers' search strategies for accessing books in public libraries *4th Information Interaction in Context Symposium* (pp. 214-223).
- Pejtersen, Annelise Mark, & Austin, Jutta. (1983). FICTION RETRIEVAL: EXPERIMENTAL DESIGN AND EVALUATION OF A SEARCH SYSTEM BASED ON USERS' VALUE CRITERIA (PART 1). *Journal of Documentation*, 39(4), 230-246. doi: 10.1108/eb026750

- 
- Peppers, D., Rogers, M., & Rogers, M. (1999). *The One to One Manager: An Executive's Guide to Custom Relationship Management*: Bantam Doubleday Dell Publishing Group.
- Pinney, J.W. & Westhead, D.R. (2006). Betweenness-based decomposition methods for social and biological networks In S. Barber, P.D. Baxter, K.V.Mardia, & R.E. Walls (Eds.), *Interdisciplinary Statistics and Bioinformatics*, pp. 87-90. Leeds, Leeds University Press.
- Preece, J. (2001). Sociability and usability: Twenty years of chatting online. *Behavior and Information Technology Journal*(20), 347--356.
- Purvis, Andy, & Hector, Andy. (2000). Getting the measure of biodiversity. *Nature*, 405(6783), 212-219.
- Rafols, Ismael, & Meyer, Martin. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287. doi: 10.1007/s11192-009-0041-y
- Ross, Sheldrick Catherine. (1999). Finding without seeking: the information encounter in the context of reading for pleasure. *Information Processing & Management*, 35(6), 783-799.
doi: [http://dx.doi.org/10.1016/S0306-4573\(99\)00026-6](http://dx.doi.org/10.1016/S0306-4573(99)00026-6)
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3), 75-174.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159-169. doi: 10.1016/j.jretai.2004.04.001
- Simonson, Itamar. (2005). Determinants of Customers' Responses to Customized Offers: Conceptual Framework and Research Propositions. *Journal of Marketing*, 69(1), 32-45. doi: 10.1509/jmkg.69.1.32.55512
- Small, H., & Sweeney, E. (1985). Clustering the science citation index ® using co-citations. *Scientometrics*, 7(3-6), 391-409. doi: 10.1007/BF02017157
- Small, Henry. (2009). Critical thresholds for co-citation clusters and emergence of the giant component. *Journal of Informetrics*, 3(4), 332-340.
doi: <http://dx.doi.org/10.1016/j.joi.2009.05.001>
- Spertus, Ellen, Sahami, Mehran, & Buyukkokten, Orkut. (2005). *Evaluating similarity measures: a large-scale study in the orkut social network*. Paper presented at the KDD '05: Proceeding of the eleventh ACM SIGKDD international

conference on Knowledge discovery in data mining, New York, NY, USA. <http://dx.doi.org/10.1145/1081870.1081956>



- Stirling, Andrew. (1998). On the Economics and Analysis of Diversity. *Electronic Working Papers Series*, 28.
- Stirling, Andy. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, 4(15), 707-719. doi: 10.1098/rsif.2007.0213
- Svensson, Martin, Höök, Kristina, Laaksolahti, Jarmo, & Waern, Annika. (2001). *Social navigation of food recipes*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, Washington, United States.
- Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest & Clifford Stein. (2009). *Introduction to Algorithms* (pp. 621-622). Massachusetts, MA: MIT press.
- Tang, Muh-Chyun, Ting, Pei-Hang, & Sie, Yi-Jin. (2012). *Exploring evaluation criteria of social navigational tools on social media: a case study of aNobii*. Paper presented at the Proceedings of the 4th Information Interaction in Context Symposium, Nijmegen, The Netherlands.
- Taylor, Robert S. (1962). The process of asking questions. *American Documentation*, 13(4), 391-396. doi: 10.1002/asi.5090130405
- Xu, Yunjie. (2007). Relevance judgment in epistemic and hedonic information searches. *Journal of the American Society for Information Science and Technology*, 58(2), 179-189. doi: 10.1002/asi.20461

二. 中文部分

- 丁培涵. (2011). 網路書櫃使用者偏好結構與瀏覽尋書行為之研究. (碩士), 臺灣大學. Available from Airiti AiritiLibrary database. (2011 年)
- 李書萍. (2012). 網路書評發表者之書寫動機、平台選擇與使用經驗探討. (碩士), 臺灣大學. Available from Airiti AiritiLibrary database. (2012 年)
- 林珊如, 劉應琳. (2001). 休閒閱讀找書策略與影響因素之探討：以台大 BBS Books 版愛書人為例. 資訊傳播與圖書館學, 8(2), 23-37.
- 林頌堅. (2010). 以詞語共現網絡分析探勘資訊傳播學領域的研究主題與關係. 圖書資訊學研究, 4(2), pp.123-148.
- 謝宜瑾. (2012). 從透鏡模式探討影響讀者尋書滿意度之因素—以 aNobii 為例. (碩士), 臺灣大學. Available from Airiti AiritiLibrary database. (2012 年)