

研究論文

社交媒體使用中使用者內容生產的同質化 ——基於新浪微博樣本的使用者 相似度分析

徐翔、周詩凡、靳菁

摘要

社交媒體使用者在內容的生產和表達中，儘管包含著豐富的內容流動和亞文化、去中心性、多樣化主體，但是也隨著社交媒體使用程度的加深而表現出使用者在內容上的不斷加強的均勻同質化。不同於個體之間或者社群等小尺度範圍上的同質化，這種同質化指的是與該媒介全域內容、全體使用者生產內容的趨同。不是與特定個體的同質化，而是受到大規模乃至全體使用者場的內容引力後的「平均相似度」提升。以新浪微博的7,825個使用者及其內容生產為樣本，本研究從帖子內容同質化、個體同質化、範本同質化這三個層面，考量用戶與微

徐翔，同濟大學藝術與傳媒學院教授。研究興趣：社交媒體、計算傳播。電郵：xuxiang@tongji.edu.cn

周詩凡，同濟大學藝術與傳媒學院碩士。研究興趣：社交媒體、計算傳播。電郵：894625444@qq.com

靳菁，清華大學新聞與傳播學院博士生。研究興趣：計算傳播、政治傳播。電郵：644731833@qq.com

論文投稿日期：2020年10月19日。論文接受日期：2021年4月1日。

《傳播與社會學刊》，(總)第60期(2022)

博整體的平均趨同度的提升；通過句向量的餘弦相似度和「類平均法」對同質化程度的測量，分別從用戶在微博的捲入度、活躍度、影響度作為使用程度的主要變數並將其劃分為不同的「使用者層級」，進而考察以下兩者之間的關聯效應。結果顯示，隨著微博使用程度的加深，用戶與整體以及範本的同質化和平均的相似度，都顯著加強。

關鍵詞：社交媒體、微博、同質化、標準化、計算傳播學

Research Article

Homophily of User-Generated Content in Social Media Usage: User Similarity Analysis Based on Sina Weibo

Xiang XU, Shifan ZHOU, Jing JIN

Abstract

Social media contain rich information. Although they have been described as decentralized and diversified subcultures, it has been demonstrated that users continue to strengthen homophily in the process of using social media. Such homophily is not generated by a community of individuals. Instead, it focuses on user-generated content on a large scale by all users. Instead of considering homophily aimed at specific individuals, this study focuses on the promotion of an “average similarity” driven by content gravity in the entire field. The study sample comprised 7,825 users of Sina Weibo and their generated content. The phenomenon was examined on three types of homophily: similarity in content, similarity between individuals, and similarity in imitating top users. By calculating the cosine similarity of the sentence vector and the unweighted pair-group method with arithmetic (UPGMA), we measured the degree of homogeneity. Three variables—involvement, activity, and impact—were

Xiang XU (Professor). College of Arts and Media, Tongji University. Research interests: social media, computational communication.

Shifan ZHOU (Master of Arts). College of Arts and Media, Tongji University. Research interests: social media, computational communication.

Jing JIN (Ph. D. candidate). School of Journalism and Communication, Tsinghua University. Research interests: computational communication, political communication.

《傳播與社會學刊》，(總)第60期(2022)

obtained to categorize different users. Our results showed a correlation between the above two aspects and demonstrated that the three types of homophily were strengthened significantly by involvement in social media.

Keywords: social media, Sina Weibo, homophily, standardization, computational communication

Citation of this article: Xu, X., Zhou, S., & Jin, J. (2022). Homophily of user-generated content in social media usage: User similarity analysis based on Sina Weibo. *Communication and Society*, 60, 123–156.

致謝

本文為國家自然科學基金項目「社交網絡互動中用戶『資訊窄化』機理分析：基於微博的數據挖掘」(項目批准號：71804126)和上海市級科技重大專項「人工智慧基礎理論與關鍵核心技術」(2021SHZDZX0100)的階段性成果。

引論：問題的提出

當今全球範圍內，社交媒體強勢崛起，成為使用率和普及率最高的應用之一，並對傳統的媒介環境和媒介生態構成重大的衝擊與改變。「人人都有麥克風」、「人人都有傳聲筒」、「草根」和多元化群體廣泛崛起的社交媒體背景下，網路被認為是「去中心」和開放的，內容是共創和共用的。新型的社交媒體作為媒介技術，為人們在社交媒體的使用、表達、資訊和文化接受方面帶來更多的主動性和能動性，而不必受制於馬克·波斯特 (Mark Poster) 所指出的「第一媒介時代」(The First Media Age) 及此類中心化的傳播方式。馬克·波斯特 (2000) 認為，以互聯網為代表的新媒介的出現意味著「第二媒介時代」(The Second Media Age) 的到來，多樣的媒介形式和多元的傳播方式與傳統的大眾媒介時代完全區分開來。尼葛洛龐帝 (Negroponte) 的「我日報」(the daily me) 等著名概念的提出，隱含著主體被智慧推送和匹配技術所包裹的環境，也表達著一種對於主體在網路中日趨差異化的想像 (Negroponte, 1995)。

對於網路去中心性和多元化的特徵，也有不少研究通過實證的方法加以關注。如陳然 (2012) 通過對網路論壇活躍分子的「關注」資料分析，提出論壇活躍群體的社交網路呈現較為扁平化和去中心化的特點。張一璐、倪靜、郭強以及劉建國 (2014) 將資訊熵作為對線上使用者多樣性的檢測標準，熵值越大表示這類用戶的興趣越多樣，其通過實證分析認為一般的用戶確實存在興趣多樣的現狀。Van Alstyne 和 Brynjolfsson (1996) 對網路巴爾幹化的分析表明資訊技術消除地理空間巴爾幹的同時，構築起基於學科專業、社會階層、個人偏好、社會文化的分化和相互之間差異性的壁壘。一些研究認為，從多樣性和演算法的關係來看，不同於「回音室效應」(the echo chamber effect) 可能造成的有偏好傾向的傳統觀點，較多的研究認為演算法推薦機制正在促進新的內容多樣性。這種多樣性不再局限於新聞來源或是政治觀點的多元，而是更加廣義地涉及了主題、體裁，甚至是書寫語氣的多樣性分佈 (Flaxman, Goel, & Rao, 2016; Helberger, Karppinen, & D'Acunzio, 2018)。

《傳播與社會學刊》，(總)第60期(2022)

事實上也有針對網路巴爾幹化的實證研究表明，網路的異質化作用並沒有我們想像那麼顯著。有研究通過對日本和美國網民的調查資料研究互聯網使用對政治資訊暴露的影響，發現網路瀏覽確實有助於接觸與自身態度一致的資訊，但是完全排斥與自身立場相反的資訊卻並沒有得到實證的支援。因此他們認為對使用互聯網的選擇性導致網路巴爾幹化 (cyber balkanization) 的恐懼是被誇大的 (Kobayashi & Ikeda, 2009)。其他針對網路巴爾幹化的研究，許多基於立場相悖的預設。比如有研究針對政治立場不同的群體 (Gross, 2015)，也有研究針對持有不同宗教信仰的群體 (Pons-De Wit, Versteeg, & Roeland, 2015)。但實際上，網民不單單只有一種政治身份或宗教身份，他們也可能成為不同局部同質的組成部分。如有研究表明如今的網民是一種「多重歸屬」(multi-homing) 的使用者，在網路中他們也具有多重選擇性。而即使用戶的興趣多樣，他們仍然會關注相同的資訊來源 (Doganoglu & Wright, 2006)。

同時，網路仍然被認為是一個具有強大影響力的媒體。阿爾弗雷德·李和伊莉莎白·李夫婦曾提出七種強有力的宣傳策略，其中的樂隊花車效應 (the bandwagon effect) 就提出人類害怕在社會中被孤立，因而有向社會其他多數趨同的心理傾向 (Lee & Lee, 1939)。而這一特徵，在網路社會中仍然被證實。魏靜、朱恒民、宋瑞曉、蔣世兵 (2016) 採集真實網路輿情傳遞線上資料，研究發現網路輿情傳遞個體間的「樂隊花車效應」是輿情得以傳遞的重要驅動力。也有學者從資訊演化的博弈模型研究網路意見的趨同現象 (劉錦德、劉詠梅, 2013; 劉亞洲、潘曉中、付偉, 2018)。總的來說，個體的表達傾向仍然會受到整體的意見氣候的影響。

另一方面，媒介和網路文化中的同質化 (homophily) 甚或標準化 (standardization) 的現象和問題亦是具有深刻內涵的命題。在法蘭克福學派 (Frankfurt School) 等的理論闡述中，「文化工業」(culture industry) 認為，文化產品生產的同質化導致了文化消費和藝術審美變得單一。即便今天文化產品的種類因技術進步而得以繁榮，大眾文化多樣而複雜的局面已經超出了阿多諾文化工業批判的理論體系，文化產品的同質化產出依然值得關注。在關於媒介文化標準化的研究中學者指出，

社交媒體使用中使用者內容生產的同質化

在工業化生產模式及商業邏輯的支配下，大眾媒介生產製作出的媒介文化產品呈現出標準化、齊一化、程式化或趨同化的特徵。曾一果(2016)認為今天的傳媒文化工業已經從霍克海默(Max Horkheimer)、阿多諾(Theodor Wiesengrund Adorno)等人所說的雷同性、同一性和複製性，發展成為具有模擬性、虛擬性和超現實性的新媒體文化工業。因此，媒介與媒介文化對人的操控問題在今天依然值得重視。

因此，在社交媒體形塑使用者的媒介環境中，有必要再對多樣個體的線上交互與多元內容的文化「狂歡」(carnival)進行再審視：社交媒體的使用，是否真的助推著用戶「個性化」(personalization)的差異表達與使用者內容生產？在社交網路媒體日趨加深的使用中，對多元需求的滿足是否造就著各有特色、去中心化、差異化而非同質化的使用者內容表達？正是基於這樣的困惑，本研究想要探討社交網路中的內容傳播機制究竟是什麼樣的圖景。我們已經習以為常的資訊傳播方式，是否並沒有說明我們形成個性，反而是以某種難以察覺的方式將我們變得與他人越來越相似。

研究回顧與分析

社交媒體的「文化工業」再審視

「文化工業」的概念最早見於霍克海默和阿多諾(2006)的《啟蒙辯證法》(*Dialektik der Aufklärung*)，其矛頭指向以報紙、雜誌、廣播、電視、電影等為傳播媒介的大眾文化。文化工業對人實行的「異化」(alienation)主要是通過大眾文化的商品化和文化生產的標準化得以實現。文化產品的價值不再是滿足人審美上的需求，而是實現利益的最大化。別無選擇的大眾，也只能在「偽個性」(pseudo-individualization)的文化薰陶下，成為偽個性的人。

作為法蘭克福學派的代表人物之一，馬爾庫塞(Herbert Marcuse)(2008)則深度思考了現代文明「單向度的人」(One-Dimensional Man)的多樣性缺失和異化問題。一方面科技發展極大地促進了生產力的提高，工人階級和其他階級共用繁榮社會的產物，好似自身與他人並無

《傳播與社會學刊》，(總)第60期(2022)

不同而逐漸趨於滿足、感到幸福。另一方面資產階級利用媒介技術不斷地製造「虛假的需求」(artificial need)來教唆人們盡情娛樂、消費。在勞動中感受到壓抑，無法獲得創造的樂趣的人們，將勞動視為追逐其他物質滿足的手段。越是得到滿足，越依附於現有的制度，深陷拜物教的泥沼，想要獲得解放的思想被消磨殆盡。媒體影響下的用戶處於一種工業體系式文化「流水線」(pipeline)的媒介生態中，成為越來越相似的「流水線」上的一顆顆螺絲釘。

在SoLoMo (Social Local Mobile) 社交本地移動的特點下，網路應用逃不開「社交」(social)這一屬性的設置，「關係」(relationship)貫穿和影響整個社交網路的運轉。今天的數位化環境中，被放大的社交像黑洞一樣持續地拖住人們的注意力。按照原本的設想，使用者已然是媒介的主人，重獲了被侵佔的主體性(subjectivity)。然而，高頻的使用讓人們養成了對技術的依賴、對不間斷社交的渴望，一個個網路技術依賴者應運而生。這種數位化的實踐方式對人的主體性進行多層面的弱化，使用者的資訊選擇的傾向性被技術放大，高頻的瀏覽帶來了視野的窄化。網路中，什麼樣的內容特徵得到注意力價值的認可，這樣的特徵會趨於成為內容生產的尺度和標準。對媒介資訊技術的依賴讓媒介對人的主體性形成「異化」作用。

受眾的內容在傳媒的「文化工業」體系和資訊「流水線」機器中走向同質化乃至標準化和勻質性，其相應的實證檢驗還有待豐富與深化。尤其是在方興未艾的社交網路新媒體背景下，作為新時代「內容工業」(content industry)和資訊生產體系的社交網路，是否以及如何生產、在怎樣的條件和程度下生產著相似化的用戶，是一個期待將理論維度與實證大資料分析良好相融的問題場域。

社交媒體使用與同質化生成

儘管社交網路中存在多樣化、「參與文化」(participatory culture)等理論和機制，但是同質化依然是重要的研究命題之一。McPherson和Cook (2001)認為「相似的鳥聚集在一起」(birds of a feather flock

社交媒體使用中使用者內容生產的同質化

together)，這一同質化的原則構成了各種類型的社會紐帶。他們將同質化定義為社交網路連接中的使用者在社會屬性上的個體相似性。他們認為同質化限制了人們的社交領域，這種方式對他們的資訊接收、態度、形式以及他們所經歷的互動具有強大的影響。De Choudhury、Sundaram、John、Seligmann和Kelliher (2010) 將社交網路的同質化定義為社交網路中的使用者更傾向和自己相似的人有更多聯繫。通過考察這種同質性是否建構了使用者的社交網路並影響了他們的交流行為，發現使用者屬性之間的趨同性與資訊擴散之間存在相互作用。這些研究表明社交網路中的使用者本身就會選擇與自己相似的個體進行連接，形成同質化的社交網路。

由此，對於社交網路的同質化研究考察使用者多維屬性和使用行為顯得更為重要。鄭志蘊、賈春園、王振飛和李鈍 (2017) 在對微博用戶相似度的計算研究中，將微博使用者屬性作為分析的依據。其中微博使用者屬性既包括使用者註冊時所填寫的背景資訊 (如性別、年齡、地理等)，同時包括線上的互動資訊 (如關注、粉絲等)。李夢潔、邵曦 (2018) 在對微博使用者屬性相似度的檢測中分別對比了背景相似度 (地點、設備、時間相似度) 和興趣相似度 (文本、關係、轉發相似度)，發現基於文本相似度的相似性分析能更準確的衡量用戶之間的相似性。

同時，社交媒體的使用者是否會傾向於關注相同的資訊這一命題也受到學者們的關注。Weng、Lim、Jiang和He (2010) 分析了Twitter用戶的關注興趣和價值觀的同質性。從新加坡提取的100萬條推文的語料庫中他們發現了Twitter上的朋友傾向於共用興趣的證據。Conover等人 (2011) 嘗試將同質性與政治取向聯繫起來，通過對1,000位用戶的抽樣調查，他們發現有證據表明Twitter上的用戶傾向於從有著相同政治背景的那些用戶轉發更多推文。還有研究則根據英國2010大選的Twitter資料發現黨派結構意義重大，黨員們更可能轉發其政黨的材料 (Boutet, Kim, & Yoneki, 2012)。

通過梳理發現這些研究多側重於個體如何朝向某些重要節點或某些局部使用者發生同質化，而並未充分涉及個體如何發生朝向於全域內容和全體使用者節點的「平均同質化」(average homophily)。事實上，

《傳播與社會學刊》，(總)第60期(2022)

局部的同質化是否必然匯出整體的同質化，是非常有必要論證的問題。因為若局部在網路巴爾幹效應下變成分裂的孤島，那我們將難以得出全域同質化的結論。因此這一命題亟需經驗材料證實。本研究中，我們提出的同質化的首要特徵，便是全域意義上的趨同。即第一個研究假設：

H1：使用者所生產、表達的帖子內容會隨著對社交媒體使用程度的加深，而和社交媒體中整體帖子內容的平均內容相似度越來越高。

這些研究主要是針對人口統計學和關係網絡。這關係到同構結構網絡，即以人為中心的社交網絡在不同的社會、人口、行為和個人內在特徵方面通常是同質的。而這樣的結果是De Choudhury等人(2010)所言的「個體傾向於在人際關係上相互聯繫」，從而影響人的資訊傳播與流通，以及他們所從事的社會角色。由於主要強調人口和社會群體，以及社會連接中的相似性，我們往往容易忽視社交媒體中潛藏的用戶及其與整體同質性的問題。對使用者同質化的研究也需要考慮到個體與整體的關係，由此我們提出第二個假設：

H2：使用者所生產、表達的帖子內容會隨著對社交媒體使用程度的加深，而變得與全體使用者的內容平均相似度越來越高。

從H1和H2的推斷來看，隨著社交媒體使用程度的加深，用戶同質化的程度也可能越來越深，那麼使用程度的差異是否會影響社交網路中同質化現象的趨同規律，也是需要關注的問題。在趨同的勢頭中，處於頂部的用戶，是否是一種「標杆式」的存在？這與「文化工業」的標準化問題暗合。個體與其他個體趨同是否真的遵循著某種標準？整體之中的「排頭兵」是否為同質化最明顯的部分？其他部分是否都在向他們看齊？因此我們在兩個假設的基礎上提出第三個假設：

H3：使用者的使用等級越高，該等級與最高使用等級的使用者之間平均的內容相似度也越高。

這個問題實際上是在拷問，用戶是否會與「標準化範本」(standardized template)形成趨同化演變。它與前兩個問題是一致的。如果前兩者成

社交媒體使用中使用者內容生產的同質化

立，即用戶整體的相似度越來越高，那麼使用度最高的用戶群，也應該和整體的相似度最高；而其他用戶也應該逐漸趨近於最高使用度的用戶。這是三個研究假設在邏輯上的連貫性和一致性。

同質化的內涵界定

首先，前文關於使用者相似性的局部式和微觀性的研究，與理解宏觀、全域意義上的同質化趨向存在差異。後者更多的是批判理論所關注的整體性與社會文化尺度。前者注重從個體之間、高連接用戶之間的相似性測度，而缺乏對於使用者和總體內容、宏觀和大尺度範圍用戶群的相似性的測量及其生成分析。這導致部分使用者之間、局部圈層的相似性並不同於使用者和整體媒介內容場之間趨同性和均勻化，局部使用者的相似性如何生成也不等同於局部和整體之間的相似性如何生成及演化。第二，本研究所指的使用者的UGC內容生產的同質化，是關切於「文化工業」的資訊產品趨同、擬態環境內容「標準化」相關的內容維度，而不是關切於使用者的社會身份、「網路巴爾幹化」、社會網連接的社會人和網路人維度。其三，前述研究對於社交網路使用者在內容、興趣方面的研究，多從使用者外在於媒介使用的條件和因素展開，例如用戶的線上或線下的身份屬性、現實圈層以及網路行為等人口學和社會學特徵，而使用者對於社交網路的媒介使用度如何加深這種同質化，如何參與和影響線上使用者主體化的複雜進程，有待豐富的詳細實證檢驗與考察。

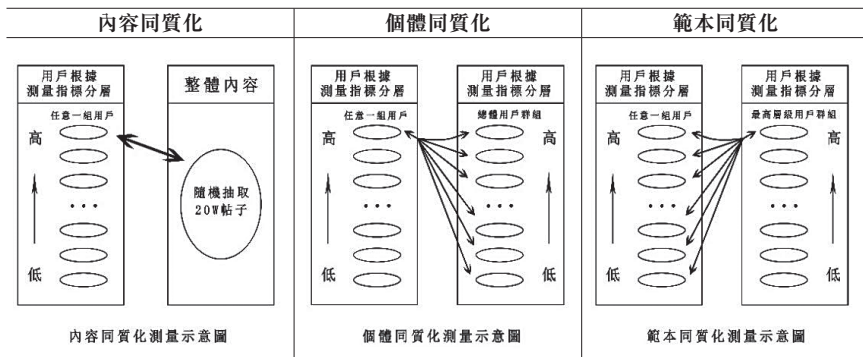
本研究所謂的同質化：其一，不是在較多的實證研究中所常見的個體之間兩兩的相似化、「物以類聚、人以群分」，也不是在社會群體、政治集群、亞文化(sub-culture)、群體「回音室」尺度的特定群體的同質化，而是在該媒介環境意義上全域化的相似性，是使用者與媒介全域各帖子或各用戶的平均相似度。它更接近於批判學派意義上的面向社會整體和宏觀尺度上的同質化，更接近於「標準化」意義上的同質化。其二，這裡的同質化，是指用戶所生產的帖子和資訊內容層面的同質化，而非使用者的人口屬性、社會身份、地位階層、網路行為、

《傳播與社會學刊》，(總)第60期(2022)

社交關係等社會特徵上的同質化，對後者的度量和分析暫不在本文的探究能力範圍內。

綜合上述同質化的內涵，本研究對社交媒體的同質化特徵分為一個條件和三個層面進行考察。首先，一個條件是指用戶對社交媒體的媒介使用程度；其次，在這一條件下探討三個層面，分別是：(1) 內容同質化、(2) 個體同質化、(3) 範本同質化(具體測量思路如圖一)。

圖一 同質化測量示例圖



研究方案與實施路線

研究選取作為中國具有代表性的社交媒體新浪微博。發展至今該平台活躍用戶數量超過四億，已經成為中國互聯網2.0時代的代表性社交媒體。它基於使用者關係形成資訊分享、傳播、獲取機制；即時更新形式多樣的海量內容；促進用戶實現多種交互和實踐活動。其使用者在媒介場域和公共領域中具有重大的影響力和黏性，參與主體具有廣泛性和多元性。由於缺乏進入門檻，草根和明星、大咖共同構築了活躍的資訊流通空間。

同質化的計算方法

根據前文所指的核心問題、一個條件、三個子問題，本研究所針對的使用者內容生產「同質化」最主要的內涵和度量方法是：社交媒體

社交媒體使用中使用者內容生產的同質化

用戶隨著對社交媒體的使用程度提高，其內容的生產和表達，會與其所處的媒介環境中的全域內容、全域使用者趨於平均相似度的提高。

對於內容上的相似度，根據在兩組矩陣中計算平均距離、平均相似度常用的「類平均法」(UPGMA, unweighted pair-group method with arithmetic)展開。基本公式是：

$$d_{pq}^2 = \frac{1}{n_p n_q} \sum_{x_i \in p} \sum_{x_j \in q} d_{ij}^2 \quad (\text{式 } 1)$$

其中設 d_{ij} 為 i 和 j 之間的距離或相似度 (例如，歐式距離、餘弦相似度等)，類均法得到兩組樣本 n_p 、 n_q 的所有元素的兩兩距離之間的平均值，這個值越大，表明各包含多個子元素的物件 n_p 、 n_q 之間平均相似度越高。

本文對使用者的內容相似度計算，涉及到的最為基本的單位為帖子。選擇在文本挖掘、語句相似度計算、自然語言處理中比歐式距離更為常用、也具有高度穩健性的餘弦相似度，來衡量兩條帖子的距離。設任意兩條帖子 (帖子 m 和帖子 n) 之間的餘弦相似度表示為 $S(m, n)$ ，也即上文所示的類平均法計算中所需的 d_{ij} 。對兩條帖子 m 、 n 分別轉換得到兩個向量 A 、 B 之後， A 、 B 的餘弦相似度也即兩個向量 A 、 B 之間夾角 θ 的餘弦，該值範圍在 $[-1, 1]$ ，值越大表明這兩個向量或兩條帖子之間越相似：

$$\cos(\theta) = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (\text{式 } 2)$$

在 $S(m, n)$ 的基礎上，對於用戶與其他用戶或者與其他帖子的平均相似度，分別從以下環節加以表示：

(1) 任意兩個用戶 U_{x1} 和 U_{x2} 之間的內容相似度。設用戶分別為 U_{x1} 和 U_{x2} ，它們分別具有的帖子樣本數量為 n_1 和 n_2 ，則這兩個使用者的內容相似度可表示為：

$$R(U_{x1}, U_{x2}) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} S(i, k) \quad (\text{式 } 3)$$

其中 U_{x1} 和 U_{x2} 都可以有且只有一條帖子，在此情況下：單條帖子 (m) 與單條帖子 (n) 之間的兩兩相似度 $S(m, n)$ ，成為式 3 中 n_1 和 n_2 分別

《傳播與社會學刊》，(總)第60期(2022)

都為1時的特例。對於兩個用戶之間的相似度，通過其所發帖子的相互之間的平均相似度來反映和測量：相似程度越高，表明這些內容之間兩兩的趨近、類同乃至重複程度越高；異質化的內容越多，內容之間差異越大，則這些內容的平均相似程度就會越低。

(2) 任意一個使用者和任意一組帖子之間的內容相似度。設用戶為 U_x ，其中包含 n_1 條帖子；一組帖子構成 U_p ，其中包含 n_2 條帖子。則該用戶與該組帖子相似度為：

$$R(U_x, U_p) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} S(i, k) \quad (\text{式 4})$$

也即這裡把使用者看作是一組內容(帖子)的特定集合，由此式4和式3在計算過程實質上是一致的。

(3) 任意一組用戶 G_1 和另一組使用者 G_2 的內容相似度。可以表示：

$$H(G_1, G_2) = \frac{1}{n_1 n_2} \sum_{x_i \in G_1}^{n_1} \sum_{x_k \in G_2}^{n_2} R(U_{x_i}, U_{x_k}) \quad (\text{式 5})$$

其中 G_1 或 G_2 都可以有且僅有一個使用者，這種情況下也即：式1中所計算的個體與個體之間的兩兩相似度 $R(U_{x_1}, U_{x_2})$ ，成為式5中 n_1 和 n_2 分別都為1時的特例。

式5在本研究中的直接應用：一是計算H2中關於個體同質化的部分涉及到的，不同的「使用度等級」的使用者，與全體用戶之間的平均相似度；二是計算H3中關於範本同質化的部分涉及到的，不同的「使用度等級」的使用者，與最高使用等級的「頂部使用者」之間的平均相似度。

(4) 任意一組用戶和任意一組帖子之間的內容相似度。該情況是式4的推廣情況，也即對於式4重複多次後，對每次計算得到的結果再求其均值，從而得到多個用戶和一組帖子之間的平均相似度。設一組用戶為 G_1 ，其中包含 n_1 個用戶；設一組帖子為 U_p 。則結合式4， G_1 和 U_p 之間的平均相似度為：

$$F(G_1, U_p) = \frac{1}{n_1} \sum_{x=1}^{n_1} R(U_x, U_p) \quad (\text{式 6})$$

式6在本研究中的直接應用：計算H1中關於和全體帖子內容同質化部分涉及到的，各個「使用度等級」的使用者，與全域帖子整體之間的平均相似度。

社交媒體使用中使用者內容生產的同質化

表一 同質化現象的測量內涵及對應公式

同質化現象	內涵	測量方法(涉及到的上述公式)
內容同質化	用戶的各個使用層級在其所生產和表達的內容維度,與該媒介環境中總體意見氣候、總體帖子內容的趨同和相近。	任意一組用戶和任意一組帖子之間的內容相似度(式6),其中一組帖子即隨機抽取的能夠代表整體意見氣候的20萬帖子。
個體同質化	個體同質化包括某一使用層級的用戶和所有用戶形成的整體之間的內容相似度顯著和加強。	任意一組用戶 G_1 和另一組使用者 G_2 的內容相似度(式5),其中 G_2 是所有層級的用戶構成的總體用戶群組。
範本同質化	範本同質化可以理解為使用者隨著其媒介使用度的提升,從而和最高使用度層級的「範本性意見階層」其內容相似度越來越強。	任意一組用戶 G_1 和另一組使用者 G_2 的內容相似度(式5),其中 G_2 指的是「最高使用層級」的用戶構成的群組。

基於 word2vec 詞嵌入對於帖子語義相似性的度量

上述計算過程中,式3、式4、式5、式6都涉及到最為基本的兩條帖子之間的內容相似度計算,也即 $S(m, n)$ 。這涉及到對於帖子、語句轉為向量。對於語句的向量化及其相似度計算,學界從詞形(呂學強、任飛亮、黃志丹、姚天順,2003)、句法特徵(藍雁玲、陳建超,2011)等方面進行了考量。但這基於傳統統計方法,從詞面特徵計算文本相似度,缺乏語義的測定。本研究對於帖子內容的向量化,採取 word2vec 進行轉換。word2vec 是 2013 年提出的淺層神經網路語言模型(Mikolov et al., 2013a; Mikolov et al., 2013b),計算成本低、速度快、精度高,在學界和業界使用廣泛。研究顯示,簡單等權平均得到的句子向量,具有良好的效果,在某些任務中不遜色於複雜深度神經網路的自然語言處理方法(Parikh et al., 2016; Shen et al., 2018; Vaswani et al., 2017)。

首先,輸入語料庫,由於現有的中文語料庫對於社交媒體文本訓練的針對性較弱,我們採用 26G 的中文語料庫進行訓練,內容來源包括媒體新聞庫、網路論壇帖子、經典名著文本等。然後我們通過「連續詞袋模型」(Continuous Bag-Of-Words, CBOW),設定反覆運算次數,通過語料庫詞彙形成對某特定詞彙的概率分佈,訓練出詞彙查找的矩陣。那麼任何一個單詞的向量化表達(one-hot representation)乘以這個矩陣都將得到自己的詞嵌入結果(word embedding)。這就將稀疏的高維形如(0,0,0……0,1,0,0)的矩陣變成了低維實數矩陣。然後我們通過

《傳播與社會學刊》，(總)第60期(2022)

直接對帖子中的每一個詞，取 word2vec 中的對應詞向量，相加後求平均池化的結果，即可得到每條帖子的句向量結果。

word2vec 對於帖子的語義表示較為準確，其建模原理是根據詞語之間在上下文語境的共現關係來對詞語進行向量化的表示，而不是簡單地根據帖子中的詞頻矩陣、TF-IDF 或 LSA、LDA 等基於統計方法的建模。兩條帖子的內容、語義越相近，其句向量的夾角就越小，從而其餘弦相似度就越高。最高的情況下，句子和自身的餘弦相似度為 1。而最差的情況下，兩個句子的餘弦相似度為 -1。由於 word2vec 的建模原理，這些帖子的相似度是基於內容的相似度計算，而不是基於態度、情感的相似度。同時，由於 word2vec 不同於 LDA 主題模型，所以所得到的相似度也不是句子在主題分佈上的相似度。總體而言，word2vec 轉換得到的句向量，對語句模糊含義的把握準確和靈敏，往往可以不需過多地依賴於相同的詞但卻能把握到不同句子的意思，而且這些相似句子的提取也可以在一定程度上兼顧情感色彩，因為詞語在語義的建模過程中，語義中包含的情感成分會由於其相似的上下文語境而不被 word2vec 的人工神經網路一道提取。

對於用戶的微博使用程度的操作性界定

本研究中涉及到的兩個關鍵方面的變數，一是新浪微博的使用程度；二是對於內容同質性的衡量與量化呈現。

本研究中所指的使用程度，主要是指用戶對於社交媒體(新浪微博)的捲入、投入和產生的實際效果等。結合微博中可實際獲取得到、具有可用性的使用者個體資訊，我們選用了以下代表性指標，來考察和轉換用戶的微博使用程度：一是用戶對微博的接受、跟進的歷史情況和捲入程度。本處用用戶在微博的註冊時長來表示，註冊時長越長，代表使用微博的時間越久。同時這也表示該用戶與微博平台發生接觸、交互、作用的可能性以及受到該媒體空間的塑造作用的時間和程度也可能越深。二是用戶在微博僅是註冊、及時的接受與跟進還不夠，還需要在這個媒介空間中有充分的使用頻率、活躍性以及在此媒

社交媒體使用中使用者內容生產的同質化

體中的實際使用行為。本處用以下微博平台自身的官方指標進行衡量：

(1) 用戶的微博經驗值：完成相應的「成長任務」即可獲得相應經驗值。這些任務包括：在微博用戶端啟用手機通訊錄訪問許可、保持每天都登錄微博、轉發或者發佈微博、通過充值開通微博會員等。由此可見，這些經驗值「成長任務」多指向用戶對於微博平台投入的時間、金錢、精力甚至信任等多方面的成本。

(2) 用戶的關注者數量。關注他人的數量在一定程度上反映著用戶在社交媒體中使用的主動性和交互行為的活性。

(3) 對於微博用戶具有實際的使用行為的基礎上，我們還需要考量使用者的使用效果。本處採取用戶的最為直觀可見的指標之一「粉絲數」進行衡量，這也體現著用戶的影響力，微博空間中對於多數用戶而言，高度的捲入、對於其他使用者有吸引力的內容、有效的社交網路行為、活躍的投入等使用因素都與其可能獲得的影響力和粉絲數之間存在著關聯。

上述三個層面，是分別遞進的層面：最先的媒介接觸和媒介捲入（註冊時長）→媒介使用過程的活性（用戶微博經驗值、關注者數等）→媒介使用獲得的實際效果和使用影響（粉絲數）。

選這三個方面的維度及其可操作性的四個指標，一定程度上可刻劃出用戶在微博中的使用和捲入的差異。此外，由於量化指標的獲取需要具有可操作性，是實際能獲取而且成本也在可控範圍的，因此選擇了便於獲取的粉絲數、微博等級等官方資料，而未採取其他更複雜，或需要經過問卷和二次計算等方式才能得到的指標。同時，採取的這些子指標也具有較好的經驗直觀性與非黑箱化的易理解性。

微博資料獲取與預處理

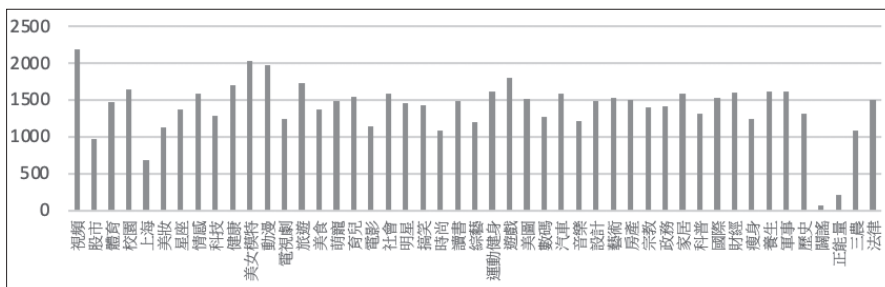
I. 資料獲取與準備

首先，課題組運用開源網頁文本抓取工具「八爪魚」，以及自行用python和selenium編寫的動態網頁抓取程式，抓取新浪微博的使用者資料及其發佈的帖子。樣本使用者來自於對新浪微博首頁每天熱門微博

《傳播與社會學刊》，(總)第60期(2022)

的累積抓取，持續一個月每天抓取2次帖子，一個月共得到67,362條樣帖(去重後為65,650條「種子」帖)，從這些樣帖中成功抓取到使用者資訊和文本的有10,037個發佈者。以上帖子廣泛地涉及47個門類(如圖二)，本研究出於規模和成本所限，未採用大規模的隨機漫步等抽樣方法，但由於採樣時間持續了一個月，並非某個短時間內的抽取；而且樣本的選取結合了新浪微博自身的內容分發系統，廣泛而大致均衡地分佈在47個大的內容類別型版塊，因而也體現出較大覆蓋面和良好程度的代表性。

圖二 種子帖在47個版塊中的數量分佈圖



採集這些發佈者的URL資訊，並在此基礎上，進一步採集10,037個用戶的用戶名、性別、所在地、粉絲數、關注數、發佈微博數、註冊時間、等級、會員資訊等多種資訊。根據此前採集的使用者URL位址，對用戶發佈的微博帖子內容進行抓取，初始抓取得到微博數量34,892,987條。發佈者廣泛涉及到各個層面，粉絲數從低層的草根到高端的「中V」或「大V」(在微博獲得個人認證，擁有眾多粉絲的微博用戶)，因此樣本具有良好的覆蓋面和代表性。粉絲數範圍在各個流量層級上均有分佈，具體分佈類型如表二所示。考慮到抓取的規模和硬體設施，本研究資料抓取一共歷時兩個月，資料抓取完畢的時間是2018年11月，每人抓取的帖子都是自抓取時間起，按時間的先後順序抓取在此之前發佈的最多4,500條帖子，每個用戶由於發佈的帖子數量並不相同，因此再從所抓到的每個用戶的最新帖子中，一律簡單隨機抽取3,000條。

社交媒體使用中使用者內容生產的同質化

由於初始抓取中，考慮到抓取時的網路環境穩定性等因素，每個用戶被抓取到的帖子總體穩定在3,000條至4,000條，因此在二次處理中，我們統一選擇了3,000條作為每個用戶被抓取的帖子數量，這是因為樣本量過小，難以完全把握使用者的總體特徵；樣本量過大，受採集條件限制難以做到每個用戶均採集條數超過5,000條，因此從「察其帖，觀其人」的角度，我們選擇每位用戶3,000條帖子作為分析樣本。這一步驟是通過random模組在python中完成的。最後得到的有效用戶為7,825個。中文分詞工具採用常用的jieba，詞語和句子向量、矩陣的計算採取常用的numpy和sklearn，用python規則運算式對文本進行簡體漢化、標點符號清洗、分詞等處理。

II. 使用者資料分佈情況

另外，由於用戶使用程度的指標，具有量綱不一致的問題，例如微博官方自帶等級是定序單位，而粉絲數等則是從數百到成百上千萬的差異巨大定距、定比數值。因此，本研究一律將其轉為定序層次的序數來考量，每個指標分為30個等級。例如，儘管每個用戶的粉絲數差異巨大，但是一律將其平均分為30個等級，每個等級按照分位元數切分因而具有同等的個體數量；其他指標的層次劃分也依此類推。

這一方面有助於克服不同指標量綱不一致以及數值差異過大的問題；也有助於不必受擾於對使用程度的切分點問題，用使用程度的高低位序來反映不同用戶和不同層級的差異；另一方面，將使用者等分為不同的層次，也有助於在每個層級使用者的數量均衡的情況下，更好地考察各層次與內容、其他使用者的相似情況和分佈特徵。同時，本研究基於對範本同質化的研究需要，因必須將不同層級的用戶，在分層之後與最高層級的用戶其平均相似度分別進行計算和比較，所以也將使用者採取相等樣本數的等分，考慮其前後高低的「定序」指標和均勻使用者規模切分，而避免因採用其定距、定比的更為精細的指標帶來量綱不一致、切分不均衡等問題，影響計算結果。

本研究將用戶在不同指標上分成若干層級，考察不同層級和總體文本相似度的關聯，因此有必要展示抓取到的資料在不同層級上的分佈情

《傳播與社會學刊》，(總)第60期(2022)

況。按照30層級的劃分，第一層級視作該指標上的最小數值，第30層級為該指標的最高範本用戶。以此類推。如表二，所抽取的7,825個用戶，在微博經驗值、註冊時長、粉絲規模等方面，各指標的覆蓋面都很充分而廣泛，例如粉絲規模從幾十到上千萬不等，註冊天數從數十天到數千天不等，使得「分層」後的分析具有良好的可行性與代表性。

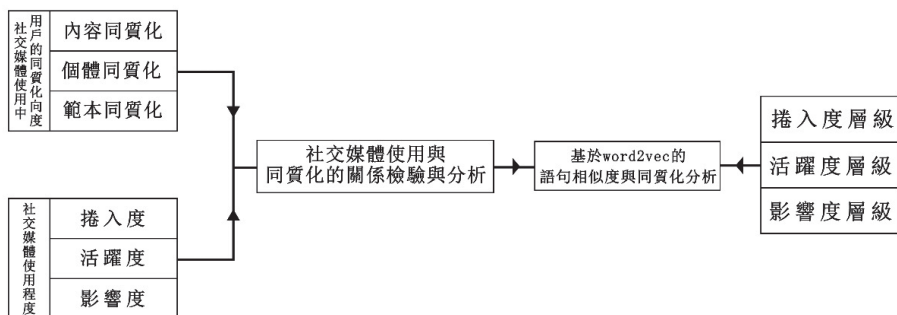
表二 原始資料在各個指標不同層級的分佈情況

指標	註冊時長(天)	微博用戶經驗值	用戶關注其他帳號數量	用戶粉絲數量
分層級數	28	18	30	30
平均數	2,391	23,549	908	808,239
中位數	2,598	25,860	522	3,718
標準差	691	8,248	1,207	3,476,931
極小值	466	60	0	0
極大值	3,539	31,080	20,000	104,187,335
分位數	25	1,870	20,010	265
	50	2,598	25,860	522
	75	2,954	31,080	1,057

整體思路與實施路線

整體研究思路如圖三：

圖三 研究思路圖



社交媒體使用中使用者內容生產的同質化

實證分析結果

本節根據上文所論述的資料準備和技術路線，基於 word2vec 的語句（帖子）餘弦相似度測量，考察用戶隨著使用度、活躍度、影響度等的微博使用差異，是否與其內容同質化、個體同質化、範本同質化之間具有相關性，以及具有怎樣的相關性：

關於內容同質化。用戶捲入和使用新浪微博的程度，以及用戶和社交媒體中整體內容的平均相似度，兩者之間的皮爾森、斯皮爾曼相關係數顯著且為正相關。

關於個體同質化。用戶捲入和使用新浪微博的程度，以及用戶與社交媒體其他用戶個體的平均相似度，兩者之間的皮爾森、斯皮爾曼相關係數顯著且為正相關。

關於範本同質化。既然用戶隨著使用新浪微博使用程度的加深而其同質化的程度也越來越深，那麼，不同使用等級的使用者會隨著層級提升出現向最高層級使用者發生內容趨同的情況。而且使用者的使用等級越高，該等級與最高使用等級的使用者之間其平均相似度也越高，兩者之間的皮爾森、斯皮爾曼相關係數顯著且為正相關。

內容同質化：每個層級使用者內容和總體中隨機 20 萬帖子文本的平均相似度

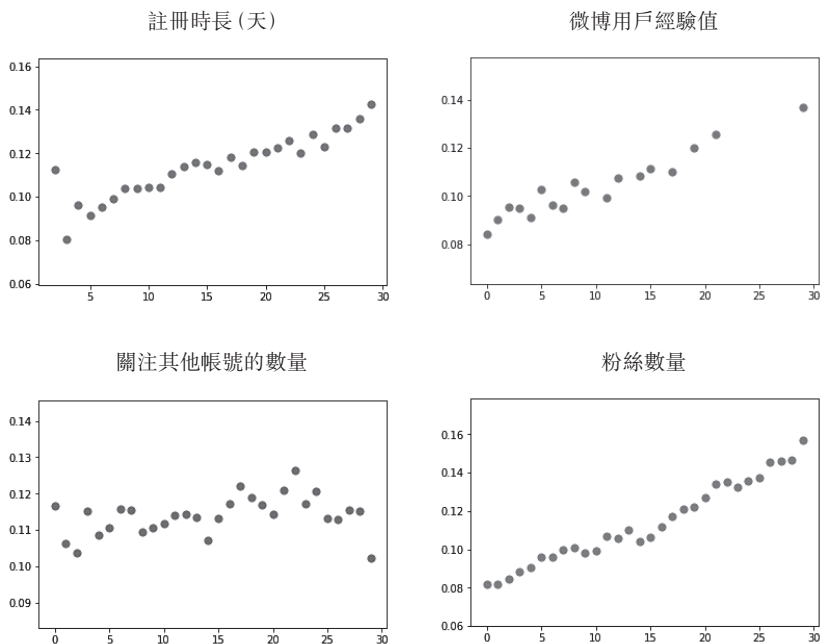
使用者與全域內容的趨於同質化的計算，從 34,892,987 條帖子中隨機抽取 20 萬帖子文本，這種隨機抽取的帖子文本能描述意見氣候分佈，而層級與整體意見氣候的相近也充分能夠說明使用者內容的同質化。

通過把微博使用度的三大方面、四小層面，根據其值大小進行等頻化的「分層」，得到了每層用戶的總體趨同化程度。如圖四，每個子圖描述的一種使用度指標。圖的橫軸是用戶被等頻「分箱」化處理後所處的使用度層級，縱軸是由式 6 也即 $F(G_i, U_p)$ 計算得到的用戶層級與全

《傳播與社會學刊》，(總)第60期(2022)

域20萬帖子的內容相似度。 U_p 由從全體帖子樣本中隨機抽取的20萬條帖子的句向量表示，也即 $200,000 \times 300$ 的矩陣。

圖四 內容同質化檢測——每個層級使用者與整體內容趨同化程度



縱軸的所有數值取值區間0.06~0.16。從橫軸即用戶使用度層級來看，雖然個別指標(關注其他帳號的數量)沒有呈現出充分顯著的線性相關，但大多數指標都表明：用戶使用微博的程度越深(即用戶所在層級越高)，其與總體內容的平均趨同化也越顯著。為了進一步確認這種關聯性的係數，我們進行統計學檢驗。註冊時長、微博用戶經驗值和粉絲數量指標均呈現高度鮮明的正相關($r > .90$; $p < .05$)。這一統計學檢測結果與圖四的發現一致。

用戶在微博中的粉絲量、關注他人數、經驗值等指標，越容易和整體的內容保持平均化的同質化。我們未判斷在何種「絕對」指標上，使用者達到趨於全域內容的高「同質化」臨界水準，但是可以判斷出用戶的「相對」變化趨勢。可以予以統計檢驗的是：無論用戶的同質化程

社交媒體使用中使用者內容生產的同質化

度是高還是低，都在隨著對媒介使用程度的提升，而表現出同質化的相對應的提升，這兩者具有顯著甚至高度的線性正相關。在社交網路中捲入程度較淺的用戶尚存觀點的奇葩性和獨特性，捲入程度較長的用戶則更容易被整體意見氣候影響，甚至由社交網路的整體意見氣候重新塑造和刻劃，成為更加典型的勻質化和標準化用戶。

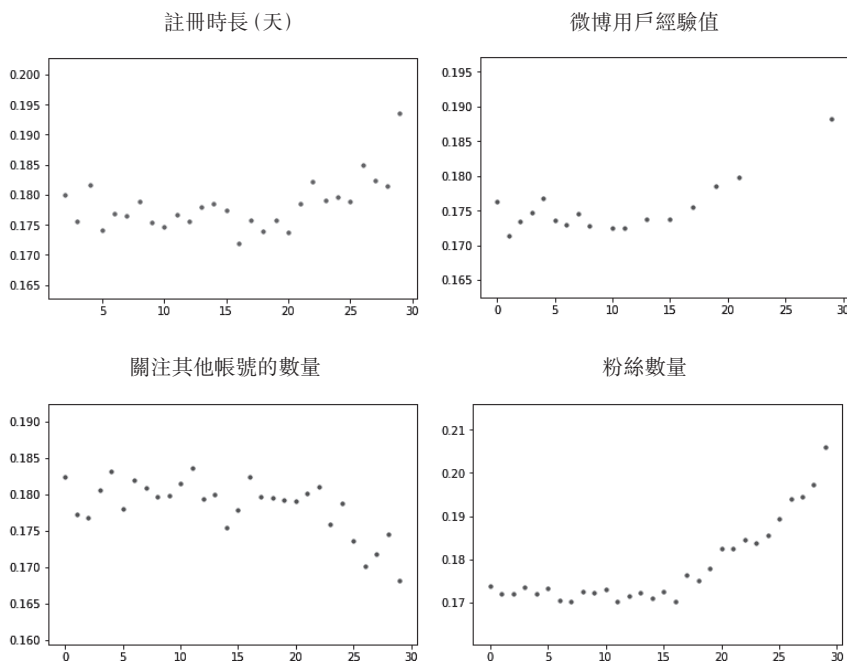
在這個過程中，並非指用戶可能偏向於哪些特殊的內容，資料並無顯示他們是否傾向於哪些特定的主題、類型、領域，而是指他們所生產的內容，變得和全域的帖子具有更高的平均相似度。這體現為一種被全域的「吸附性」(adsorption) 或「話語規訓」(discourse discipline) 產生的去奇葩性和去獨特性，使用者所表達的內容和全域內容越來越勻質性地趨同化，但這並不一定等於和全部內容中的某些特定類型或局部的內容越來越趨同。

個體同質化：每個層級使用者內容和總體使用者內容的相似程度

考察微博中每一層級樣本使用者和全體樣本使用者的內容相似度，是對於用戶同質化和趨同性的有力支持維度之一。該部分的計算方法為式5，即設某一層的用戶為 G_1 ，全體用戶為 G_2 ，則該層用戶與全體用戶的個體同質化程度由 $H(G_1, G_2)$ 計算得到。該式計算的是 G_1 中每位元使用者的帖子內容與 G_2 中每位元使用者的帖子內容，兩兩之間的相似度的均值，其中兩位元使用者之間的帖子內容相似度由式5中的子部分 $R(U_{xi}, U_{xk})$ 也即前文所述的式3計算。實際過程中， G_2 如果採用全部的7,825個使用者則矩陣計算總量過大，因此為了簡化而從中隨機抽取800個用戶代替 G_2 。根據以上計算方法，得到個體同質化方面的結果。

《傳播與社會學刊》，(總)第60期(2022)

圖五 個體同質化檢測——每個層級用戶的個體同質化程度



圖五中，隨著橫軸(即用戶四種使用程度的層級)的增加，個體同質化的程度或升或降，趨勢不一。但是總體看來，個體同質化呈現上升態勢的指標更多，包括註冊時長、微博經驗值和粉絲數量。在相關性分析中，這種態勢表現得更加明顯。註冊時長、自帶經驗值和粉絲數量與個體同質化的關係更加顯著($r > .40$; $p < .05$)。

我們發現，關注其他帳號的數量這一指標似乎總是「叛逆」的。這一指標要麼呈現出統計學上微弱的解釋效力，要麼比如在本次檢驗中，它呈現出與其他指標不同的態勢($r = -.62$; $p < .05$)。那麼該指標與其他指標有何種不同呢？我們認為，與粉絲數量偏向用戶的「傳播者」屬性不同，關注他人數量更強調使用者的「接收者」屬性。作為受眾的使用者，即便在社交網路中深層捲入，也沒有形成明顯的同質化現象。甚至，關注他人數量最多的用戶，他們的個體同質化程度相對來說是最弱的。然而，當用戶作為「傳播者」的時候，隨著影響力的增加，用戶卻呈現出較為明顯的同質化現象。這種用戶角色的轉換與其中蘊含著的張力，都值得我們對微博和社交媒體產生更多思考。

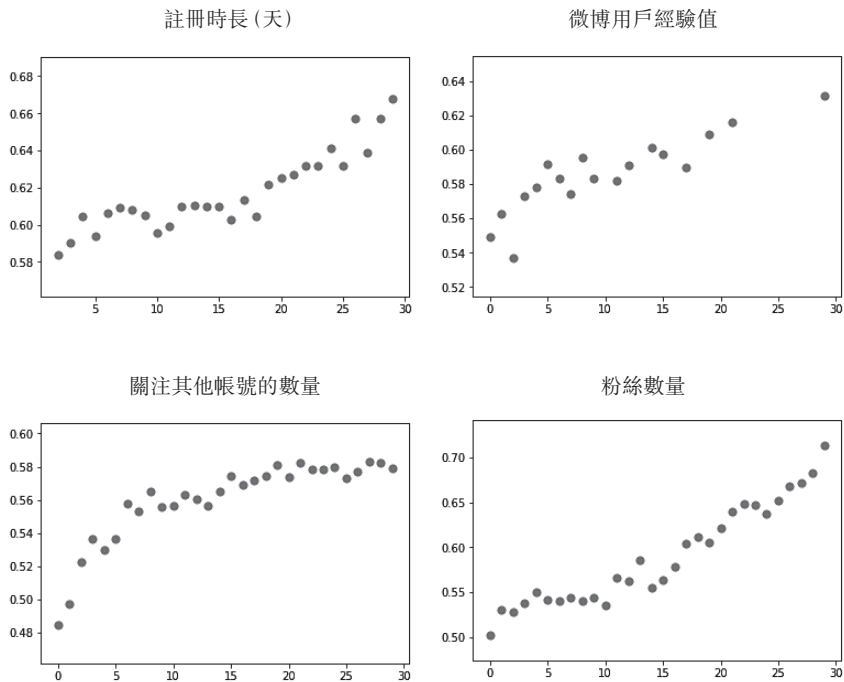
社交媒體使用中使用者內容生產的同質化

範本同質化：每個層級使用者內容和最高層級使用者的相似程度

範本同質化的計算方法與個體同質化的計算方法式5一致。設某一層的用戶為 G_1 ，最高層級的頂部用戶為 G_2 ，則該層用戶與頂部層級的同質化程度可由 $H(G_1, G_2)$ 計算得到。

這種計算方法與個體同質化不同的是， G_2 不是整體的全部使用者，而是只選擇每一指標下的最高等級的使用者，例如最高粉絲數的用戶層，或最高經驗值的用戶層，藉以探索普通使用者和最高等級使用者之間的關係，以及用戶在使用程度增長中的演變規律。範本同質化可以看作層級內部的傳遞機制，這種「傳遞」可以通過每一層級與最高層級的比較，並逐漸增加趨同化來完成。每個層級用戶的範本同質化程度如圖六。

圖六 範本同質化檢測——每個層級用戶的範本同質化程度



《傳播與社會學刊》，(總)第60期(2022)

圖六的結果鮮明支持了這種趨向範本的「同質化」現象。所有指標，包括上文所述的有些「叛逆」的關注他人的帳號數量指標，全部呈現出隨著橫軸指標的增加而較為明顯的線性上升的態勢。說明範本同質化與用戶使用程度的確存在明顯和高度的關聯。越是高等級的使用者，趨向於最高範本的相似度就越高，這個相似度程度明顯超過了高等級使用者和全域使用者的相似度。統計學分析中，所有相關係數均在0.80以上，且通過了顯著性檢驗。這一結果說明，社交網路工業的「流水線」中，使用程度越深的用戶更容易受到使用度頂層群體的影響。這可以在一定程度上說明我們理解網路社群中趨同傾向的產生原因。「意見領袖」(opinion leader)到達範圍更廣，粉絲受眾擁躉者更多，其影響力和話語聲量自然擴大。另一方面要想獲得更多話語權力，就應該多向「範本」看齊，學習他們的話語方式，這在一定程度上與網路時代的「流水線」式生產暗合。

使用者屬性對於三種同質化類型的解釋程度

發帖數量表示用戶在微博空間中的話語聲量，這一變數天然地和捲入度、活躍度、影響力產生關聯。發帖數量高的用戶，捲入度、活躍度和影響力都比較高。因此發帖數量可能隱含在使用者屬性和同質化水準的關係之中。當我們剔除發帖數量這一變數的影響，再去分析使用者的屬性和同質化現象的影響，可以捕捉因變數對於自變量的淨作用。

表三說明，即便控制發帖量的影響之後($N = 7,825$)，用戶的活躍度、捲入度、影響力等指標依然能夠對於同質化水準產生影響。控制使用者發帖量的影響之後，內容同質化和註冊時長、微博用戶經驗值、粉絲數量均存在正向關聯，且具有統計學意義。粉絲數量對於範本同質化的影響較為顯著。從整體來看，在控制變數的情況下，使用者捲入度、活躍度、影響力越高，同質化水準就越高這一結論是能夠成立的。

社交媒體使用中使用者內容生產的同質化

表三 控制變數下的使用者屬性與同質化現象的偏相關分析

控制變數	內容同質化	個體同質化	範本同質化
發帖數量			
註冊時長(天)	.24**	.21**	.09**
微博用戶經驗值	.22**	.18**	.03**
關注其他帳號的數量	.04**	.07**	-.03**
粉絲數量	.37**	.13**	.33**

註：** $p < .01$

結論

本研究試圖考察：微博中的使用者生產內容是否趨於和總體平均的「同質化」，以及在此中表現出「主體性的消逝」(the disappearance of subjectivity)及其在社交網路中的標準化再生產。在進行了內容同質化、個體同質化和範本同質化的分析和檢驗後發現，這種全域、平均的同質化現象在上述分析中的確存在，而且與用戶的使用層級有著密切的關聯。特別是，用戶和最高層級的範本用戶的同質化會隨著層級的增加形成明顯的「看齊」(emulate)和「模仿」(imitate)，甚或產生內容上的「使用者複製」(user duplication)和「用戶模因」(user meme)擴散。在對於用戶使用微博的指標進行量化的過程中我們發現，註冊時長、微博經驗值、粉絲量這些指標基本上包含了微博用戶作為一個傳受者，接受與發佈資訊、在平台內進行多種互動的基本使用習慣。這些指標與同質化的關聯較為明顯。但是當使用者作為受眾關注資訊和他人帳號的時候，同質化現象似乎並沒隨層級增加而變得嚴重。

總體而言：

(1) 用戶隨著微博使用程度的加深，其所生產、發佈、表達的內容，與微博其他使用者所生產的內容，在總體上的平均相似度會越高。用戶的所言所表，不是像較多的理論研究與實證分析認為的那樣變得越來越多樣、差異化，也不是加大各種「亞文化」間的溝壑和「參與文化」的「狂歡」，而是與總體平均內容越來越相似和趨同。這種內容趨同化，在使用度、活躍度的指標都表現得較為鮮明，甚至是高度的內容同質化。長期以來，人人都能發表意見的微博被視為象徵民主的新場域，而從內容相似性的分析上看，微博的意見氣候分佈是趨於同一的。這種趨同的主流化甚至在多數時候並未被宣稱自主或個性的個體

《傳播與社會學刊》，(總)第60期(2022)

所意識到，用戶在看似注重個性和自主性、草根參與的環境下，恰恰在求異之中卻馳往存同的趨向。

(2) 用戶的個體同質化在捲入度和活躍度方面都是顯著相關。說明用戶作為獨立個體，從其表達、發佈的視角考察相似度同樣存在同質化現象。作為獨立個體，使用者特徵還可以從其發佈內容的特徵解構使用者在社交網路中扮演的角色，這種從內容到使用者的分析路徑更客觀地描述了使用者在社交網路海洋中所扮演的角色。用戶的主體性被宏觀的同質性和平均化張力漸漸拉扯和吞沒，在日復一日的網路實踐中被社交媒體流水線上的「機器」(machine)加工成一個「標準化用戶」(standardized user)。

(3) 用戶的範本同質化展示出了與使用者的社交網路使用之間高度強烈的正相關。這讓我們不禁追問微博空間中趨向最高使用層級的用戶並建立聯繫是否存在某種「範本效應」(template effect)? 即越模仿最高等級使用者的發佈內容越能夠獲得更多影響力。這種「範本效應」不僅與傳播效率的提升有關，更使我們對於微博空間的資訊互動產生了更多思考。本研究邏輯並非僅關注最高使用層級的深度用戶，但內容同質化和個體同質化的顯著，都會導向作為一個「社會標準化」(social standardization)「仲介」(intermediary)的「典範階層」(template class)。重要的不是誰能成為這樣的階層，而是一種邏輯應然性，即均質趨同的標準化媒體空間，勢必導向這樣的「意見典範」(opinion template)。在微博這樣的社交平台中，話語權的中心往往主宰著意見氣候的走向，成為一種「範本」、「標準」。一個用戶在微博中「浸泡」的時間越長，越容易向標準看。社交網路捲入程度的深入，使其受教化的程度加深。主體性被限於既定的框架內。借用傅柯(Michel Foucault)(2003)的概念，展現出好似被某種「規訓」(discipline)或「規則」(regulation)培養後趨於標準化的景象。

第三子問題中的「範本同質化」，開啟的思考是從「頭部用戶」到「頂部用戶」的延拓問題。在微博傳播研究中，具有高影響力的意見領袖、「大V」，作為「頭部用戶」吸引著學界和業界關注的焦點。然而在此之外，有一個同樣可能重要的維度被忽視了，那就是雖然未必是高流

社交媒體使用中使用者內容生產的同質化

量、高影響力，但是卻有高使用度的那部分「頂部使用者」。本研究鮮明地顯示，這些頂部使用者具有強大的作用效應，隨著使用度的提升，用戶們朝向這種使用度「頂部」的用戶趨近、相似的現象穩定而明顯。我們現有的材料暫無法斷定：是從底部用戶到頂部用戶都在某種更超越、更高的共同方向趨近，還只是較為淺顯地——底部用戶不斷地朝頂部用戶趨近和蛻變。但是至少可以斷定的是，頂部使用者的內容擁有不亞於「頭部用戶」和意見領袖的在微博生態圈中的節點效能，起著其作為典範的作用。這是在內容的層面，但不是意見認知的擴散，而接近於用戶的複製和再生產。這種現象指向從「意見領袖」到「意見典範」的研究進路。

本文分析的使用者內容生產的同質化，是受到總體內容、全域使用者的內容吸附或「拉扯」而下沉於總體的均勻同質性。這種同質化是一種全域性的同質化，是朝向全域內容的平均同質化。局部同質化和局部相似性可能帶來越來越深的網路「巴爾幹化」及其各種社會壁壘。而本研究分析的全域、拉勻後的同質化則是帶來網路空間整體的內容表達趨同性。在一定意義上，這種同質化不是生成網路中的「巴爾幹化」，而是使網路本身整體「巴爾幹化」。這一點，本研究也為網路巴爾幹化與同質化之間的關係提供了新的思考。

對媒介技術的批判，伴隨著大眾傳媒的發展一直備受學者關注。本研究通過計算傳播學實證的方法，驗證了批判學派的觀點，探討了媒介技術的使用中對人、主體的異化作用。法蘭克福學派以及其他批判學派對於文化工業的「單向度」和同一性、文化和傳媒「標準化」等的研究，主要基於整體和宏觀理論，和注重局部的實證研究時有相悖之處。而新的計算傳播學、文本挖掘的多種研究手段，使得對於媒介內容的研究更好地建立在對於內容基本的微觀單位的分析基礎上，搭建了通往宏觀的橋樑。本研究對社交網路同質化的研究也更強調同質化作用在更為全域的發生，給同領域的其他研究提供了更為宏觀的視角。在實際的檢測中，我們發現的社交網路同質化效應隨著使用程度提高而逐漸加深，這一點也為社交網路同質化的形成機制和演化提供了一些參考。

《傳播與社會學刊》，(總)第60期(2022)

本研究還存在一些方面的不足之處。為了充分計算使用者的內容、擴大所選取帖子的覆蓋度、增強計算結果的健壯性，我們對每個用戶選取了3,000條帖子。雖然最終所選取的7,825個樣本使用者中，其粉絲數、微博經驗值、關注數等多種指標，在分佈上都廣泛涉及到各種低、中、高的多樣化用戶，可以一定程度上減弱這種推廣結論的約束。但是這個限定仍然可使得本研究的結論在無可靠證據的情況下，適用範圍暫不能有效推廣到發帖活躍度在此閾值以下的用戶群，這是今後有待補充和加強的方面。此外，本文對word2vec模型的使用是基於上下文語境關係所蘊含的語義特徵的計算，會在一定程度上反映在語境中所蘊含的情感、態度、意向等成分，但並沒有專門、集中地對這些語義之外的其他要素進行分析。同時，本文也沒有探討除了社交媒體使用度之外的其他線下因素的影響，例如用戶的現實身份、地位、社會關係和社會文化屬性等等。這些都有待今後的繼續挖掘。

目前對於互聯網等新興媒介技術的研究日益增多，互聯網的普及賦予人們更多能動性和未來的可能性，而互聯共通的網路滲透至社會各個角落卻帶來了諸如資訊異化和同質化的擔憂。本研究雖以新浪微博為樣本，但這個問題在當今的社交化和新媒體時代，並非僅限於微博。

參考文獻

中文部分 (Chinese Section)

- 呂學強、任飛亮、黃志丹、姚天順(2003)。〈句子相似模型和最相似句子查找演算法〉。《東北大學學報》，第6期，頁531-534。
- Lü Xueqiang, Ren Feiliang, Huang Zhidan, Yao Tianshun (2003). Juzi xiangsi moxing he zui xiangsi juzi chazhao yansuan fa. *Dongbei daxue xuebao*, 6, 531-534.
- 李夢潔、邵曦(2018)。〈基於文字屬性的微博用戶相似度研究〉。《電腦技術與發展》，第5期，頁23-28。
- Li Mengjie, Shao Xi (2018). Jiyu wenzi shuxing de weibo yonghu xiangsidu yanjiu. *Diannao jishu yu fazhan*, 5, 23-28.
- 馬克·波斯特(2000)。《第二媒介時代》(范靜曄譯)。南京：南京大學出版社。(原書Poster, M. [1995]. *The second media age*. Oxford: Polity Press.)

社交媒體使用中使用者內容生產的同質化

- Make Bosite (2000). *Dier meijie shidai* (Fan Jinghua, Trans.). Nanjing: Nanjing daxue chubanshe. (Original book: Poster, M. [1995]. *The second media age*. Oxford: Polity Press.)
- 馬克斯·霍克海默、希歐多爾·阿多諾 (2006)。《啟蒙辯證法》(渠敬東、曹衛東譯)。上海：上海人民出版社。(原書Horkheimer, M., & Adorno, T. W. [1947]. *Dialektik der Aufklärung: Philosophische fragmente*. Vilnius: Margi rađtai.)
- Makesi Huokehaimo, Xiouduoer Aduonuo (2006). *Qimeng bianzheng fa* (Qu Jingdong & Cao Weidong, Trans.). Shanghai: Shanghai renmin chubanshe. (Original book: Horkheimer, M., & Adorno, T. W. [1947]. *Dialektik der Aufklärung: Philosophische fragmente*. Vilnius: Margi rađtai.)
- 陳然 (2012)。〈網路論壇活躍群體社交網路研究——從「關注」行為的視角〉。《新聞界》，第18期，頁51-55。
- Chen Ran (2012). Wanglu luntan huoyue qunti shejiao wanglu yanjiu—Cong “guan Zhu” xingwei de shijiao. *Xinwen jie*, 18, 51-55.
- 張一璐、倪靜、郭強、劉建國 (2014)。〈線上用戶興趣多樣性的實證研究〉。《電腦應用研究》，第11期，頁3250-3252。
- Zhang Yilu, Ni Jing, Guo Qiang, Liu Jianguo (2014). Xianshang yonghu xingqu duoyang xing de shizheng yanjiu. *Diannao yingyong yanjiu*, 11, 3250-3252.
- 曾一果 (2016)。〈批判理論、文化工業與媒體發展——從法蘭克福學派到今日批判理論〉。《新聞與傳播研究》，第1期，頁26-40。
- Zeng Yiguo (2016). Pipan lilun, wenhua gongye yu meiti fazhan—Cong Falankefu xuepai dao jinri pipan lilun. *Xinwen yu chuanbo yanjiu*, 1, 26-40.
- 米歇爾·傅柯 (2003)。《規訓與懲罰：監獄的誕生》(劉北成、楊遠嬰譯)。北京：生活·讀書·新知三聯書店。(原書Foucault, M. [1975]. *Surveiller et punir: Naissance de la prison*. Paris: Gallimard.)
- Mixieer Fuke (2003). *Guixun yu chengfa: Jianyu de dansheng* (Liu Beicheng, Yang Yuanying, Trans.). Beijing: Shenghuo dushu xinzhishi sanlian shudian. (Original book: Foucault, M. [1975]. *Surveiller et punir: Naissance de la prison*. Paris: Gallimard.)
- 赫伯特·馬爾庫塞 (2008)。《單向度的人：發達工業社會意識形態研究》(劉繼譯)。上海：上海譯文出版社。(原書Marcuse, H. [2002]. *One-dimensional man: Studies in the ideology of advanced industrial society*. Boston: Beacon Press.)
- Hebote Maerkusai (2008). *Dan xiangdu de ren: Fada gongye shehui yishi xingtai yanjiu* (Liu Ji, Trans.). Shanghai: Shanghai yiwenshu chubanshe. (Original book:

《傳播與社會學刊》，(總)第60期(2022)

- Marcuse, H. [2002]. *One-dimensional man: Studies in the ideology of advanced industrial society*. Boston: Beacon Press.)
- 鄭志蘊、賈春園、王振飛、李鈍(2017)。〈基於微博的用戶相似度計算研究〉。《電腦科學》，第2期，頁262–266。
- Zheng Zhiyun, Jia Chunyuan, Wang Zhenfei, Li Dun (2017). Jiyu weibo de yonghu xiangsidu jisuan yanjiu. *Diannao kexue*, 2, 262–266.
- 劉錦德、劉詠梅(2013)。〈基於不完全資訊演化博弈模型的網路輿情傳播羊群行為〉。《國防科技大學學報》，第5期，頁96–101。
- Liu Jinde, Liu Yongmei (2013). Jiyu bu wanquan zixun yanhua boyi moxing de wanglu yuqing chuanbo yangqun xingwei. *Guofang keji daxue xuebao*, 5, 96–101.
- 劉亞洲、潘曉中、付偉(2018)。〈基於節點從眾博弈的社交網路謠言傳播模型〉。《電腦工程》，第44期，頁303–308。
- Liu Yazhou, Pan Xiaozhong, Fu Wei (2018). Jiyu jiedian congzhong boyi de shejiao wanglu yaoyan chuanbo moxing. *Diannao gongcheng*, 44, 303–308.
- 藍雁玲、陳建超(2011)。〈基於詞性及詞性依存的句子結構相似度計算〉。《電腦工程》，第10期，頁47–49。
- Lan Yanling, Chen Jianchao (2011). Jiyu cixing ji cixing yicun de juzi jiegou xiangsidu jisuan. *Diannao gongcheng*, 10, 47–49.
- 魏靜、朱恒民、宋瑞曉、蔣世兵(2016)。〈個體視角下的網路輿情傳遞鏈路預測分析〉。《現代圖書情報技術》，第1期，頁55–64。
- Wei Jing, Zhu Hengmin, Song Ruixiao, Jiang Shibing (2016). Geti shijiao xia de wanglu yuqing chuandi lianlu yuce fenxi. *Xiandai tushu qingbao jishu*, 1, 55–64.

英文部分 (English Section)

- Boutet, A., Kim, H., & Yoneki, E. (2012). What's in Twitter: I know what parties are popular and who you are supporting now! *Social Network Analysis and Mining*, 3(4), 1379–1391.
- Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2011, July). *Political polarization on Twitter*. Paper presented at Fifth International AAAI Conference on Weblogs and Social Media, Barcelona.
- De Choudhury, M., Sundaram, H., John, A., Seligmann, D. D., & Kelliher, A. (2010). “Birds of a feather”: Does user homophily impact information diffusion in social media? *Modern Rheumatology*, 17(4), 279–282.
- Doganoglu, T., & Wright, J. (2006). Multihoming and compatibility. *International Journal of Industrial Organization*, 24(1), 45–67.

社交媒體使用中使用者內容生產的同質化

- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320.
- Gross, B. (2015). What makes someone a cyber-Balkan? Finding the linkages between social psychology and self-selectivity in U.S. politics online. *Atlantic Journal of Communication*, 23(4), 225–236.
- Helberger, N., Karppinen, K., & D’Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information Communication & Society*, 21(2), 1–17.
- Kobayashi, T., & Ikeda, K. (2009). Selective exposure in political web browsing. *Information Communication & Society*, 12(6), 929–953.
- Lee, A., & Lee, E. B. (1939). *The fine art of propaganda*. New York: Harcourt, Brace & World, Inc.
- Mcperson, M., & Cook, S. L. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a, May). *Efficient estimation of word representations in vector space*. Paper presented at Proceedings of Workshop at ICLR, Arizona Country.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b, December). *Distributed representations of words and phrases and their compositionality*. Paper presented at Proceedings of Workshop at NIPS, Lake Tahoe.
- Negroponte, N. P. (1995). *Being digital*. New York: Random House Inc.
- Parikh, A., Tckstrm, O., Das, D., & Uszkoreit, J. (2016, November). *A decomposable attention model for natural language inference*. Paper presented at Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin.
- Pons-De Wit, A., Versteeg, P., & Roeland, J. (2015). Contextual responses to interreligious encounters online. *Social Compass*, 62(1), 89–104.
- Shen, D., Wang, G., Wang, W., Min, M. R., Su, Q., Zhang, Y., Li, C., Henao, R., & Carin, L. (2018, July). *Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms*. Paper presented at Proceedings of Workshop at ACL 2018, Melbourne.
- Van Alstyne, M., & Brynjolfsson, E. (1996). Could the internet balkanize science? *Science*, 274(5292), 1479–1480.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, December). *Attention is all you need*. Paper presented at Proceedings of Workshop at NIPS, Long Beach.
- Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010, February). *Twitterrank: Finding topic-sensitive influential Twitterers*. Paper presented at Proceedings of the Third International Conference on Web Search and Web Data Mining, New York.

《傳播與社會學刊》，(總)第60期(2022)

本文引用格式

徐翔、周詩凡、靳菁(2022)。〈社交媒體使用中使用者內容生產的同質化——基於新浪微博樣本的使用者相似度分析〉。《傳播與社會學刊》，第60期，頁123-156。