

研究論文

# 平台「熱趨同」：基於新浪微博的內容熱度與相似度網絡中心度關聯結構分析

徐翔、孫天怡

## 摘要

社交媒體的熱門內容看似主題多元，實際上是否在隨著其熱度增長而有規律地越來越趨同？本研究基於新浪微博7,869,861條帖子樣本聚類後的類相似度矩陣，對微博資訊內容在相似度矩陣中的中心性與其熱度間的關係與結構進行實證分析，以考察新浪微博內容隨著熱度增長而表現出的趨同態勢和演變結構。研究發現：平台中資訊內容類在類相似度矩陣中的中心度與其類熱度成正比；進一步分析顯示，微博平台輿論存在頭部資訊隨著熱度升高而成正比收窄的「熱凝聚」現象；平台的 $N \times N$ 內容相似度矩陣具有「核心-邊緣」結構；平台任意兩個內容類之間的相似度，與這兩類的熱度之和成正比、與這兩類的熱度之差絕對值成反比，並可經由線性迴歸方程進行預測。本文嘗試打

徐翔，同濟大學藝術與傳媒學院副院長、傳播系教授，大數據與計算傳播研究中心主任。研究興趣：社交媒體、計算傳播。電郵：[xuxiang210089@163.com](mailto:xuxiang210089@163.com)  
孫天怡，同濟大學藝術與傳媒學院傳播系碩士研究生。研究興趣：社交媒體、算法文化。電郵：[2131683@tongji.edu.cn](mailto:2131683@tongji.edu.cn)

論文投稿日期：2022年7月19日。論文接受日期：2022年11月11日。

《傳播與社會學刊》，(總)第65期(2023)

破社交媒體平台內容碎片化、去中心化的觀點，強調了分化背後的潛在統一性，揭示了平台內部封閉性和整體性的趨同結構，且這種趨同性隨著熱度增長而亦發生梯度增長。

關鍵詞：社交媒體、平台內容、中心度、熱度

---

Research Article

## **The “Hot Convergence” of the Platform: An Analysis of the Correlation Structures of the Centrality and Content Popularity and of the Similarity Network Based on Sina Weibo**

Xiang XU, Tianyi SUN

---

### **Abstract**

“Hot” content in social media seems to be based on multiple themes. However, has such content become more similar because of its increasing popularity? This study aimed to refute the view of social media content as fragmented and decentralized. A class similarity matrix was obtained by clustering 7,869,861 posts on Sina Weibo. An empirical analysis of the relationships between matrix centrality and the popularity of Weibo content was conducted to investigate the convergence and evolving structure of Sina Weibo content based on its increasing popularity. The results showed that the centrality of the class of content in the similarity matrix was proportional to the prevalence of the content. Further analysis revealed the phenomenon of “hot convergence” in the blog platform. The results of the trial showed that the main topics converged in proportion to the increase in their popularity. The results

---

Xiang XU (Professor & Vice President). Department of Communication, College of Arts and Media, Tongji University; Director of the Big Data and Computing Communication Research Center, Tongji University. Research interests: social media, computational communication.

Tianyi Sun (Master Student). Department of Communication, College of Arts and Media, Tongji University. Research interests: social media, algorithm culture.

*Communication and Society*, 65 (2023)

also showed that the  $N \times N$  similarity matrix of platform content had a “core-periphery” structure. Moreover, the similarity between any two content categories on the platform was proportional to the sum of the “hotness” of the two categories but inversely proportional to their absolute differential values, which was predicted by linear regression. The findings of this study showed that potential unity underlies divergent content, which indicated a closed and aggregated convergence structure that increased incrementally in proportion to increasing trends within the platform.

**Keywords:** social media, platform content, centrality, hotness

**Citation of this article:** Xu, X., & Sun, T. (2023). The “hot convergence” of the platform: An analysis of the correlation structures of the centrality and content popularity and of the similarity network based on Sina Weibo. *Communication and Society*, 65, 125–154.

## 致謝

本文是上海市「科技創新行動計劃」軟科學研究項目(23692110600)系列成果之一。

## 引言

隨著網絡的普及，各大平台已融入人們的日常生活，多樣化的平台推送了海量的資訊，引發了內容「碎片化」、「去中心」的討論。與此同時，社交媒體又似乎正以平台為單位形成內部封閉的資訊社區。本文的核心觀點是：社交媒體平台的內容若劃分為 $N$ 類，任意某一類內容在 $N \times N$ 的內容相似度網絡裏的中心度，與該類內容在平台的傳播熱度成正比。這種關聯結構會進一步引發平台中的一系列資訊傳播後果，使得資訊流動陷入多樣性趨於塌縮的「平台資訊黑洞」。現有研究在總結高可見度內容的主題時，對其內部聯繫和本質屬性的關注並不充分，產生了熱門內容彼此孤立的錯覺，以至於忽略了平台內容中心度與熱度之間的關係。

其一，平台熱門內容看似多種多樣、「百花爭春」的「碎片性」背後，是否隱藏著高熱度內容的「中心性」(centrality)? 社交媒體平台的流行內容傾向於網絡巴爾幹化(cyber-balkanization)亦或同質化一直存有爭議。一些研究認為平台展現了媒介話語權的去中心化和平民化特徵。借助平台，人們能夠自主獲取和分享資訊，這將使網絡空間分裂為裂痕越來越大的各式興趣社區(Van Alstyne & Brynjolfsson, 1996)。然而事實上在這些多樣化的內容生產中，其高熱度的內容可能具有窄化性。Goldszmidt與Yemini (1995)提出了中心化與去中心化範式能夠且應當共存的觀點，社交網絡平台呈現出一種在狹窄的資訊源中消費並選擇性地暴露於熱門內容的傾向(Nikolov et al., 2019)，Lee、Hoang及Lim (2017)認為用戶通常會對某些主題顯示出某些社交媒體平台的偏好，而非跨平台共用主題。平台有何偏好、熱門內容是否以及如何窄化，這些問題需要深挖平台推送的頭條內容隱藏的「中心性」來回應。

其二，平台熱門內容類型之間，看似相互分化、獨立，實則是否擁有某種反直覺的強相似性和強統一性? 現有研究已指出了新浪微博的熱門內容所包括的類型，例如「休閒心情」、「時尚娛樂」、「時事新聞」等(王國華等，2014；王國華、陳靜、鍾聲揚，2015；李彪，2013；張文汐，2014)，但僅停留於主題概括，就會忽視類別之間潛在的相似性和統一性，轉而突出了彼此的差異性和分化性。這些研究雖顯示出熱

《傳播與社會學刊》，(總)第65期(2023)

門內容只圈定在少數幾種範圍內，卻未能徹底展示出不同類型相互融通、彼此影響的可能性。有學者也具體討論了微博熱門的特點，如「娛樂化」、「情緒化」或「負面性」(曹林，2011；Zhao, 2021)，為本研究所關注的熱門內容同質化趨勢提供了重要的參考和借鑒，但缺乏對於這些屬性內部的一個分化性的關注，以負面性為例，憤怒、悲傷等不同的分化屬性如何被兼容在負面類熱門內容裏？另外這些研究也缺乏對於熱門和冷門內容之間在趨同程度上的比較和關注。可以看出，相較於邊緣主題，看似獨立的熱門實則是更彼此相似的。本研究認為熱門內容類之間「形不同而神一致」，本質上存在一種規律性的關聯結構。如若實證結果顯示高熱度、高回饋的內容彼此間確實存在緊密的相似性和統一性，且與其內容的中心性有關，那麼熱門資訊很可能正向平台中心收攏，不斷被吸收、同化，從本質上失去了多樣性，這種機制值得注意。

其三，在社交媒體多元化的表達空間中，是否任意類型的內容都能成為「熱門」？學界在這方面存在爭議，一方認為，結合個性化機制，平台傳播可能隱藏著「資訊偏食風險」(曾子倩，2019)，另一些學者則指出熱門有規律性，只有達成共識的內容才能獲得更顯著人氣(Figueiredo et al., 2014)。偶發的「爆冷」、「黑馬」讓人產生一種錯覺，即每個聲音都有平等的成名機會。韋路和胡雨濛(2014)通過新浪熱門微博實證研究發現，內容是影響微博熱度的決定性因素，在日常狀態下，熱門微博的議題呈現零散、瑣碎狀態，但沒有深挖分化類型背後的統一性。如果平台資訊的熱度隨著其中心度增加而增加，就意味著熱門內容在本質上具有普遍性，越普遍化而非特異化的內容，才越可能具有流行性。本研究認為，熱門內容之間存在強關聯。不同的內容之所以共同成為熱門，因為都在向中心趨同。本文通過檢驗平台頭部內容的相似程度，驗證熱門內容是否與中心性有某種密切關聯。

其四，對於在某個社交媒體平台中未出現過的新內容，是否能夠通過比對資訊特徵的中心性來預估其熱度？為了預測內容流行度，不同學者已提出各種模型，但面對未曾見過的全新內容，機器學習的效果往往欠佳。翁小蘭和王志堅(2018)總結了現有的解決方法，包括：

隨機推薦法、平均值法、人口統計學方法和相似性度量法等。一些研究雖處理了如何基於早期流行度預測最終熱度，但無法更早一步對於新內容進行預測，難以解決諸如「冷啟動」等問題或幫助內容發布者提早完善作品。也有一些學者找到了可用的文本屬性，如Bandari、Asur及Huberman (2021) 構建了一個基於文章屬性的多維特徵空間，Hong、Dan及Davison (2011) 基於消息內容、時間和用戶等文本特徵構建主題模型。但他們沒有關注這些文本屬性在中心性網絡中的重要程度，即新內容的這些資訊特徵與平台熱門內容圈層之間的內在聯繫。如果能夠更加充分地把握住內容中心性與熱度的關聯結構，就能夠避免向平台投放全新內容時「水土不服」。

綜合來看，平台中熱門話題或許對應著平台的某「中心點」。內容的中心性與熱度是否有關聯，它們之間是怎樣互相影響的，會對平台傳播產生怎樣的後果？為了回答這些困惑，本研究針對社交媒體平台，以新浪微博的樣本進行實證分析。如果熱門內容實際上被圈定在一個狹小的範圍內，平台就會成為一個整體趨同的熱門牢籠，將異質性的內容排斥在熱門之外，這種結構性的後果值得關注。

## 文獻綜述與假設提出

本研究對平台的帖子通過聚類方法將其劃歸為若干具有共性的類別，並使用「內容類」來指代每一組由此得到的內容類別。本文通過四個假設，嘗試分析高熱度內容類的共同特徵和內在聯繫結構。

### 議程集中：熱度與相似度的正相關 (H1、H2)

在平台資訊流的循環之中，所有議題正隨著輿情發展而不斷向平台中心熱點聚攏。高熱度報導出現後，媒體本能地貼合受眾偏好，尋找相似新聞(易豔剛，2015)，而平台用戶也會用自己的話語體系來處理、加工和傳播資訊。部分媒體大規模推送同類資訊，來迎合公眾的選擇性理解、接觸和記憶，從而產生「熱點搭便車」效應。實證檢驗發現，

《傳播與社會學刊》，(總)第65期(2023)

存在媒體框架設置等問題致使公共議程走向單一化(Pinto et al., 2019)。新媒體經常複製著名傳統新聞媒體的已有內容；記者經常相互觀察並模仿報導；此外，類似的新聞規範和傳統被統一應用於各類事件(Tan & Weaver, 2013)。據調查，日本主要傳統媒體的新聞內容都有相似之處，因此可以預期在線媒體平台的議程也存在同質性(Takeshita, 2006)。

一些相關研究提到了平台內容的趨同傾向。Airoldi、Beraldo及Gandini(2016)通過社會網絡分析(social network analysis)檢測了同個平台裏50個緊密連接的視頻集群，發現它們在不同的相似軸上具有強大的內部同質性。這種向某一中心點靠近的同質化趨勢使得平台內部形成一個趨同結構，越靠近中心，內容越集中、越多連接點，因而越可能被關注。媒體風暴(media storms)的存在加劇了模仿的過程。生產者只發布可能受歡迎的內容，發布內容的影響因素主要為趨勢、發布者興趣、鄰居興趣，且發布的內容具有與其他內容互相影響的外部性和收斂性(Masson, Hayel, & Altman, 2017)。雖然媒體風暴現象僅僅概括了短期內的輿論反應，但也有力佐證了議程正走向收窄的同質化。

從長期看，依託於技術的社交媒體平台依舊展示出一種向熱門類目靠近的現象。用戶在平台的規則體系下進行內容生產，以召回率為目標的推薦算法帶來了內容選題、類型和風格的系統化傾向，直接影響到內容生產者的旨趣；同時，內容召回和用戶留存的前提在於內容的高效產出，這也促使內容生產活動變得越加職業化和標準化(翟秀鳳, 2019)。據此，提出假設一：

**H1：越是相似於最高熱度內容的類別，該類別的熱度也越高。**

如果H1成立，即與平台最熱門的內容類別越靠近，就越會獲得熱度，那麼相應的，熱度越低的內容之間，是否彼此越分化？是否在高熱度的內容圈層之中，會出現內容的集中現象？Zhu(1992)曾提出議程設置的零和理論，由於公眾議程承載力的有限性，媒體議題增加意味著不同議題在有限的空間相互競爭，多一種議題，另外的議題的影響力就會減弱，其結果是受眾接受的議題總量並不增加，而每一種議題的影響力卻相應減弱。Guan、Liang及Zhu(2022)也認為社交媒體對於同一事件的集中議程可以有效降低資訊冗餘，從而提升傳播效果。



零和理論和集中議程都認為，公共議程的數量受到時間、注意力和資源等約束力的限制，資訊和輿論會被盡可能整合成一個共同的声音發出。於是在短時間內，為了實現最優化的傳播效率，高熱度的內容逐漸趨於集中，用相似的模式進行產出，以避免受眾注意力分散，可見性降低。在更長的時間跨度中，Jennings、Bevan及John (2011)認為受到更多關注的高熱議程會對長期議程的多樣性產生負面影響。徐翔和王雨晨(2021)通過今日頭條案例的實證研究證實，社會資訊趨同的現象及其隨熱度演化的曲線函數變化在五天到120天的24種週期長度內的差異都很小，即在長週期中內容類趨同的現象依舊存在。由此看來，無論週期長短，高熱度內容都呈現出明顯的類型有限化、收窄化現象(陳昌鳳，2014)。平台的承載量和用戶的注意力始終都是有邊界的，平台熱門議程走向集中有其內在的趨勢性。據此提出假設二：

**H2：處於越高熱度的內容類別，彼此之間的相似度越高。即，內容熱度越高，彼此越趨同。**

通過相似度矩陣的維納多樣性指數(Weiner Index)可測量類和類之間的平均相似程度，但這類指標計算過程相對簡單，本研究採用在社會網絡分析中較為穩定的平均群聚系數(average clustering coefficient)(Eggemann & Noble, 2011)來考察熱度層級內部帖子類之間的趨同度。該系數可以反映某帖子的鄰接(adjacent)帖子之間的相互連接，即各內容類傾向於接近、相似在一起的緊密程度。平均群聚系數可以全面體現出內容類之間的相似程度，該值越高代表帖子類之間越相似，反之則代表帖子類之間彼此離散、分化的程度高。

內容同質：相似內容的相似熱度(H3)

一方面，同一平台的用戶消費畫像具備一定的共性，例如「性別」、「階層」、「興趣」等。互聯網的「匿名性、群體凝聚和群體沉浸」(anonymity, group cohesion, and group immersion)(Tajfel, 1986)等特點共同促進了群體標準的一致性。用戶一直在主動選擇「主流資訊」。Hongan (2011)認為在社交媒體平台中，人們傾向於更多地與所有人都

《傳播與社會學刊》，(總)第65期(2023)

感興趣的內容交互。在微博中，不同影響力「能量」級的用戶具有朝向「意見典範」趨同的演化機制，存在一種動態而確定性的社會內容同質化現象(徐翔、劉佳琪、靳菁，2021)。

另一方面，平台作為一個內容「黑洞」，隱形地塑造了受眾的需求。楊洸和余佳玲(2020)的研究顯示，用戶越沉浸於平台中，得到推薦的資訊同質化越高、觀點多樣性越低。在數量和產品組合效應的作用下，個性化推薦反而創造了與他人的內容共性。另外，社交網站通過「點讚」等機制對成員進行「顯式行為矯正與規訓」(explicit group identification and deindividuation) (Mikal et al., 2016)，讓想成為平台一員的用戶感知到相關的群體規範，並試圖按照這些規範對內容進行整合。長此以往，高可見度的類別之間形成了彼此相似度更高的富人俱樂部，所有平台用戶擁有一個共同關注的內容中心，而這一中心實際上束縛了用戶的視野。據此提出假設三：

**H3：相似的內容類別之間具有相似的熱度。換言之，內容類型的熱度越高，與之相似的內容類型熱度也越高。**

追隨熱門：熱度與中心度的關係(H4)

同一社交網絡平台具有某種共通的溝通標準，它導致用戶的內容生產趨向於和該標準保持一致。例如在口號為「隨時隨地發現新鮮事」的微博上，「個體工作或生活實錄、感言」(夏雨禾，2010)在微博達人所發起的對話中佔71.1%。這在一定程度上證明，平台意見領袖在生產過程中並不獨立，而隱藏著一個共同的內容中心，人們正在對平台熱門內容進行無意識的模仿行為。隨著一個線上平台的用戶逐漸了解該網站的文化和參與規則，就會開始預測社區對某些原創內容類型的歡迎程度，以修改自己的內容發布，符合感知到的平台文化規範(Mikal et al., 2014)。

這種追隨使得平台中心內容被壓縮和凝聚在一個很小的範圍。社交網絡平台的意見領袖不僅彼此之間有著更強的連接度和更強的相似度，而且也會和普通用戶、甚至邊緣群體緊密相連。於是，平台受眾

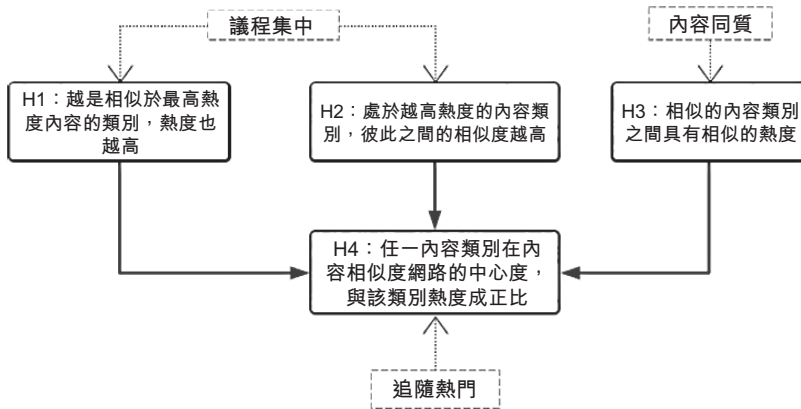
大多數時間只接觸熱門資訊；高可見的內容幾乎完全由核心用戶群體生產；這些生產者在追隨熱門內容，彼此比較和模仿。這個循環模式會加速平台內容的同質化。

從這個角度說，社交媒體平台在重要的內容約束方面是霸權的，而公眾的關注模式是重疊的。資訊的內容特徵與其傳播熱度具有密切聯繫，平台中的熱門內容呈現出明顯的窄化傾向，而其中重疊的流行資訊會始終成為關注的最中心。據此現象，若H1、H2、H3成立，則越和熱門相似，越能擁有熱度；越擁有熱度，則彼此越相似。由此推測假設四：

**H4：媒體平台中，任一內容類別在內容相似度網絡的中心度，與該類別熱度成正比。**

四個假設和論據的關係如下圖（圖一）：

圖一 假設之間的關係



## 研究設計與方法

本文的核心問題是探討媒介平台中，內容中心化程度與內容熱度之間是否存在關聯規律；若存在，究竟是何種結構？結合理論和文獻

《傳播與社會學刊》，(總)第65期(2023)

分析，本文從四個關鍵假設入手，基於新浪微博的數據樣本進行實證檢驗。

由於粉絲基數等差異，同一內容經由不同的發布者很可能產生截然不同的熱度，因而會對實證結果造成干擾。因此，本研究需要通過聚類後的相似度矩陣計算，得出相似內容的熱度均值，以稀釋個別發布者帶來的過度影響，反映該類內容的穩定水準。

### 樣本抓取與預處理

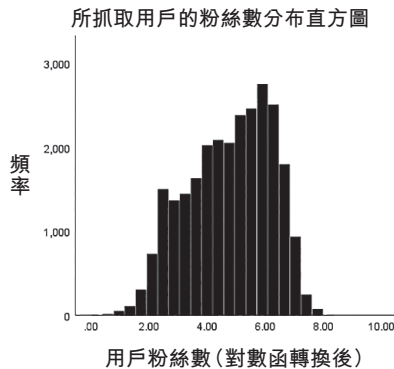
本研究選擇新浪微博為樣本，根據新浪微博2021用戶發展報告，截至四季度末，微博月活躍用戶達到5.73億，日活用戶為2.49億。數據抓取採用Selenium編寫網絡爬蟲並配合「八爪魚」軟體，使用滾雪球的抽樣抓取方法進行。滾雪球的數據抓取方法在社會網絡研究領域中具有可資借鑒的使用，例如Ashokkumar與Pennebaker(2021)以Reddit作者的個人網絡進行滾雪球。Patton、Frey及Gaskell(2019)的一份權威研究以一個幫派成員作為種子用戶，對其Twitter好友列表進行滾雪球，該過程收集了具有廣泛性的可用數據，對結論形成良好支持。

首先，運用「八爪魚」軟體以及Python和Selenium編寫動態網頁抓取程式進行網絡數據抓取。從新浪微博首頁47個內容版塊(社會、國際、科技、科普、數碼、財經、股市、明星、綜藝、電視劇、電影、音樂、汽車、體育、運動健身、健康、瘦身、養生、軍事、歷史、美女模特、美圖、情感、搞笑、闢謠、正能量、政務、遊戲、旅遊、育兒、校園、美食、房產、家居、星座、讀書、三農、設計、藝術、時尚、美妝、動漫、宗教、萌寵、法律、視頻、同城)中，各個版塊每天早晚各抓取一次，持續一個月，獲得65,650條帖子。本次抓取歷時一個月，避免了短時間內大量抽取的樣本偏差；並藉助微博自身的分區設置，使得樣本廣泛而大致均衡地分布在47個不同版塊中，具有較大的覆蓋面和良好程度的代表性。

從這些帖子的發布者中，隨機抽取500個用戶，作為「種子」用戶；對每個種子用戶從其最多前五頁被關注者中隨機抽出一名，也即500個

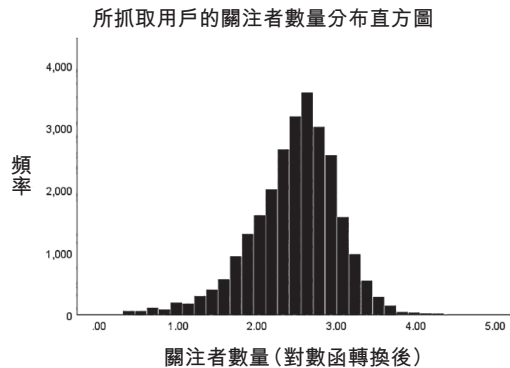
種子用戶得到下一輪的500個樣本用戶；將此過程循環100輪得到500個 $\times$ 100輪=50,000個用戶。樣本中，用戶的粉絲數最小值為1，最大值達到99,261,000，平均值1,372,919.50，中位數為84,000，標準差為4,479,776.81；用戶的關注者數（關注他人的人數）最小值為0，最大值為20,000，平均值為576.51，中位數為337，標準差為979.83；用戶的發微博數最小為0，最大為874,762，平均值為7,814.83，中位數為1,718.5，標準差為21,144.53。總體上看，這些用戶從無人問津的「草根」到「中V」、「大V」以及各種不同活躍度的用戶，都有廣泛分布，減少了樣本窄化的風險。具體如圖二至四所示。由於分布值域太廣而且各值差距過大，經對數函數轉換後窄化分布範圍予以呈現。

圖二 粉絲數分布直方圖



註：橫軸粉絲數量= $\log_2$ (粉絲數原始值+1)

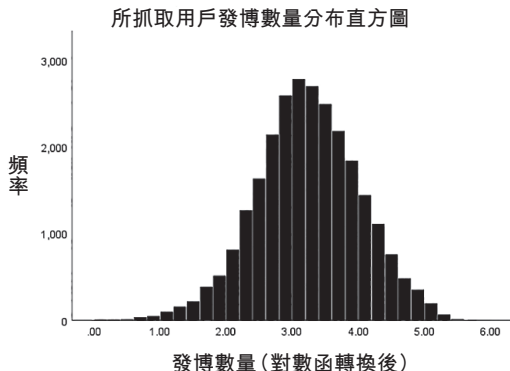
圖三 關注者數量分布直方圖



註：橫軸關注者數量= $\log_2$ (關注數原始值+1)

《傳播與社會學刊》，（總）第 65 期（2023）

圖四 發博數量分布直方圖



註：橫軸發博數量 =  $\log_2(\text{發博數原始值} + 1)$

對上述用戶抓取其歷史發帖，抓取工作的實施時間為 2021 年 10 月到 2022 年 1 月，該階段共成功抓取得到帖子三千多萬條。考慮時間跨度問題，只選取 2017 年 1 月 1 日到 2019 年 12 月 31 日這三年的帖子。其一是發現回溯歷史過早後，每天的帖子逐步零星化，數量不夠充分和穩定；其二是 2020 年和 2021 年爆發的新冠疫情讓輿情具有特殊的主題性，單一事件在輿情中影響和佔比過大，可能對本研究的結果引入過大偏差，因而進行剔除。最後得到的樣本帖為 7,869,861 條。對帖子的評論數、點讚數進行統計分析，具體如表一所示。這些樣本帖子廣泛涉及各個熱度層，具有良好的覆蓋度和代表性。

表一 帖子評論數和點讚數的描述統計數據

	評論數 (對數轉換後)	點讚數 (對數轉換後)
最大值	19.97	23.07
最小值	0.00	0.00
中位數	2.00	4.10
方差	10.73	14.10
標準差	3.28	3.75
平均值	3.07	4.55
四分位數		
25	0.00	1.00
50	2.00	4.10
75	5.10	7.20

## 帖子的向量化表示及其聚類

首先，採用 Word2vec 詞向量及其平均池化 (word vector and its mean-pooling) 的方式對帖子內容進行向量化轉換。將詞向量進行等權求平均向量的方法與迴圈和卷積網絡 (word-embedding-based RNN/CNN models) 進行比較發現，在大多數情況下前者表現出更高性能 (Shen et al., 2018)。Word2vec 是一種淺層神經網絡語言模型 (Neural Network Language Model, NNLM) (Mikolov et al., 2013a; Mikolov et al., 2013b)。通過無監督訓練將文本數據轉化為低維實數向量，通過語言模型訓練對語義和語法資訊進行訓練並投影到多個向量空間，從而完成文本的結構化處理。向量空間模型 (Vector Space Model) 是常用的文本表示模型，對於特徵覆蓋性較全，但它並不能解決辭彙相似度和「辭彙鴻溝」(vocabulary gap) 的問題，而 Word2vec 基於神經網絡的詞嵌入則可以很好地解決這個問題。

對於 Word2vec 模型的搭建和訓練，採用開源模組 Gensim (Řehůřek & Sojka, 2011) 實現。訓練 Word2vec 所使用的語料採用 26G 的中文語料庫，來源包括媒體新聞庫、網絡論壇帖子抓取、經典名著和文學文本等，該詞向量訓練結果包括 5,830,979 個辭彙的嵌入式表示，訓練效果良好。

Word2vec 詞向量及其平均池化的方式計算方式如下。Word2vec 的詞向量對帖子進行等權平均計算形成句向量，對 Word2vec 的參數設定 300 維空間。將樣本中的帖子，按照分詞後的結果，把每一個單詞轉換成一個 300 維的詞向量，然後計算每個詞向量的平均值，得到該帖子、短文本的語句向量。具體來說，逐一將樣本中的帖子 T1 中的每一個詞 (A / B / C / D) 各表示為形如  $V=[0.01, 1.002, -0.02, \dots, 2.26, 0.05, 3.97]$  的 300 維的向量。接著計算 T 所有詞向量的等權平均值 (也即平均池化)：

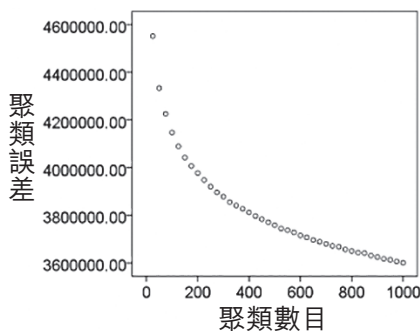
$$VT_i = (V_A + V_B + V_C + V_D) \div n \quad (1)$$

在此基礎上，對帖子向量採取  $k$  均值方法聚類，聚類工具採用 scikit-learn 中的  $k$  均值函數。 $k$  均值聚類算法 ( $k$ -means clustering

《傳播與社會學刊》，(總)第65期(2023)

algorithm) (Hartigan & Wong, 1979) 解決的問題是，在事先不知道如何分類的情況下，讓程序根據距離的遠近，把 $N$ 個對象最優的劃分為 $k$ 個類，是常見的一種無監督算法。聚類前對向量通過scikit-learn的normalize函數作L2的規範化預處理。計算聚為不同數目下的誤差(kmeans.inertia)，由於只是比較聚類誤差的變化趨勢，因此反覆運算次數統一為六次，其誤差如圖五顯示：聚類超過400類時，誤差減少緩慢，再增加聚類數目無助於有效減少聚類誤差，因此本研究最終選擇把所有帖子聚為400類。

圖五 聚類誤差與數量圖示



其後對全部帖子進行400類數量下的計算迭代，通過 $k$ 均值聚類算法迭代60次得到最終的聚類結果。400類主題的分布的百分比例中，中位數是0.0021，標準差為0.0021。其後提取每一類中TF-IDF值最大的10個關鍵詞，TF-IDF可以用來評估某一字詞對於全文的重要程度，其運行邏輯是該字詞出現的總頻次高且在其他文檔中很少出現，因此有良好的區分能力。本批次樣本聚類後，序號前十類內容類的TF-IDF值最大的十個關鍵詞如表二所示。



表二 聚類後序號前 10 位的內容類的關鍵詞 (每類中 TF-IDF 值最高的 10 個詞)

類標號	詞1	詞2	詞3	詞4	詞5	詞6	詞7	詞8	詞9	詞10
類0	上午	明年	就要	敬請期待	上線	開播	截止	預售	打卡	週五
類1	營業	摸魚	共勉	畫師	精選	操作	明日	嘻嘻	夏天	打卡
類2	青春	不想	表達	眼淚	善良	感情	總是	長大	溫暖	一輩子
類3	三位	簽名	隨機	私信	小可愛	詳情	這條	中獎	每人	平分
類4	旋律	qq	聲音	鋼琴曲	主題曲	這首	舞蹈	樂隊	演唱	吉他
類5	場次	坐車	場面	龍蝦	均勻	坐等	坐火車	坐標	坐姿	坐地鐵
類6	Hhhh	最右	毛病	呵呵	救命	爆笑	有毒	嘻嘻	神經病	恍恍惚惚
類7	表演	老外	槽寶	臥槽	倉鼠	尷尬	內涵	瞬間	操作	微笑
類8	樣子	學會	真正	遇見	珍惜	總是	過去	失望	決定	明白
類9	公園	music	hip	嘻哈	day	The park	prpr	use	Music vide	rea
類10	大獎	年度	編劇	預告片	觀影	青年	電影展	北美	單元	戛納

## 帖子熱度計算

在微博樣本中，評論數和點讚數原始數據值較大而且單個數值之間的差異大，因此分別將它們的每個指標原始數值加 1 後再經對數函數 (以 2 為底) 轉換，其後進行 min-max 歸一化處理為 [0,1] 區間內的值。分別處理後，以帖子的點讚數和評論數的歸一化結果等權相加作為帖子熱度值。經過統計分析，所有帖子熱度的平均值是 0.17，標準差是 0.16，中位數是 0.14。在單條帖子熱度基礎上進一步計算得到每類帖子熱度的平均值。

## 帖子之間的相似度和趨同度計算

對於每類帖子之間的相似程度，根據「類平均法」展開。將任意兩個帖子向量 ( $t_i$  和帖子  $t_j$ ) 之間的「餘弦相似度」(張振亞等，2005) 表示為  $R(t_i, t_j)$ 。其中帖子向量由前文所述的 Word2vec 得到，帖子類型則由前文所屬的聚類方法得到。分別擁有  $n_1$  條和  $n_2$  條帖子數量的任意兩類帖子之間的相似度，表示為兩類中每條帖子向量兩兩餘弦相似度的均值，見式 (2)：

《傳播與社會學刊》，(總)第65期(2023)

$$\frac{\sum_i^{n_1} \sum_j^{n_2} R(t_i, t_j)}{n_1 n_2} \quad (2)$$

接下來計算帖子的平均群聚系數，以考察帖子之間的趨同度。在式(2)的基礎上計算出某層內 $m$ 類兩兩之間的 $m \times m$ 相似度矩陣，並進一步把矩陣中的相似度值進行歸一化的預處理，然後對該矩陣採用開源模組Networkx中的average\_clustering函數計算其平均群聚系數。

### 內容類在相似度矩陣中的中心度

中心性是社會網絡分析的重點之一，本研究採用特徵向量中心性(eigenvector centrality)和接近中心度(closeness centrality)兩個指標來判定不同內容類在相似度矩陣中居於怎樣的中心地位。根據式(2)的計算公式，可以算出任意兩類的相似度，重複 $400 \times 400$ 次，可以得到一個 $400 \times 400$ 的相似度矩陣。將每個相似度的值都進行歸一化處理為[0,1]區間內的值，然後計算該矩陣裏每個類的中心度。

「特徵向量中心性」(Bonacich, 2007)依據其相鄰節點中心性對該節點的中心性進行函數計算，表現了該點在網絡中的重要性。用 $EC_i$ 來表示 $i$ 的特徵向量中心性， $X_i$ 代表 $i$ 的重要性，見式(3)：

$$EC_i = X_i = c \sum_{j=1}^n a_{ij} x_j \quad (3)$$

其中， $c$ 為一個比例常數， $a_{ij} = 1$ 當且僅當 $i$ 與 $j$ 相連，否則為0。

「接近中心度」(汪小帆、李翔、陳關榮，2012)可以用來描述網絡中某一點與其他各點的接近程度，當節點與網絡中其他節點的最短距離之和越小，該點的中心性程度就越高。用 $CC_i$ 來表示 $i$ 類的接近中心度，計算方式如式(4)：

$$CC_i = \frac{n-1}{\sum_{j \neq i} d(i, j)} \quad (4)$$

其中， $d(i, j)$ 表示類 $i$ 到類 $j$ 的距離。基於該類到網絡中其餘所有類的最短路徑之和進行歸一化處理，得到該類到其他所有類的平均最短距離。距離越小，中心性越大。

## 研究結果與假設檢驗

利用 SPSS26.0 統計軟體，對處理後的樣本數據進行皮爾遜相關性分析 (Pearson correlation analysis)，從而對各假設進行檢驗。

### H1 檢驗

對於聚類後得到的 400 類帖子，首先，將各類帖均熱度和它「與最熱的  $N$  類內容的相似度」之間進行相關性分析。此處採用皮爾遜相關分析方法，並進行雙尾檢驗 (two-tailed test)。

結果表明，微博中，某類帖子的平均熱度及其與最熱門的一類、十類和三十類內容的相似度的相關系數分別為 .51 ( $p < .01, N = 400$ )、.71 ( $p < .001, N = 400$ )、.71 ( $p < .001, N = 400$ )，顯著性符合要求，證明了 H1：越是相似於最高熱度的類別，其熱度也越高。

### H2 檢驗

將微博數據的類均熱度與它所處的「同熱度層的相似度網絡平均群聚系數」(average clustering coefficient of similarity network) 之間進行相關性分析，同樣採用皮爾遜相關分析方法，並進行雙尾檢驗。

結果表明，微博中，某類帖子的類均熱度與同熱度圈層十類帖子的相似度網絡平均群聚系數、同熱度圈層三十類帖子的相似度網絡平均群聚系數的相關系數分別為 .79 ( $p < .001, N = 400$ )、.86 ( $p < .001, N = 400$ )，顯著性符合要求，證明了 H2：處於越高熱度的圈層之中的內容類別，彼此之間的相似度越高，即高熱度內容具有圈層集中性，彼此趨同。

### H3 檢驗

採用皮爾遜相關分析方法，並進行雙尾檢驗，以對於某類帖子自身的類均熱度同「與之最相似的  $N$  類熱度」之間的關係進行相關性分析。

《傳播與社會學刊》，(總)第65期(2023)

結果表明，微博中，帖子自身的類均熱度與和該帖最相似的一類、十類和三十類帖子熱度的相關系數分別為.48 ( $p < .01, N = 400$ )、.58 ( $p < .01, N = 400$ )和.56 ( $p < .01, N = 400$ )，顯著性符合要求，證明了H3：熱度越高，相似的內容的熱度也越高，即相似的內容之間具有相似的熱度。

#### H4 檢驗

最後，將微博中某類帖子的特徵向量中心性、接近中心度與平均熱度三個變數進行相關性分析。特徵向量中心性反映了內容類的重要程度，而接近中心度越小在網絡中越處於核心地位，因此，若假設成立，則帖均熱度應當與前者成正比，與後者成反比。

此處首先採用皮爾遜相關分析方法，並進行雙尾檢驗。結果表明，微博中，內容的接近中心度、特徵向量中心性與帖子熱度的相關系數分別為-.68 ( $p < .001, N = 400$ )和.68 ( $p < .001, N = 400$ )，顯著性符合要求。

接下來，為了進一步明晰變量之間的關係，本研究引入兩個線性迴歸分析進行檢驗。共線性診斷顯示變量之間基本不存在共線性，殘差直方圖和正態P-P圖均滿足了線性迴歸模型對於殘差正態分布的要求。簡單線性迴歸結果提示，接近中心度與熱度之間存在線性關係  $F(1,398) = 345.76$  ( $p = .000 < .001$ )， $R^2 = .46$ ， $Beta = -.41$  ( $p < .001$ )；特徵向量中心性與熱度之間也存在線性關係  $F(1,398) = 344.71$  ( $p = .000 < .001$ )， $R^2 = .47$ ， $Beta = 13.90$  ( $p < .001$ )。其分析結果與皮爾遜相關分析吻合，內容的特徵向量中心顯著正向預測內容相似度，接近中心度顯著負向預測內容相似度。具體的迴歸方程如下：

帖子熱度 =  $0.862 + (-0.408 \times \text{接近中心度})$ ；

帖子熱度 =  $-0.524 + (13.9 \times \text{特徵向量中心性})$ 。

因此可以認為，類別內容中心度與類別熱度成正比，H4成立。

## 進一步的討論

若前文假設成立，可得到一系列推論。這些推論可以進一步明晰平台內容的熱度和中心度分布規律，檢驗「熱趨同」的特徵和結構。下述三條推論全部檢驗成立，既進一步表現了本文核心假設的合理性，也表現出這些假設對於預測某些現象的有效性。

### 平台輿論隨著熱度增加而表現出「熱凝聚」現象

H2 指向高熱度內容的集中性，內容熱度越高，所在圈層內容相似性越高。換而言之，高熱度的輿論內容具有範圍的有限性，而不是任意多樣的，由此，位於平台熱度最高的「頭部」領域的每天輿論內容，就會呈現出更集中的分布趨勢和更明顯的相似程度。據此得出推論一：平台的每天頭部輿論，若其平均熱度越高則頭部的輿論內容越是集中和封閉。對此也可簡稱為「熱凝聚」。熱凝聚現象是對平台輿論的高熱度內容向「中心」收攏趨勢的反映。

本文所使用的 7,869,861 條微博樣本，分布於 2017 年到 2019 年的 1,095 天。在每天的帖子中隨機抽取 2,334 條，然後從中選取熱度最高的 5% (116 條) 作為當日「頭部」內容，計算頭部平均熱度、頭部內容相似度和頭部內容基尼系數 (Gini index)。其中熱度的計算方法與前文相同；頭部內容相似度為前文經 Word2vec 平均池化後得到的帖子向量之間的彼此相似度，計算方法等同於式 (2) 中某類和自身類的相似度；頭部內容的基尼系數為 116 條帖子在聚類後的類型分布集中程度，採用基尼系數計算，該值越高表明頭部的帖子越是集中於少數類型。

「熱凝聚」現象的皮爾遜相關性檢驗結果顯示，當日頭部的平均熱度與其內容的相似度和基尼系數的相關性分別為 .25 ( $p < .01$ ,  $N = 1,095$ ) 和 .46 ( $p < .01$ ,  $N = 1,095$ )，因而，當日的頭部內容熱度越高，意味著能進入該頭部領域的類型越少，從而該類型基尼系數越高；同時內容相似度也越高。

《傳播與社會學刊》，(總)第65期(2023)

## 平台內部N類內容構成的 $N \times N$ 相似度網絡具有「核心—邊緣」結構

社會網絡分析中運用「核心—邊緣」(Borgatti & Everett, 2000)結構(core-periphery, 簡稱CP)區分網絡中不同節點之間聯繫的緊密程度並進行測量。核心—邊緣網絡結構中緊密相連的節點構成了網絡中的「核心部分」(core), 而連接較為稀疏的節點則是「邊緣部分」(periphery) (Csermely et al., 2013)。

綜合H1至H3, H4明確認為平台中某類內容的中心度與熱度之間呈現出正向的相關關係。內容越靠近平台中心, 越可以獲得關注熱度; 越在平台中擁有熱度, 也就越相似於熱門類型。換言之, 高熱度的內容之間具有高相似性; 低熱度內容彼此差異較大。這樣可以將平台內容清晰地劃定為「核心—邊緣」兩個不同的圈層, 兩者不僅在內部各類的連接距離上擁有較大差異, 也在整體熱度上區別明顯。得出推論二: 平台存在「核心—邊緣」結構; 高曝光、高熱度的內容彼此之間更加趨同和稠密, 而邊緣內容則較為分散, 內容越遠離平台中心, 越難成為熱門。

利用Ucinet軟體, 對於抓取到的400類微博內容類彼此相似度形成的400行 $\times$ 400列相似性矩陣數據, 進行核心—邊緣結構分析。類之間的相似度計算方式見式(2)。利用連續模型測算每個內容類的核心度(coreness), 並將coreness值 $\geq 0.50$ 的視作核心類, 將coreness值 $< 0.50$ 的視作邊緣類。在SPSS中通過獨立樣本T檢驗, 比較核心類和邊緣類的平均熱度差異。通過箱圖和Shapiro-Wilk檢驗判斷所有內容類的帖均熱度統計值符合正態分布( $p > .05, N = 400$ ), 且不存在顯著的異常值。結果如表三所示, 核心類和邊緣類的帖均熱度存在統計學差異, 前者較後者的熱度顯著更高。

表三 核心—邊緣內容的熱度均值T檢驗

	平均值	<i>t</i>	Sig (雙尾)
帖均熱度	核心類 0.21	16.10	.00
	邊緣類 0.14		

## 內容類之間表現出「趨同引力」結構

H3指出相似的內容具有相似的熱度，即相似度越高的兩個帖子，它們的熱度越會趨於同步。那麼可以反推出，內容之間的相似度越低，彼此之間的熱度差異就越大。因此，任意兩個話題的相似程度，與熱度之差的絕對值成反比。同時，依據H1和H2，越相似於熱門，越可能成為熱門，而越熱門的內容之間的相似度又越高。越有熱度的資訊擁有更多的共性特徵，彼此之間有很大的吸引力，越趨向於同質化，即是說，任意兩個內容類型的相似程度與其熱度之和成正比。將以上兩點概括為推論三：平台存在內容的趨同引力結構，任意兩類內容的相似度與其熱度之和成正比、與熱度之差成反比。

利用已經抓取到400類微博內容類的數據，計算兩兩之間的相似度、熱度之和以及熱度之差的絕對值，共計 $400 \times 400 - 400 = 159,600$ 條數據（去掉400類和自身的相似度）。通過SPSS軟體，對該數據進行相關性檢驗與迴歸分析。

在相關性檢驗中，兩兩內容類彼此的相似度與熱度之和及熱度之差的絕對值的相關性分別為.53 ( $p < .01, N = 159,600$ )和-.18 ( $p < .01, N = 159,600$ )。任意兩類內容的相似度與其熱度之和、熱度之差的絕對值都在.01的顯著性水準上相關，說明內容引力結構存在。

在迴歸分析中，殘差直方圖和正態P-P圖滿足了線性迴歸模型對於殘差正態分布的要求。變異數分析 (analysis of variance, ANOVA) 中顯著性 $p$ 為.000 < .001，模型成功建立。共線性診斷顯示引數之間基本不存在共線性。其分析結果與皮爾遜相關分析吻合，熱度之和 ( $\beta = .56, p < .001$ ) 顯著正向預測內容相似度，熱度之差的絕對值 ( $\beta = -.12, p < .001$ ) 顯著負向預測內容相似度。迴歸方程的 $R^2 = .29$ 。可以認為，兩個議題越有熱度，它們之間就越相似，內容類之間有一種趨同引力結構關係。

## 結論

與一些過往研究有所不同，本研究的結果反對社交媒體平台內容去中心的觀點，指出熱門內容向中心趨同。研究以新浪微博為樣本，

《傳播與社會學刊》，(總)第65期(2023)

對微博資訊內容在相似度矩陣中的中心性與其熱度間的關係與結構進行實證分析，旨在考察社交媒體平台的內容是否正在形成一種隨著熱度增長而梯度增長的趨同態勢，並揭示這種趨同的梯度漸變結構。

本文的主要結論是：(1) 社交媒體平台中看似多樣的熱門內容擁有「中心性」。平台本質上是一個資訊內容的「黑洞」，只有少部分中心內容可以獲得高可見性。(2) 平台的熱門內容存在內在關聯性，而非彼此孤立的類型。以微博為例，熱門議題實際上共存於一個收窄的資訊帶寬之中，通過一些固定特徵將它們綁定在一起。(3) 頭部的內容類並不多元。相似的內容類之間具有相似的熱度，處於越高熱度的內容類，彼此之間的相似度越高。平台的流行密碼，就是中心度的階梯增長。(4) 內容在平台上的熱度可以通過內容在相似度網絡的中心性程度來預測。具體而言，指出了以下可能的現實問題：

其一，平台坍塌帶來了熱門內容窄化。大量的內容呈現出長尾分布的狀態，只有少部分流行內容可以頻繁出現在人們的視野中。為了融入群體、獲得回應，用戶群會對平台中的高熱度事件予以關注並作出基本符合大眾呼聲的回應，於是每個人都在無意識地向群體中心意見聚攏。微博用戶通過關注列表裏的「大V」和熱搜榜單完成對世界的感知，在點讚、評論的交互中對自我意見進行修正，在算法推薦、分區設計和人工審核的平台混合把關機制下，被引導到某些流行的大眾選擇中去，導致平台內容從豐富走向狹隘。與此同時，熱門類目之間互相模仿、彼此趨同，逐漸形成「抱團取暖」的緊密結構，使得異質性的內容在平台內部始終受到排擠，平台越來越被封印於非常有限的邊界之中。由於這種內容機制所蘊藉的強大的窄化與同化效力，每個用戶的漫遊路徑大多囊括於既定的熱門「光譜」之中，看似五光十色的內容實則是千篇一律的內核。

其二，平台背後的資本和權力操控著社群注意力。平台的有界封印性側面證明了輿論可以被引導控制，必須警惕技術的負面效應。本文的實證結果顯示，平台本質上如同一個「資訊黑洞」，且越靠近黑洞的中心，內容就越是多連接點、高集中性的，擁有更多的能量和熱量。隨著「眼球經濟」時代的到來，在一定利益動機的驅使下，如果不



加以規制，一些資本及權力集團很容易被熱度所裹挾，一味圍繞平台中心進行內容堆砌，使得平台內容越發被圈在一個很狹小的空間之中。

本研究的結論提示我們，在利用平台獲得資訊時，需要警惕其內容封閉性結構。以微博為代表的社交媒體平台看似可以保證內容的公共性，為我們提供全面豐富甚至個性化定制的推送，但在海量資訊的背後，是一種內容隨著其熱度增加而越發向中心聚攏的結構和趨勢。如果每日固定只從某一個平台獲取內容並只關注其熱搜推薦，我們所接觸的資訊其實是非常窄化和趨同的，久而久之，會將我們困在群體性的資訊繭房之中，更危險的是，這種趨同結構的統一性是潛在的，隱藏在多元的分區和主題背後，更具有迷惑性，更容易被平台背後的「人」所設計。因此當下有必要提升公眾數字媒介素養和加強權力監管。另一方面，本文的結論也為「冷啟動」等平台運營難題提供了應對思路：與現存熱門類別進行比對，從而預測其熱度趨勢。如果中心度比較高，則可以推測其有望成為熱門；如果與熱門差異過大，則可以推斷較為小眾，從而實現對初次生產的新內容的簡單篩選和有效投放。

本研究的貢獻和創新之處主要體現在以下方面：(1) 指出了以微博為代表的社交媒體平台所存在的信息同質化的問題，不同於個體層面的「資訊繭房」，這是一種平台整體性的趨同現象；(2) 揭示了微博平台熱門內容的不同類別之間存在一種有規律的趨同演化結構，這是一種內容的中心度會隨著熱度的增長而越來越階梯強化的結構和趨勢；(3) 研究關注的重點不在於哪些內容可以聚為中心類別，不在於具體的類型，而是點明了社交媒體平台內部本身就存在這樣一種封閉性的有界封印結構，使得資訊隨著熱度增長而越發趨同；(4) 本研究發現的這種趨同結構打破了以往研究中認為媒體平台內容正走向碎片化和分化的觀點，強調了多元背後的潛在統一性。

本研究沒有探討這種熱門內容趨同的現象除了內容熱度與中心度的關聯結構以外，可能受到的其他外部因素影響。社交媒體平台並非在真空中運營，這種趨同演化的結構在現實中會與日常生活的許多方面產生接觸和互動，因而可能會受到更為複雜和豐富的因素影響和制約，例如內容的情緒維度、平台的文化背景、用戶的素養和人口特徵、

《傳播與社會學刊》，(總)第65期(2023)

推送算法機制等等。這些因素在其中扮演了怎樣的角色，起到了甚麼作用，有待其他研究進一步的深入。

社交媒體平台是大眾獲取公共資訊最主要的途徑之一，卻陷入了同質化風險中。本研究以新浪微博為例，但得出的結論並非只在微博平台適用。在新媒體時代，面對可能產生的公共性的資訊繭房的問題，該現象和結構值得我們繼續深入挖掘。熱門內容之間逐漸趨同乃至最終形成平台封印，其背後擁有何種邏輯和機理？具體是怎樣運行的？這種公共性的資訊繭房又是如何隨著時間發展而動態演化的、演化趨勢是怎樣的？這些問題有待在未來研究中回應。

## 參考文獻

### 中文部分 (Chinese Section)

- 王國華、陳靜、鍾聲揚 (2015)。〈微博熱門話題及其線下行為轉化研究〉。《電子政務》，第11期，頁59–71。
- Wang Guohua, Chen Jing, Zhong Shengyang (2015). Weibo remen huati jiqi xianxia xingwei zhuanhua yanjiu. *Dianzi zhengwu*, 11, 59–71.
- 王國華、鄭全海、王雅蕾、熊偉強、謝宏 (2014)。〈新浪熱門微博的特徵及用戶轉發規律研究〉。《情報雜誌》，第4期，頁117–121。
- Wang Guohua, Zheng Quanhai, Wang Yalei, Xiong Weiqiang, Xie Hong (2014). Xinlang remen weibo de tezheng ji yonghu zhuanfa guilü yanjiu. *Qingbao zazhi*, 4, 117–121.
- 李彪 (2013)。〈微博中熱點話題的內容特質及傳播機制研究——基於新浪微博6025條高轉發微博的數據挖掘分析〉。《中國人民大學學報》，第5期，頁10–17。
- Li Biao (2013). Weibo zhong redian huati de neirong tezhi ji chuanbo jizhi yanjiu——Jiyu Xinlang Weibo 6025 tiao gao zhuanfa weibo de shuju wajue fenxi. *Zhongguo renmin daxue xuebao*, 5, 10–17.
- 汪小帆、李翔、陳關榮 (2012)。《網絡科學導論》。北京：高等教育出版社。
- Wang Xiaofan, Li Xiang, Chen Guanrong (2012). *Wangluo kexue daolun*. Beijing: Gaodeng jiaoyu chubanshe.
- 易艷剛 (2015)。〈同類新聞為甚麼喜歡紮堆〉。《青年記者》，第24期，頁4。

- Yi Yangang (2015). Tonglei xinwen weishenme xihuan zadui. *Qingnian jizhe*, 24, 4.
- 韋路、胡雨濛 (2014)。〈中國微博空間的議題呈現：新浪熱門微博實證研究〉。  
《浙江大學學報(人文社會科學版)》，第2期，頁41–52。
- Wei Lu, Hu Yumeng (2014). Zhongguo weibo kongjian de yiti chengxian: Xinlang remen weibo shizheng yanjiu. *Zhejiang daxue xuebao (Renwen shehui kexueban)*, 2, 41–52.
- 夏雨禾 (2010)。〈微博互動的結構與機制——基於對新浪微博的實證研究〉。  
《新聞與傳播研究》，第4期，頁60–69、110–111。
- Xia Yuhe (2010). Weibo hudong de jiegou yu jizhi——Jiyu dui Xinlang Weibo de shizheng yanjiu. *Xinwen yu chuanbo yanjiu*, 4, 60–69, 110–111.
- 徐翔、王雨晨 (2021)。〈社會回音室：網絡傳播中的信息趨同及其媒介邏輯——基於「今日頭條」樣本的計算傳播學分析〉。《國際新聞界》，第7期，頁99–124。
- Xu Xiang, Wang Yuchen (2021). Shehui huiyinshi: Wangluo chuanbo zhong de xinxi qutong jiqi meijie luoji——Jiyu “Jinritoutiao” yangben de jisuan chuanboxue fenxi. *Guoji xinwen jie*, 7, 99–124.
- 徐翔、劉佳琪、靳菁 (2021)。〈微博空間中的意見典範用戶及其作用路徑研究〉。《新聞大學》，第7期，頁14–32、119–120。
- Xu Xiang, Liu Jiaqi, Jin Jing (2021). Weibo kongjian zhong de yijian dianfan yonghu jiqi zuoyong lujing yanjiu. *Xinwen daxue*, 7, 14–32, 119–120.
- 翁小蘭、王志堅 (2018)。〈協同過濾推薦算法研究進展〉。《計算機工程與應用》，第1期，頁25–31。
- Weng Xiaolan, Wang Zhijian (2018). Xietong guolü tuijian suanfa yanjiu jinzhan. *Jisuanji gongcheng yu yingyong*, 1, 25–31.
- 張文汐 (2014)。《新浪微博熱門話題的特點與規律研究》。遼寧大學傳播系碩士論文。
- Zhang Wenxi (2014). *Xinlang Weibo remen huati de tedian yu guilü yanjiu*. Liaoning daxue chuanboxi shuoshi lunwen.
- 張振亞、王進、程紅梅、王煦法 (2005)。〈基於餘弦相似度的文本空間索引方法研究〉。《電腦科學》，第9期，頁160–163。
- Zhang Zhenya, Wang Jin, Cheng Hongmei, Wang xufa (2005). Jiyu yuxian xiangsidu de wenben kongjian suoyin fangfa yanjiu. *Diannaokexue*, 9, 160–163.
- 曹林 (2011)。〈微博傳播的十大特點及對言論生態的影響〉。《新聞記者》，第9期，頁29–34。

《傳播與社會學刊》，(總)第65期(2023)

- Cao Lin (2011). Weibo chuanbo de shida tedian ji dui yanlun shengtai de yingxiang. *Xinwen jizhe*, 9, 29–34.
- 陳昌鳳 (2014)。〈同質化的信息消費：2013年我們在關注甚麼——中國四地2013年谷歌搜索熱點數據分析〉。《新聞與寫作》，第2期，頁50–52。
- Chen Changfeng (2014). Tongzhihua de xinxi xiaofei: 2013 nian women zai guanzhu shenme——Zhongguo sidi 2013 nian Guge sousuo redian shuju fenxi. *Xinwen yu xiezu*, 2, 50–52.
- 曾子倩 (2019)。《個性化推薦對網絡信息行為的窄化影響研究》。北京郵電大學經濟與管理學院碩士論文。
- Zeng Ziqian (2019). *Gexinghua tuijian dui wangluo xinxi xingwei de zhaihua yingxiang yanjiu*. Beijing youdian daxue jingji yu guanli xueyuan shuoshi lunwen.
- 楊洸、余佳玲 (2020)。〈新聞算法推薦的信息可見性、用戶主動性與信息繭房效應：算法與用戶互動的視角〉。《新聞大學》，第2期，頁102–118、123。
- Yang Guang, She Jialing (2020). Xinwen suanfa tuijian de xinxi kejianxing, yonghu zhudongxing yu xinxi jianfang xiaoying: Suanfa yu yonghu hudong de shijiao. *Xinwen daxue*, 2, 102–118, 123.
- 翟秀鳳 (2019)。〈創意勞動抑或算法規訓？：探析智能化傳播對網絡內容生產者的影響〉。《新聞記者》，第10期，頁4–11。
- Zhai Xiufeng (2019). Chuangyi laodong yihuo suanfa guixun?: Tanxi zhinenghua chuanbo dui wangluo neirong shengchanzhe de yingxiang. *Xinwen jizhe*, 10, 4–11.

### 英文部分 (English Section)

- Airoldi M., Beraldo, D., & Gandini, A. (2016). Follow the algorithm: An exploratory investigation of music on YouTube. *Poetics*, 57, 1–13.
- Ashokkumar, A., & Pennebaker, J. W. (2021). Social media conversations reveal large psychological shifts caused by COVID-19's onset across US cities. *Science Advances*, 7(39), 7843.
- Bandari, R., Asur, S., & Huberman, B. (2021). The pulse of news in social media: Forecasting popularity. *Proceedings of the International AAAI Conference on Web and Social Media*, 6(1), 26–33.
- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555–564.
- Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4), 375–395.

- Csermely, P., London, A., Wu, L. Y., & Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2), 93–123.
- Eggemann, N., & Noble, S. D. (2011). The clustering coefficient of a scale-free random graph. *Discrete Applied Mathematics*, 159(10), 953–965.
- Figueiredo, F., Almeida, J. M., Benevenuto, F., & Gummadi, K. P. (2014, April). Does content determine information popularity in social media? A case study of YouTube videos' content and their popularity. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 979–982). Association for Computing Machinery.
- Goldszmidt G., & Yemini Y. (1995, May). Distributed management by delegation. In *Proceedings of 15<sup>th</sup> International Conference on Distributed Computing Systems* (pp. 333–340). IEEE.
- Guan, L., Liang, H., & Zhu, J. J. (2022). Predicting reposting latency of news content in social media: A focus on issue attention, temporal usage pattern, and information redundancy. *Computers in Human Behavior*, 127, 107080.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm as 136: A *k*-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.
- Hogan, B. (2011). Visualizing and interpreting Facebook networks. In Derek L. Hansen, Ben Shneiderman, & Marc A. Smith (Eds.), *Analyzing social media networks with NodeXL* (pp. 165–180). San Mateo, CA: Morgan Kaufmann.
- Hong, L., Dan, O., & Davison, B. D. (2011, March). Predicting popular messages in Twitter. In *Proceedings of the 20<sup>th</sup> International Conference Companion on World Wide Web* (pp. 57–58). Association for Computing Machinery.
- Jennings, W., Bevan, S., & John, P. (2011). The agenda of British government: The speech from the throne, 1911–2008. *Political Studies*, 59(1), 74–98.
- Lee, R. K. W., Hoang, T. A., & Lim, E. P. (2017, April). On analyzing user topic-specific platform preferences across multiple social media sites. In *Proceedings of the 26<sup>th</sup> International Conference on World Wide Web* (pp. 1351–1359). International World Wide Web Conferences Steering Committee.
- Masson, A. R., Hayel, Y., & Altman, E. (2017). Posting behaviour dynamics and active filtering for content diversity in social networks. *IEEE Transactions on Signal & Information Processing Over Networks*, 3(2), 376–387.
- Mikal, J. P., Rice, R. E., Kent, R. G., & Uchino, B. N. (2014). Common voice: Analysis of behavior modification and content convergence in a popular online community. *Computers in Human Behavior*, 35, 506–515.
- Mikal, J. P., Rice, R. E., Kent, R. G., & Uchino, B. N. (2016). 100 million strong: A case study of group identification and deindividuation on Imgur. com. *New Media & Society*, 18(11), 2485–2506.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a, May 2-4). *Efficient estimation of word representations in vector space*. Paper presented at Proceedings of Workshop at ICLR 2013, Scottsdale, AZ, United States.

《傳播與社會學刊》· (總) 第 65 期 (2023)

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b, December 5–10). *Distributed representations of words and phrases and their compositionality*. Paper presented at Proceedings of Workshop at NIPS 2013, Lake Tahoe, NV, United States.
- Nikolov, D., Lalmas, M., Flammini, A., & Menczer, F. (2019). Quantifying biases in online information exposure. *Journal of the Association for Information Science & Technology*, 70(3), 218–229.
- Patton, D. U., Frey, W. R., & Gaskell, M. (2019). Guns on social media: Complex interpretations of gun images posted by Chicago youth. *Palgrave Communications*, 5(1), 1–8.
- Pinto, S., Albanese, F., Dorso, C. O., & Balenzuela, P. (2019). Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and Its Applications*, 524, 614–624.
- Řehůřek, R., & Sojka, P. (2011, August). Gensim-statistical semantics in python. *EuroScipy*. Retrieved from [genism.org](http://genism.org).
- Shen, D., Wang, G., Wang, W., Min, M. R., Su, Q., Zhang, Y., Li, C., Henao, R., & Carin, L. (2018, July 15). *Baseline needs more love: On simple wordembedding-based models and associated pooling mechanisms*. Paper presented at Proceedings of ACL 2018, Melbourne, AUS.
- Tajfel, H. (1986). The social identity theory of intergroup behavior. *Psychology of Intergroup Relations*, 13(3), 7–24.
- Takeshita T. (2006). Current critical problems in agenda-setting research. *International Journal of Public Opinion Research*, 18(3), 275–296.
- Tan, Y., & Weaver, D. H. (2013). Agenda diversity and agenda setting from 1956 to 2004: What are the trends over time? *Journalism Studies*, 14(6), 773–789.
- Van Alstyne, M., & Brynjolfsson, E. (1996). Could the internet balkanize science? *Science*, 274(5292), 1479–1480.
- Zhao, Z. (2021). Propagation structure feature of entertainment news in the Weibo online social network. *EPL (Europhysics Letters)*, 135(1), 16002.
- Zhu, J. H. (1992). Issue competition and attention distraction: A zero-sum theory of agenda-setting. *Journalism Quarterly*, 69(4), 825–836.

## 本文引用格式

徐翔、孫天怡 (2023)。〈平台「熱趨同」：基於新浪微博的內容熱度與相似度網絡中心度關聯結構分析〉。《傳播與社會學刊》，第 65 期，頁 125–154。