

## JLISR Open Peer Review Report

### Reviewer Comments and Rebuttal to the Comments

#### Open Point 開放觀點：開放式同儕評閱機制

本刊新採行「開放觀點」（Open Point）機制，以留存作者與評閱者之間論證的寶貴文字紀錄，並鼓勵雙方在同意公開大部分的評論意見與審查回覆。希冀透過「開放式同儕評閱」（Open Peer Review）模式，提供學者窺見已被刊登文章背後，許多同樣值得被理解與被引用的觀點。這項機制有助於我們的作者、評閱者、讀者享有更真實的學術傳播精華。

# 圖書資訊學研究

JOURNAL OF LIBRARY AND INFORMATION SCIENCE RESEARCH

審查文章：檔案文本探勘研究：以特種檔案為例

審稿者：

匿名評閱者 A (\*僅公開評閱意見)

匿名評閱者 B (\*僅公開評閱意見)

林頌堅

主編綜評

作者：邵僅、林巧敏

刊登卷期：20卷1期（2025年12月）

DOI：10.30177/JLISR.202512\_20(1).0002

說明：◎評閱者；★主編；●作者

## 審查階段

### 初審

審稿者：匿名審稿者 A

#### 評閱意見：

本文以數位科技進行歷史與檔案研究，利用數種不同功能的數位程式與方法來解析檔案，輔以視覺化的方式呈現結果，冀能發掘特種檔案中無法以人工直觀的重要資訊，並結合質性說明，為檔案內容的分析與詮釋提供新的方法與應用模式，進而為未來的檔案學與史學研究提供啟發，其研究方法別具創意，令人耳目一新。然而，數位人文研究的精神在於以數位方式研究人文或社會科學議題，解決人工無法處理或是發現以往無法察覺的問題，並佐以史料進行探討論證，最終做出合理的歷史批判和詮釋。可惜的是，本文較偏重於敘述如何利用新科技來統計運算，以及如何將計算結果進行視覺化之展現，卻沒能讓成果落實在歷史研究的深度裡頭。由於缺乏聚焦的主題，本文未能將研究所獲融入某個歷史議題進行妥適的深論剖析，以致推出的結論顯得散漫零亂，看不出解答了什麼歷史問題，也無從得知透過程式運算所獲之資訊到底有何具體文獻價值或學術意義。同時，全文幾乎沒有運用任何史料來佐證論述，不知僅憑程式運算、統計發生次數的歷史研究方法可能會悖離歷史真實，更遑論只利用檔案作為探索歷史真相的唯一史料，也非常容易造成扭曲及誤解。以上，作者皆未留意，致使本文的結論與發現流於空泛，在史學領域的意義與價值令人質疑，故含金量不高，對歷史研究的貢獻度相當有限。以下針對本文研究結果提出一些問題，供作者參考：

1. 頁 11-14，「一、核心詞彙分析」：不論是人物或組織，請問計算特種檔案當中兩者出現的次數頻率高低，到底代表什麼意義呢？請作者直接指出並清楚說明。例如人物經軟體計算後，發現吳鐵城的詞頻數值最高，甚

至超越蔣中正，但詞頻數值最高真正能呈顯出什麼現象呢？作者對此未加解釋，抑或是作者以此數據認為吳鐵城是當時國民黨內最活躍的核心人物，其核心地位甚至高過蔣中正或其他政壇巨擘？但如此說法真的是歷史事實嗎？相信會讓許多人難以信服，而作者也有沒有再參考、列舉任何史料來深入論析，並佐證重要發現。同樣地，組織的詞頻分析也有上述之流弊，不知有何具體的歷史意義。

- 作者回覆：詞頻多寡代表人名與組織名在文本的出現頻率，並非等同活躍程度，也不試圖以頻率高低連結重要性的解釋或以數位分析結果論斷史事，此項分析在於以量化觀察文本隱含的資訊，作為史學者進一步探究的參考。

為避免閱讀者產生誤解，本文於分析結果增列文字說明分析意義。

2. 頁 14-15，「（一）人物與人物之網絡關係」：請問作者將人物與人物連結起來所依據的社會關係究竟為何？是僚屬、親友、同學、同事、師徒、同鄉或是政敵、世仇、競爭對手、當事人、黨外人士……，還是什麼呢？其實，不同社會關係產生的連結，意義可能會各不相同，作者實在不宜將各種社會關係混為一談，無差別地作為結點之間連結的基礎。同時，僅透過特種檔案，經社群網絡分析後所謂的核心人物，到底能具有什麼樣的歷史意義呢？所得結果真正是新發現嗎？吳鐵城經社會網絡分析結果是「最核心的人物」，在歷史上究竟應該如何給予正確的評價？而事實上真的是這樣嗎？可惜作者都沒有多加闡論或查證，以致無法令人信服。

- 作者回覆：本研究社會網絡分析期望以系統性方式呈現不同類型詞彙之文本共現關係，並結合少量內文引用或是引據該時代背景脈絡解釋，提供讀者初步了解該文本資訊架構，但最終解釋權仍屬歷史研究範疇，並非試圖以數位分析結果解釋歷史。本文已補充相對文字說明。

3. 頁 15-16，「（二）人物與事件之網絡關係」：由於每位人物與事件的涉入程度不同，因此對於事件的重要性也各自不同，不知作者如何在社群網絡分析當中加以區別呢？例如蔣介石命吳鐵城前往東北勸說張學良歸順國民政府，終使東北易幟，不知在這個歷史事件中，作者認為蔣、吳何者的關聯性比較重要呢？而在結點的連結上，又需不需要設計不同的處理方式呢？

- 作者回覆：感謝審查意見，參照前二點說明，如蔣中正、吳鐵城與其他人物之網絡展現，僅止於以數位方法呈現諸位人物在文本中的位置。為避免誤解，本文中已重新說明量化資料的詮釋界限，並建議結合質性分析進行綜合研判。

4. 頁 18-21，「三、文本內容地理資訊分布」：本節透過地理資訊分析，針對特種檔案內容記錄的地名進行統整分析，然後將結果分為四個時期呈現，並在每一時期內做教科書式的文字說明，可惜各時期的發現卻顯得簡單、表面而不夠深入。建議作者於各時期的發現說明分別參考引用各種史料，旁證補述，深入論析，而非僅是利用教科書式的說明帶過，如此恐將流於主觀，甚至過度詮釋之嫌。同時，審度每個時期內作者對於所獲成果的文字說明，多半的成果似乎也能僅靠閱讀檔案內容，利用人工直觀的方式便可察知。

綜上，本文似乎只是一篇假設性或推論性的文章，研究發現或成果實際上仍然有待查證。

- 作者回覆：感謝審查意見，經重新檢視修正地理資訊分析內容，將各分期圖像分立，並增列文字說明，以增補對於各時期地理資訊分析結果對照於檔案內容的解釋。然數位人文研究偏重於研究方法和獲致結果的呈現，對於史料內容的批判和論述，將回歸史學研究者的進一步研究。

## 審稿者：匿名審稿者 B

### 評閱意見：

此研究結合多個數位人文工具分析國民黨特種檔案，並展現分析結果。研究目的在探討「特種檔案」的重要人物、組織，以及重要人物、組織、事件等詞彙間的共現關係，並結合 GIS 系統展示「特種檔案」文本內容的地分布。本研究的主要優勢，是結合 CKIP tagger、Docusky、CORPRO 等工具，並使用社會網絡分析方式進行研究。針對本文呈現的研究結果，各項建議如下

1. 在研究程序與資料分析一節中說明權威詞表建立過程時，作者指出「仍有少數意義相同但字詞不同、無法合併計算者如『中執會』與『執委會』，則分別維持原狀。」為何在確認詞彙意義相同時，仍將詞彙分別計算？以此句為例，若「中執會」與「執委會」意義相同，是指兩個字詞代表同一組織。在這個前提下，不做合併計算會影響到組織出現頻率，從而影響到後續詞頻與社會網絡分析的結果。如果作者認為不應合併，請說明不做合併的考慮因素。

● 作者回覆：感謝審查意見，已於詞頻統計表格內容進行調整，然共現分析部分，因囿於當前分析工具功能限制，未能進行合併計算處理，因此選擇保留原始詞頻，此一限制已於本文中進行文字補充說明。此做法雖有影響呈現結果之虞，但可避免工具僅計算某一數值導致的節點共現數量與實際情況不符的情形。

2. 圖三名稱為「前 30 高頻組織詞彙文字雲」，但列出的詞彙數高於此數。檢視文字雲內容時，多個詞彙有重覆列出的情形，茲舉數例：中祕處、黨部、外交部、中央祕書處、奸偽、中共。請再確認此圖內容的正確性與呈現方式。

● 作者回覆：感謝審查意見，已針對圖片內容進行更新。

3. 在核心詞彙共現分析一節的開頭，作者說明研究中將 edge 強度 100 以下

的 edge 取 1，100 以上則取 10。但從圖 4 到圖 7 的內容中，兩種類種的 edge 並沒有容易辦識的差異，視覺化上需要再調整。此外，為何使用這個方式來轉換數值，而非保留原始數值？若是考量原 edge 數值範圍過大，亦可採用如對數轉換方式處理。目前採用的方式，是將數值變數轉為類別變數，將會大幅度弱化可由 edge 強度呈現的意義。

- 作者回覆：感謝審查意見，經嘗試不同數值，本文採目前方式與規則是為能清晰呈現每張圖片之結果。若保留原始數值，會導致共現詞彙數量、線條過多而導致難以閱讀。此項考量已於文中增列文字說明，以利於閱讀理解，

4.同樣在核心詞彙共現分析一節的開頭，說明節點數的保留方法為「總節點數大於 300 者約保留前 10%，介於 60 至 300 間者約保留前 20%，60 以下者則全數保留。採總節點數不超過 60，讓網絡清晰且利於判讀為準則。」此處作者僅說明在保留節點數，但未說明是依據節點的何種屬性進行決定。請作者在文中提供說明，以利後續欲循此模式進行的研究參考。

- 作者回覆：感謝審查意見，已修改關於核心詞彙共現分析採用節點數之說明文字，以利於理解採用之數值決定。本研究繪圖原則採用之數值係經過嘗試不同關聯強度和節點數組合，以呈現最有意義且易於閱讀之核心詞彙的共現關係為主。

5.文中在呈現各個社會網絡圖時，常有文字內容被蓋住、節點過密、edge 連結的節點不明顯等問題，請再修正這些圖的視覺化細節。除此之外，圖 6 下方有部分內容消失，請一併修正。

- 作者回覆：已修正圖像，並分立為獨立圖片，以利觀察。然部分節點因受限工具功能，位置無法完全固定，故仍有小部分遮擋，已盡可能以不影響閱讀為原則進行調整。

6.進行社會網絡分析時，研究中將 1928 至 1949 年代的所有資料建立成單

一網絡。這樣並不利於結合時空脈絡來分析檔案內容。以研究中使用的檔案年代，1928 至 1937 年間為北伐後至中日戰爭全面爆發，而 1937 年至 1940 年間的中日戰爭情形，也與美國參戰後的情形有別。1945 年後的國共內戰，依戰事的變化而有不同脈絡。目前在社會網絡分析時採不做時期區分的方式，不利於結合不同年代的背景脈絡進行解讀。相比之下，本研究在進行 GIS 分析時就有依時代進行區分，這有利於討論的深化。建議採取同樣的取徑進行社會網絡的分析與討論，這會有助於判斷特種檔案在不同時期反映的人物、組織、事件網絡變化，並更易於與已有的史學研究結果呼應，進而深度揭示與驗證此研究模式的潛力。

- 作者回覆：感謝審查意見，本研究檔案雖橫跨 1928 至 1949 年，但主要集中於 1940 至 1948 年。雖可將各時期分段呈現，但經評估與整體呈現之結果可能無太大差異，且篇幅可能大幅增加。因此，本研究選擇以整體網絡為主，結合時空脈絡進行解讀，而更細緻的動態網絡分析將列為後續研究建議。

7. 圖 8 至圖 11 等 GIS 圖的解析度不足，難以做為說明文字的輔助資料。

- 作者回覆：感謝委員意見，已調整圖示，將各圖分立和放大，以利於閱讀檢視。

8. 請再依 APA 格式，修正目前參考文獻格式。

- 作者回覆：感謝委員意見，已調整引用格式。

審稿者：林頌堅

評閱意見：

本文運用數位人文方法，結合多種自然語言處理和地理資訊工具，探勘國民黨特種檔案的文本內容。整體研究設計，技術實施亦具實證參考性。仍有若干可補強之處，若能進一步釐清資料特性並深化歷史意涵的探討，將有助於提升本研究的學術厚度與跨領域對話潛力。

## 一、文獻探討部分

文獻回顧雖涵蓋多種文本探勘技術與工具應用的研究，整體資料豐富，但在內容取捨與主題聚焦上仍可進一步精簡與補強。具體建議如下：

1. 建議刪減部分與本研究核心議題關聯性較低的技術應用文獻，特別是來自其他領域（如醫療、商業文本分析等）之案例，以避免使篇幅過長並分散論文焦點。
  - 作者回覆：感謝審查意見，已刪減非歷史背景相關之內容，並加入本研究對於文獻探討內容的連結文字。
2. 補充與研究主題之歷史背景相關的重要文獻，建立文本分析所需的時代脈絡與社會語境，有助於後續分析之詮釋力與說服力。
  - 作者回覆：感謝審查意見，參閱史料文獻後，目前於地理資訊分析內容開始前，新增時代脈絡與重要事件之文字說明，幫助理解特種檔案時期之歷史背景。
3. 文獻回顧建議不僅呈現工具應用，更應說明過往研究在哪些層面未能處理特種檔案的潛在知識，從而強化本研究的問題意識與知識落差定位。
  - 作者回覆：感謝審查意見，已於文獻回顧內容中加入對於過往研究的綜述和回應，呈現本研究相對於前行研究的價值，亦增列文

字說明數位人文研究與傳統史學家閱讀文本探究特定議題的差異。

## 二、研究方法部分

清楚說明採用之數位工具與研究過程，然仍建議補充下列內容以增進方法透明度與可再現性：

1. 增加對原始文本資料的探索性分析與敘述統計摘要（如檔案數、字數、詞數、年代與來源分布），有助於讀者理解語料範圍與結構，並判斷樣本的代表性與分析基礎。
  - 作者回覆：感謝審查意見，已於研究設計與實施一節加入研究對象說明，包括檔案數、字數、年代與來源分布等基礎資訊。
2. 更清楚說明資料前處理過程，例如檔案選取／排除標準。
  - 作者回覆：各類型分析前段已補強對於分析內容之原則說明，對於檔案文本，並未進行選取或排除，以完整呈現檔案系列內容。
3. 建議將後續地圖視覺化所使用的「分期」標準於方法章節中明確說明，強化論述連貫性與方法論透明度。
  - 作者回覆：文本地理資訊相關內容中已說明時間分期之切分考慮，並於各時期增列背景描述，同時調整寫作方式，以強化論述。

## 三、研究結果與討論

結果分析詳實，圖表呈現具參考價值，然而在結果解讀與歷史意涵連結方面，尚可進一步加強。具體建議如下：

1. 多數分析仍停留在人物職稱與事件類型的描述，較難顯現資料背後的政治運作邏輯與時代結構。建議作者嘗試從共現網絡與地理分布中提出具歷史詮釋力的問題，如黨內權力結構、政權遷移與地緣政治等，促使數

據分析與歷史論述產生對話。

- 作者回覆：已增加對於分析結果理解角度之說明，並適度加入數位人文研究意義以及回歸史學家研究解釋之重要性。
2. 社會網絡圖若能結合中心性分析或模組劃分，辨識潛在中介角色或跨機構人物，將有助於提出具體的歷史解釋假說。
- 作者回覆：本研究以整體性分析為主軸，相關之中心性分析與模組劃分雖具高度潛力，惟限於篇幅與研究限制，本研究尚無法增列此部分探究，故增列說明於後續研究建議中，未來研究可基於本研究成果進一步深入探討，進行更細緻的解讀與分析。

#### 四、結論與建議

結論篇幅略長，部分內容與前文重複，且未依學術貢獻排序或對研究成果進行層次性的統整。建議如下：

1. 精簡結論段落，避免重述研究過程與圖表細節，改以明確整理本研究於方法創新、資料詮釋與實務應用等層面的核心成果。
  - 作者回覆：感謝委員意見，已修改文字，減少對於研究過程與圖表的重述。
2. 於建議與未來研究方向中，可進一步強調文本探勘在歷史檔案應用的深化潛力，說明未來如何透過更多元或進階的語言處理方法（如主題模型、語義網絡）探討歷史語境、權力結構或敘事策略等問題，以鞏固其於史學研究中的方法地位。
  - 作者回覆：感謝委員意見，已增加對於後續研究建議之文字內容。

#### 總結

本文結合數位工具應用於近代政治檔案之研究設計具潛力，資料分析面豐富、圖像呈現清楚，若能進一步強化文獻脈絡、補足方法說明，並深化歷史詮釋與理論貢獻的層次，將能提升本研究在數位人文與歷史資訊學領域

# 圖書資訊學研究

JOURNAL OF LIBRARY AND INFORMATION SCIENCE RESEARCH

的學術價值與影響力。

# 圖書資訊學研究

JOURNAL OF LIBRARY AND INFORMATION SCIENCE RESEARCH

審稿者：主編綜評

評閱意見：

請按照評閱者意見修改後重審。

- 作者回覆：三位審查意見逐項回覆說明如附檔

## 審查階段

### 複審

審稿者：匿名審稿者 A

#### 評閱意見：

作者對於各種數位工具的使用相當熟稔，確實令人激賞。然而，吾人認為數位人文研究最終似乎仍應該回歸人文層面，如此才有意義，不知作者是否贊同？誠如作者於文獻分析內所列舉的諸多案例，學者們多半是在遇到具體的「問題」後，再選擇、利用各種數位工具或方法來釐清或幫助解決「問題」；但反觀作者僅主張將特種檔案內容透過數位工具運算後的結果呈現出來而已，卻不知道本研究的具體「問題」到底何在？要解決的「問題」是什麼？復宣稱可以提供日後歷史學者研究之參考，或主張「最終之評價與解釋仍屬於歷史研究之範疇」，這樣的研究作為實在令人難表認同。本文若為原創性研究論文，誠摯建議作者必須先釐清問題意識，再針對各個問題利用各種不同的數位工具，然後確實說明經過運算後所獲致的結果對於解決問題的貢獻或幫助，相信如此才是專業且負責的研究。當然，若是本文的研究目的只是設定在幫助讀者解析特種檔案的內容，或許仍有些許參考價值，於茲亦尊重作者之修正結果。

- 作者回覆：感謝評委意見，本文進行特種檔案內容分析之初衷，在於解析特種檔案的內容，提供對於此批檔案整體的概觀，審查者提出對於研究「問題」意識的看法是值得思考的觀點，因此，本文再度修改緒論文字，增列關於使用數位工具進行文本整體內容分析希望解決之問題，至於與內容專家合作以問題導向進行資料分析的建議良善，本文列為未來研究建議。

## 審稿者：匿名審稿者 B

### 評閱意見：

本研究以緒論中指出研究目的為探討「特種檔案」文本內容的人物、組織、事件等詞彙之分布、共現關係，並經 GIS 分析其文本內容地理分布與當代地緣政治之關連。但在結論與建議部分，較多著墨於數位工具應用與數位人文分析的實作與其展現之價值。緒論中所提出的問題意識明確聚焦於特種檔案本身，而本文結論與建議展現之主要價值則是於研究工具的應用。建議作者可考慮釐清兩者的關聯，進一步強化所述研究問題與主要分析成果之間的連結。

- 作者回覆：感謝審查意見，已針對結論、建議之內容與呈現順序進行修正，以強化研究問題與分析結果的連結。

在研究方法部分，文中提及使用 Excel 格式的特種檔案目錄作為文本來源，並透過欄位篩選、格式轉換及 CKIP Tagger 進行斷詞與實體辨識。關於最終保留下來的文本欄位內容，尚未詳加說明，若能補充說明篩選依據與欄位特性，將有助於理解資料處理的基礎。

- 作者回覆：感謝提醒，已於研究對象與研究架構說明中，增列文字說明保留下來成為分析標的之文本欄位內容。

對於 CKIP 預處理的資料，本文提及會去除無意義詞語及非中文字符（頁 10）。請說明去除詞語的進行方式，以及用於判斷詞語是否屬於無意義的標準。

- 作者回覆：感謝審查意見，已增列文字說明去除詞語的方式並對於文字用詞進行修正，避免誤解。

關於權威控制的部分，研究中提及在 CKIP 預處理後，會透過人工方式回溯原文，進一步判斷權威詞表的處理方式。第 10 頁以「陳素」與「陳素電」為例，說明了意義相同但字詞不同時進行合併的情形。不過，目前尚未說明在何種條件下，研究者雖判斷字詞具相同意義，卻選擇不合併計

算。若為分析工具功能限制，為何可合併「陳素」與「陳素電」，而無法處理上述同義詞？

- 作者回覆：感謝審查意見，已調整文字說明方式，並調整不同詞彙合併判斷更為明確的案例，以減少閱讀誤解。

圖 1 中並沒有指出權威控制進行的階段。圖中在 CKIP 進行文本預處理後，就進入數位人文研究平台、CORPRO、DocuSky 進行內容分析的階段。但頁 14 中，作者指出「惟因當前分析工具功能限制，同義詞彙無法進行合併計算處理」，這似乎代表在以數位人文研究平台進行社會網絡分析時，似乎仍有未合併同義詞彙？

- 作者回覆：已於研究程序與資料分析部分增列權威控制的步驟說明。

頁 14 的「因此選擇保留原始詞頻，以避免工具僅統計特定詞彙導致節點共現數量與實際情況不符」一句，文意不明，需再做修正。

- 作者回覆：感謝審查意見，已修改語句，以期更為直觀理解。

由圖 2 的文字雲與表 3 的人物高頻詞對照觀察，部分在文字雲中顯著呈現的人名，未見列入表 3。以「金九」為例，在文字雲中，金九一詞似較表 3 列於末位的王天雄更為顯眼，但此人未見於表 3 中。建議作者再做確認。

- 作者回覆：感謝委員提醒，經重新檢視數據，已根據實際結果對於字體大小進行調節。

其它細節修正意見，列舉如下：

- 多個段落中，作者文中的單句過長，需全面再做修正。部分需修正段落，如下列：第二頁第一段、第三頁第二段、第三頁第四段、第三頁第五段。
- 第八與第九頁，小節標題編號有重覆。
- 作者回覆：感謝委員提醒，對於提示的段落文字皆已進行修正。

審稿者：林頌堅

評閱意見：

作者已回應前次意見，修訂後之稿件在文獻選擇、研究方法透明度與分析詮釋層次上皆有明顯提升，研究主題與方法的對應性亦更為清楚，整體論述更加聚焦。

但建議作者能夠進一步調整結論段落的組織結構，將各項論點依據其學術貢獻的重要性進行排序，並檢視內容中是否有與研究主旨關聯性較低的項目，例如「具備程式能力有助於分析效率」一點，是否適合保留於結論中，宜再斟酌。此舉有助於提升整體論文收束的力道與一致性。

- 作者回覆：已針對結論與建議之內容與呈現順序加以修正，俾於更貼合研究主旨，同時刪除「具備程式能力有助於分析效率」此項文字，感謝審查意見。

# 圖書資訊學研究

JOURNAL OF LIBRARY AND INFORMATION SCIENCE RESEARCH

審稿者：主編綜評

評閱意見：

評閱者建議修改，本刊再接受您的稿件。

- 作者回覆：所有評閱意見逐項回覆如附檔