

學術對談

對談計算傳播學：起源、理論、 方法與研究問題

對談人：祝建華、黃煜、張昕之



祝建華教授
(Prof. Jonathan J. H. Zhu)

「計算傳播學的出現是歷史的必然，動因大致包括三方面：一是現象、二是數據、三是方法。具體來說：互聯網的興起催生了很多前所未有的傳播行為。互聯網誕生後，人們的線下 (offline) 生活和網上 (online) 生活是此消彼長的；如今，網上生活已經超過了線下生活。這種現象帶來了新的研究數據。網上行為留下了數碼蹤跡 (digital trace)，海量數據變得唾手可得。因為有了這些海量數據，

祝建華，香港城市大學媒體與傳播系講座教授。研究興趣：計算社會科學在傳播研究中的應用。電郵：j.zhu@cityu.edu.hk

黃煜，香港浸會大學傳理學院教授。研究興趣：新聞與媒體表現、傳播政治經濟學。電郵：s03033@hkbu.edu.hk

張昕之，香港浸會大學新聞系研究助理教授。研究興趣：比較政治傳播、新媒體與社會變遷。電郵：xzhang2@hkbu.edu.hk

《傳播與社會學刊》，(總)第44期(2018)

傳統的數據收集和分析方法(比如調查、實驗和內容分析)效率不高、精度不夠，所以需要引入新的研究方法。因此，一批學者開始從事計算方法的開發和推廣。計算方法的普及，使得研究網上行為如虎添翼。總之，『現象、數據、方法』，三方面的條件在最近十幾年內相繼逐步成熟，最終導致計算傳播學的興起。」

Dialogue

Dialogue on Computational Communication Research: Origins, Theories, Methods, and Research Questions

Discussants: Jonathan J. H. ZHU, Yu HUANG, Xinzhi ZHANG

Abstract

Computational communication research is an emerging research approach. It aims to describe, explain, and predict human's communication behaviors and their driving mechanism with the availability of online behavioral data (e.g., server log analytics, web scraping, online archiving/indexing data, and mass online experiment) and computational methods (e.g., social network analysis, text mining, spatial analysis, and temporal analysis). Jonathan Zhu is a leading scholar in computational communication research. In the present dialogue, he described the conceptual, methodological, and technological profiles of computational communication research, through his first-hand observations. He also reflected his personal experience in pursuing research in this interdisciplinary area as a trained social scientist and a bridge with computing and data science, and pinpointed the opportunities, challenges, and pitfalls in this research approach in the future.

Citation of this article: Zhu, J. J. H., Huang, Y., & Zhang, X. (2018). Dialogue on computational communication research: Origins, theories, methods, and research questions. *Communication & Society*, 44, 1–24.

Jonathan J. H. ZHU (Chair Professor). Department of Media and Communication, City University of Hong Kong. Research interests: computational social science for communication research.

Yu HUANG (Professor). School of Communication, Hong Kong Baptist University. Research interests: journalism and media performance, political economy in communication.

Xinzhi ZHANG (Research Assistant Professor). Department of Journalism, Hong Kong Baptist University. Research interests: comparative political communication, new media and social change.

《傳播與社會學刊》，(總)第44期(2018)

祝建華教授簡介

香港城市大學計算社會科學講座教授。他先後獲得中國復旦大學學士、碩士和美國印第安納大學博士。曾任美國康乃狄克大學副教授和中國人民大學長江學者講座教授。現為香港城市大學媒體與傳播系講座教授、傳播研究中心主任，以及互聯網挖掘實驗室創辦人。近年來的研究興趣包括互聯網、社交網、移動網的使用及效果等。他目前兼任 *Journal of Computer-Mediated Communication* 編委、中國計算機學會大數據專家委員會委員等職，並在傳播學、計算機、物理、醫療信息等學科刊物上發表其有關網絡抽樣、網絡可控性、互聯網信息擴散等研究結果。他是國際傳播協會 (ICA) 計算方法小組 (Computational Methods) 的首倡發起人，他最近指導實驗室學生團隊在中國計算機學會 2017 年大數據和計算智能大賽中獲得三項大獎。

JZ：祝建華

HY：黃煜

XZ：張昕之

HY：請您談一談計算傳播學的起源、主要研究對象和方法的特點。

JZ：計算傳播學 (computational communication research) 不是一門新的學科或新的理論，而是一種新的研究取向 (research approach) 或一套新的研究方法，可以大致定義為：通過收集和分析網上行為數據 (online behavioral data)，描述、解釋和預測人類傳播行為及其背後驅動機制的一系列計算方法 (computational methods)。其主要數據來源不再是傳統的內容分析、問卷調查和控制實驗，而是來自於服務器日誌數據挖掘 (server log analytics)、網頁數據抓取 (web scraping)、在線檔案數據 (online archiving/indexing data)、大規模在線實驗 (mass online experiment) 等；其數據分析方法除了經典的統計分析之外，還包括社會網絡分析 (social network analysis)、文本挖掘 (text mining)、空間分析 (spatial analysis)、時間分析 (temporal analysis) 等。為了幫助理解，我經常用以下這個表來界定計算傳播學的邊界。

對談計算傳播學：起源、理論、方法與研究問題

表一 什麼是計算傳播學、什麼不是？

研究方法	研究場景	
	線下內容、使用及效果	網上內容、使用及效果
傳統方法 (如定量方法)	1. 傳統傳播學研究	2. 傳統傳播學研究
計算方法	3. 計算傳播學研究	4. 計算傳播學研究

表一中的第1和第4格非常清楚，不需多說。第2格(如社會化媒體研究中常見的通過問卷調查研究大學生使用Facebook)為什麼不是計算傳播學研究？因為這類研究與使用問卷調查研究讀者閱讀印刷版《紐約時報》的行為在本質上沒有區別，只是研究場景不同而已。反之，第3格又為什麼算作計算傳播學呢？我這裡專指採用計算方法來收集和分析線下傳播行為，但這些行為記錄一定要被數碼化了，不然無法用上計算方法。如Google nGrams把幾百萬本書掃描成文本數據，在此基礎上的分析研究就是計算傳播學了(Michel et al., 2011)。當然，這一格內的案例不多，很值得計算傳播學者去關注和發掘的。以下的對話中提到的計算傳播學，主要是指第4格。

計算傳播學的出現是歷史的必然(儘管這過程中也有一些偶然的因素，後面會提到)。必然的因素有三方面：一是現象、二是數據、三是方法。具體來說：互聯網的興起催生了很多前所未有的傳播行為。互聯網誕生後，人們的線下(offline)生活和網上(online)生活是此消彼長的，今天，網上生活已經超過了線下生活。同時，這種現象帶來了新的研究數據。網上行為可以留下數碼蹤跡(digital trace)，比如網站登陸、在線消費等，海量數據變得唾手可得。因為有了這些海量數據，傳統的數據收集和分析方法(比如調查、實驗和內容分析)效率不高、精度不夠，所以需要引入新的研究方法。因此，一批學者開始從事計算方法的開發和推廣。計算方法的普及，使得研究網上行為的學者如虎添翼。總之，「現象、數據、方法」，三方面的條件在最近十幾年內相繼逐步成熟，最終導致計算傳播學的興起。

當然，計算傳播學發展速度之快，遠遠超出人們的預料。2012年夏天，我參加了第465次北京香山會議，專門討論是否需

《傳播與社會學刊》，(總)第44期(2018)

要建立數據科學(當時還不流行大數據的說法)。香山會議是小規模的閉門「神仙會」，屬於非正式會議，但討論結果對高層決策往往有較大影響。那次會議的一個具體成果就是在中國計算機學會內設立大數據專委會，從此揭開大數據在中國的推廣、普及(甚至炒作)之浪潮。可以說，我從一開始就身在源頭，但想不到浪潮這麼快就沖到了傳播學。

HY：這個「現象－數據－方法」的過程，有先後出現的特定順序嗎？

JZ：這個過程在邏輯上應該是先後發生的，但在事實上是平行出現的。早期，新的傳播現象並沒有引起社會科學研究主流的足夠重視。其實，用計算方法對人類媒體的使用行為進行記錄和分析的工作很早就開始了，只不過當時的研究者不以「計算社會科學」或「計算傳播學」作為名稱而已。比如，早在互聯網之前的1970年代，就開始了對電視收視的在線行為的研究，稱為電視收視分析(TV ratings analysis，見Webster & Lichty, 1991)，可惜這些方法及其成果並沒有得到學術界的足夠重視。我早年也研究過電視收視的個體行為，以後有了第一台互聯網服務器後，就很自然地沿用收視率的概念和方法來分析用戶網上行為，所以比較瞭解計算傳播學的來龍去脈。

如果一定要從時間上來劃分的話，計算傳播學的發展大致上有三個階段。第一階段，針對用戶使用行為的研究，從電視收視分析的時候開始就有了。第二階段，應該是到2005年前後，隨著社會化媒體平台的出現和用戶自創內容(user-generated content，簡稱UGC)的出現，數據來源從後台(服務器)變到了前台(網頁)。在第一階段時，很多數據是依賴後台提供的；而到了第二階段，可以直接從前台抓取，學界和業界對UGC這類研究開始感興趣。第三階段大概在2010年前後開始，越來越多的學者開始採用不同於傳統研究方法的一系列計算方法，計算傳播學這個研究取向獲得了進一步的發展。這其中，大量可得數據的出現，是貫穿始終的一個驅動因素。

HY：「計算傳播學」成為一個約定俗成的名詞是什麼時候開始的？有沒有標誌性的事件？

對談計算傳播學：起源、理論、方法與研究問題

JZ：我並不主張把「計算傳播學」叫做一個學科或一個獨立的研究領域，而更傾向將其稱為「計算社會科學在新聞傳播研究中的應用」。我和幾位學生在2014年發表了一篇文章，就是用這一說法做標題的（祝建華等，2014），文中按照傳播學「5個W」（Who [誰]、says what [說什麼]、in what channel [通過何種渠道]、to whom [對誰說]、with what effect [產生何種效果]）的順序，逐一介紹應用計算方法研究傳播學的經典案例。這篇文章，也許算是中文文獻中的開始介紹計算傳播學的一個標誌性事件。而說到「計算傳播學」這個名稱，應該是我的博士生王成軍首先提出的。我個人比較保守，最初並不主張用這個名稱，因為計算傳播學研究的依然是5W的「老問題」。這裡當然還有一個不同語言之間的翻譯問題。英文比較準確，叫computational communication research，但譯成中文時加了一個「學」字，似乎就成為一個專門的學科了。現在約定俗成了，大家都用計算傳播學這個名稱，我也只能從眾，但每次都要強調這是一種新的研究取向或一組新的研究方法，而不是一門新的學科。也許有些年輕後輩對此並不滿足，認為新取向、新方法背後有其新理論、並且已經有了新發現，可以叫新學科了。我當然也希望這個取向有朝一日能夠孵化、催生出一門新的學科，但目前才剛剛起步，來日方長。

通俗地說，計算傳播學的主要研究對象是「社會化媒體」（social media）。目前很多研究學者並不是傳播學出身，而是來自物理、計算機、工程等學科的學者。這些人一直在使用計算方法。從最早的數據挖掘（data mining），到機器學習（machine learning），到現在的深度學習（deep learning），計算方法至少有這麼三個發展階段，（與上面講的計算傳播學發展三階段並不一一對應）。而這三個階段的發生和演變，有必然、也有偶然的因素，如深度學習的興起、成為計算方法的最新階段，就是偶然的。這是題外話，以後有機會可以專門講講這個過程對社會科學基本理念的深刻啟示。總之，任何技術的興起，都和滿足某種社會需求有關。

在國際傳播學術圈內，一個標誌性的事件則應該是2014、2015年之間，我和彭泰權、梁海等人在國際傳播學會（Inter-

《傳播與社會學刊》，(總)第44期(2018)

national Communication Association, ICA) 內倡議建立一個計算方法 (Computational Methods) 的興趣小組。作為首倡者和發起人，我草擬了小組的申請報告、對計算傳播學做了前面提到的定義和說明。我們徵集了歐美亞太等地50多ICA會員作為發起人，兩年多來已經增加到200多位成員了。首屆小組主席由彭泰權擔任。這段歷史也說明，中文語境下的計算傳播學與國際範圍的計算傳播學之間的發展是同步掛鉤、互相推動的，這種情形在傳播學其它各分支中還是罕見的。

我們起初想把這個小組命名為「computational communication research」(簡稱CCR)，但是鑑於ICA已經有若干涉及計算研究的分會或興趣小組，比如傳播與技術 (Communication and Technology, CAT) 分會和移動傳播 (Mobile Communication) 興趣小組。為了避免名稱上的重合，我提出兩個方案，一個叫 social media analytics，另一個就是 computational methods。最後經過投票，大家在這兩個名字中選取了後者，因為前者顯得太應用，而後者更加學術。所以說，我們一開始就把計算傳播學定位為研究方法，希望通過在ICA中建立一個專門小組來為傳播學研究方法的創新作出貢獻。如果說任何一個研究都涉及 what (做什麼) 和 how (怎樣做) 這兩部分的話，我們更希望為「how」這個問題提供一個討論的場所。推廣開去，任何一個成熟的學科都有一個專門討論方法論的刊物。最近聽說德國的一些學者在推動ICA成立一個傳播研究方法小組，專注經典方法的討論。從時間上看，ICA先有新的計算方法小組、再有經典方法小組，則是一件與眾不同、值得回味的趣事。

HY: 那麼在您看來，計算傳播學與傳統的量化研究方法或者傳播理論分析有什麼重要的不同之處？背後的研究理念是什麼？

JZ: 相同之處在於都是研究人類傳播的現象和行為，但是由於不同的記錄方法，產生了不同的數據，從而導致了一系列的區別。

不同之處主要有三：

首先，是數據的結構不同。傳統方法(專指定量方法，下同)用的多是「矮胖型」(flat data) 數據，觀測的個案 (cases) 數目少，

對談計算傳播學：起源、理論、方法與研究問題

但是變量 (variables) 很多，有的社會調查可以問到成千上百個問題。統計學裡有一個分支專門討論當 k (變量數) 大於 n (個案數) 的分析方法 (比如你只有 200 個個案，但是每個個案回答 1,000 多個問題，這違反了經典統計分析中 $k < n$ 的原則)。而大數據時代，我們不再需要為此煩惱，因為數據都是「高瘦型」(tall data)，尤其是網上獲取的行為數據，個案很多 (可以達到數億甚至數千億)，但是變量很少 (通常只有三五個變量，如用戶在什麼時間、什麼地點、做了什麼事情這幾項而已)。

其次，變量的內容和品質也不一樣。過去的調查、實驗等方法提供的數據是受訪者自己報告 (self-reported) 的認知、態度以及行為意向等，統稱為動機 (motivations)。即使是研究購買、選舉、媒介使用的行為，其實也只是測量了動機意向而已。我們知道，人的記憶很可能出錯，或者因為從眾心理 (social desirability)，人可能有意無意地說謊。而網絡數據中觀測到的數據是人類的行為，基本上沒有記憶、說謊等人為誤差 (當然有其他誤差，如機器人行為、水軍行為、廣告軟文等等)。

第三，觀測的時間粒度 (time resolution) 不同。過去傳統方法的時間粒度，往往非常粗略，如「過去」、「現在」，或「今年以來」、「本月 (周) 內」等等。最精確的是日記調查，也只能做到以小時為單位。網絡數據在時間上非常精確，一般以毫秒 (秒的千分之一) 為單位。這是大數據之所以「大」的一個原因。如果一個用戶的手機 24 小時開機 (我估計很多人是這樣的)，那麼他 / 她每天就有近一億條記錄！當然目前來講，我們並不需要或並不善於使用這麼精確的時間粒度，所以大數據存在著嚴重的誇大和浪費 (祝建華，2013)。我們這裡不談這些技術細節 (有興趣者可參見我下面提到的「時段」研究)。總的說來，精准的時間粒度是利遠大於弊，是對傳統研究數據的顯著進步。

XZ：對研究對象進行如此詳實而精確的測量和記錄，是否會有研究倫理方面的考量 (比如研究對象的知情同意、研究是否可能對研究對象帶來傷害等等)？

JZ：我每次談計算傳播學，都有人問這個問題，可見其受到普遍關心

《傳播與社會學刊》，(總)第44期(2018)

的程度。問題涉及很多層面和側面，這裡無法展開詳談，只講兩個重點：第一，迄今的討論都是基於提問者的個人感覺(anecdotal evidence)，缺乏嚴謹的數據，如問題的廣度(少數偶然還是普遍必然事件)和來源(演算法錯誤還是人為錯誤，如「內鬼」倒賣數據)等，所以亟需科學地加以評估；第二，如果不用這些數據和方法，還有更好的備選方案(alternatives)嗎？沒人能夠回答。我們即使不考慮後者的數據質量，其收集方法同樣問題多多，如入戶面訪涉及到更多的安全和隱私問題、電話調查容易被當作推銷或電信詐騙等等。Salganik (2018)在他的新書*Bit by Bit*中提倡「向前看」的態度，即現狀確實不盡人意，但可以逐漸改進。我完全認同他的說法。以下是我最近對一個批評算法的貼子的回覆，用來總結我的基本看法：

「我們需要理性討論這個問題。算法確實是一種打分系統。民主不就是基於打分嗎？如選舉，公投，議會，等等，不都是打分？打分的反面是精英的主觀判斷。哪個更好？當然，現在的算法打分，很多做的並不好，我們既不要盲目迷信，更不要因噎廢食。需要科學地評估，不斷改進，而不是回歸精英說了算。」

XZ：在傳統傳播學的研究中，問題的設計是由研究者決定的，研究者設立了問題，然後才去問這些問題。而在計算傳播學的語境下，對於研究對象的記錄都是被動的，那麼動機方面的問題如何加進去？是否應該是先有動機，才有行為？

JZ：你問了兩個問題：一是研究程序。傳統定量研究，確實是先假設、後檢驗的驗證性(confirmatory)研究為主，當然也有從數據中找感覺的探索性(exploratory)研究，那是支流，難為主流期刊接受。計算傳播學同樣有驗證或探索的兩種程序，各自所佔的比例還不清楚，但直覺上後者數量不少，很可能超過探索程序在傳統研究中的比例。探索性研究的價值及其局限，下面還會涉及，我先回答你的第二個問題：動機和行為的關係。

理想狀態下，我們同時擁有這兩者的數據。但事實上，如前面說的，傳統研究往往只有動機數據而網絡數據只有行為數據。計

對談計算傳播學：起源、理論、方法與研究問題

算傳播學研究中，是從人的行為去推測其動機。因此，從這個角度看，計算傳播學的一個基本（和隱含）假定是人的行為比動機更重要，因此研究行為更有意義和價值。以至於有一種觀點認為，動機不重要。因為人要做成一件事，光有動機還不夠，還需要有各種條件，如社會環境、個人能力等，才能有預定的結果。當然這就回到了社會科學的一個基本問題了。在社會學中有社會結構 (structure) 與個體能动性 (agency) 的辯論，在傳播學中有主動受眾 (如使用與滿足論) 與被動受眾 (如大眾受眾論，見 Webster & Phalen, 2013) 兩派的長期對峙 (Biocca, 1988)。

我們至少可以這麼說，基於行為數據的計算傳播學研究，不管是按驗證性還是探索性程序展開，其結果都為一些長期爭執不休的基本理論問題提供了前所未有的新證據。

如何把動機數據加到行為數據中去？如果有相同個體的動機數據 (如問卷調查)，如何整合只是個技術問題。現在主要的困難是缺乏這種動機數據，巧婦難為無米之炊。

XZ : 過往若干心理學、政治傳播等社會科學中使用的概念，比如人格 (personality)、態度 (attitude)、知識 (knowledge)、信任 (trust)、效能 (efficacy)、乃至意識形態 (ideology) 等，這些測量的都不是行為，那麼這些概念和測量還重要嗎？這方面的研究應如何與計算傳播學進行對話？

JZ : 現在已經有人開始用行為數據去推測動機 (包括人格、態度等，如 Eftekhari, Fullwood, & Morris, 2014; Wei et al., 2017)。我也一直有這方面的興趣，主要從用戶行為的時間數據入手，來推測人的社會角色、生活方式、動機、人格等等。我們最近剛剛完成了一項基礎性的研究，通過測量手機使用的「時段」(sessions) 來給用戶的生活方式畫像 (Zhu et al., 2018)。總之，你問動機研究與行為研究如何對話，一是很有必要、二是正在悄悄進行、三是很多最終結果 (如知行之間的相似度) 可能不是大家預期的。也許這正是包括計算傳播學在內的計算社會科學的魅力所在之處。

HY : 讓我們來談另一個相關話題，「計算社會科學」(computational social science) 是什麼時候興起的？

《傳播與社會學刊》，(總)第44期(2018)

JZ：「計算社會科學」(computational social science)的興起是一個漸進的過程，因為不同學科的學者都一直在進行相關的研究，只不過用了不同的名詞。如以英國為首的歐洲學者曾稱之為「e社會科學」(e-social science)，Wikipedia上有這個條目(https://en.wikipedia.org/wiki/E-social_science，核心部分還是我寫的)。而計算機學者則喜歡使用「社會計算」(social computing)的概念(我也參與其間，一直擔任「中國社會計算年會系列」的共同主席和組織者)；人文學者也有自己的體系，如「數碼人文」(digital humanities)。這些概念及其團體，至今還很活躍，並沒有完全被整合到計算社會科學的旗下。這很正常，科學研究是個思想自由市場(free marketplace of ideas)，任何品牌的興衰，都取決於大多數研究者的理性選擇。對我個人來說，經過深思熟慮之後，覺得計算社會科學最合適，「社會科學」是我們的身份認同(identity)、「計算」是我們區別於傳統方法的一個特徵標籤。

現在回到你的問題，如果要見諸文字的話，計算社會科學的「正式誕生」，應該就是David Lazar等16位學者於2009年發表在*Science*上的一篇題為「計算社會科學」的論文為標誌(Lazar et al., 2009)。這篇文章其實不是一個原創研究，而是一個「定位宣言」(position paper)，強調人們生活網絡化及其記錄(網絡數據)使得通過這些數據而研究社會的必要性和可行性。或者說，這篇文章的最大貢獻，就是其標題中的三個英文詞(computational social science，計算社會科學)，在各自為政的學術江湖中樹起一面富有召喚力和凝聚力的旗幟。

HY：我問這個問題，其實是想和傳播學研究的發展過程進行比較。我們都知道傳播學有「四大奠基人」。那麼到了計算傳播學這一研究取向的歷史進程會是怎樣的？

JZ：計算傳播學的進程，和傳播學早年的發展有點類似。我們可以引用施拉姆(Wilbur Schramm)曾經提出的兩個經典比喻來形容這兩者的相似之處。第一個比喻是在1960年代提出的，當時他說傳播學是一個「cross-road」，即十字路口，各學科的人騎著駱駝(即順著他們的研究興趣)路過此地，停留一陣之後又重新上路，

對談計算傳播學：起源、理論、方法與研究問題

各奔東西。到了1980年代，他改變了看法，提出了第二個比喻，叫做「oasis」，綠洲。意思是隨著傳播學的成熟，各個大學相繼培養了一批傳播學博士，比如他在伊利諾大學厄巴納－香檳分校 (University of Illinois Urbana-Champaign) 創辦的傳播研究所 (Institute of Communications Research)，就是最早的一個傳播學博士點，所以傳播學不再是其它學科的旅行者暫時客居之地，而是由傳播學原生學者為主的永久性學術共同體。

這段歷史與計算傳播學的發展很相近。到今天為止，計算傳播學的大部分研究是由其它學科的學者完成、並在傳播學之外的期刊和會議上發表。這些「遊客」中有些已經留下來了，在大學傳播院系內任教。這樣的例子在今後幾年內會越來越多。當然，他們是否能夠成為永久居民，還有待雙方的磨合，當中存在各種新問題、新挑戰。與此同時，傳播學自己培養的第一批計算學者也已經問世，其中有些就是從我這裡畢業的。他們可以算是計算傳播學的原生學者。當然，他們也面臨很多挑戰，與外來的計算學者相比，他們在計算上並沒有優勢；与其它傳播學者相比，他們在傳播學理論上也沒有明顯的優勢。但是，這些同時又是他們的價值所在：連接計算科學（也被稱為數據科學）與傳播學的橋樑。或借用社會網分析的術語來說，他們是跨學科的structural holes（結構洞）。對於他們來說，如何借助這個歷史性的機遇，協調和發揮計算學者和傳播學者雙方的互補作用，將是計算傳播學這片小綠洲能否持續發展的關鍵。

HY：剛才我們談到了傳播史，今天對傳播行為的理解和50年前不一樣。傳播學裡把傳播分為人際傳播、組織傳播、大眾傳播、國際傳播這些，您覺得這樣的區分在今天還有意義嗎？您認為何種行為才是核心的經典的傳播行為？

JZ：5個W是傳播學研究歷來的核心。可惜有些學者往往忽視、甚至藐視這5個W，感覺不夠「高大上」。我認為計算傳播學的研究範圍和主題，基本上就是傳播學研究的5個W。針對5W模型的批評，我想進一步談兩個看法。首先，有些學者批評5W是線性的、單向的，沒有包括反饋（feedback）。這個批評是對的。早期學者不研究

《傳播與社會學刊》，(總)第44期(2018)

反饋，主要因為缺乏相關的數據。現在隨著數據量的增加，研究反饋(如社交媒體上的用戶評論、點讚、轉發等等)已是一項標準動作了。但是另外一個批評則未必有道理了。有些人認為5W模型沒有考慮具體的語境，沒能做到語境化(contextualization)。我認為，這種看法恰恰就是傳播學多年以來進步不大的一個重要原因，我把這個現象稱做「過分語境化或具象化(over-contextualization)陷阱」。你剛才把傳播分為人際、組織、大眾，某種意義上，也是從contextualization的邏輯出發的。我認為，過分強調傳播的語境，就是強調事件的特殊性。5W模型的可貴之處，恰好就是超越了傳播現象中具體的語境(context)和範圍(level)，把傳播現象上升到概念(concept)的高度，從而使這個模型具有普適性。我們並不知道當時拉斯韋爾(Harold Lasswell)提出這個模型的時候是否想到了這點。但是他的思想來源是香農(Claude Shannon)的信息論，信息論原來研究的是工程問題，把信息傳播中所有材料的物理介質(機械、聲波、電子等等)都被抽象掉了，只剩下這5個W。抽象是科學化必須的一步。反之，強調語境的研究，更多的是具體的應用性研究，而非科學的抽象研究。

我們在最近的一項文獻分析中發現，隨著計算傳播學的普及，越來越多的人參與進來，同時也出現了過分語境化或具象化的趨勢(Zhu et al., 2017)。學術研究中，最難的部分就是既要從具體的現象出發，又要超越具體的現象。在用慣的術語中，最具象的、最具體的叫做事件，然後是現象(phenomenon)，最抽象的是現象背後的形成機制(driving mechanisms)。用到傳播學研究裡，現象是指對媒體的使用和效果、機制則指為何使用和為何有效。我們需要從具體的媒體、具體的用戶類別、內容中涉及的具體事件等現象抽象虛化掉，才能找到背後的傳播過程和機制。普遍的誤區是，把具體的場景當成現象、把現象當成機制。我們總是比真正的科學研究更「具象」一步。這是傳播學發展受阻的一大問題。

HY：可否舉幾個例子？

JZ：比如很多人研究茉莉花革命中的社交媒體，另外一批人研究佔領

對談計算傳播學：起源、理論、方法與研究問題

華爾街運動中的社交媒體。但是雙方各自被事件的獨特性(如社會、經濟、文化等等背景和一些機構、人物的特殊作用等等)所迷住了，所以分別寫成了茉莉花革命史和華爾街運動史，而社交媒體的使用及效果只成了其中的註解。

再舉一個我親身經歷的例子。我讀博士時，曾經把一篇研究中國民眾政治意識的文章投給 *China Quarterly*，結果未被接受，編輯說我的研究不夠 contextualize。我請教了一位研究中國問題的專家，他說你研究的是普遍問題，應該投給 *Asian Survey*，那也是一個不錯的期刊，後者果然接收了。我從中體會到，研究傳播問題有兩種取向，一是研究個性(如中國國情)，二是研究共性(如變量之間的普適關係)。前者應該投給 *Area Study* (地區研究) 的期刊(如 *China Quarterly*)，他們很清楚自己研究的就是具體的語境以及特定的案例。後者應該投給傳播學期刊(或政治、社會等期刊，取決於所涉及的理論問題)。換言之，可以發表在 *Journal of Communication* 的文章，未必可以發表在 *China Quarterly* 上，反之亦然。因此，我們則需要清楚區分，究竟一個研究是在研究具體的特例，還是研究抽象的現象及其背後的機制。

XZ：您剛提到研究需要超越具體的 context，那麼如果是比較研究呢？在社會科學上有 large-N 和 small-n 兩種取向，前者將 contextual 的變量(如國家、社區層面的變量)與個體的變量(個人層面的變量)區分開來，再使用多層模型(multilevel models)進行建模；而後者是在有限個案中考察變量的異同(如 Skocpol [1979]、Liebes & Katz [1993] 這些學者在政治社會學、跨文化傳播等領域的論著)。這些似乎都離不開對 context 的考察。比較傳播研究如何與計算傳播進行對話？

JZ：我在 2001 年的一篇文章(祝建華，2001)中，有一段對理論化與本土化的關係的描述，認為兩者是一個連續變量(continuous variable)的兩端。這個變量可以指的是「抽象化」的程度(context-free vs. context-specific)，或者叫做國際化的程度(internationalization vs. localization)。這個變量的一端是不受任何時空限制的人類傳播原理，另一端是某一個特定社會在某一特定時期的特定傳播現象。

《傳播與社會學刊》，(總)第44期(2018)

計算傳播學也應該強調這個變量的一端(普遍的人類規律)，而不應太過重視另一端(具體的現象)。

HY：計算傳播學現在的研究遇到什麼問題嗎？

JZ：現在計算傳播學還沒有達到傳統定量研究那樣成熟的地步，還處於初步發展階段。如，計算傳播學尚無一本公認的教科書；很多院校還沒有專門的課程、系統講授計算傳播學的理念、研究設計等核心內容，更常見的做法是在定量研究方法中添加幾個章節或講座，或者教一些專門的分析技術，比如text mining、online experiment等等。如果用羅傑斯(Everett Rogers)的創新擴散理論來看，現在處在第二個階段，就是快速增長(growth)的階段，參與者都算是早期採用者(early adopters)，但大部分研究還是零打碎敲式的，而不是在一個理論架構下的有機組合。有些朋友建議出一本《計算傳播學研究手冊》(*Handbook of Computational Communication Research*)之類的書，我則覺得為時過早。因為手冊應該具有相對穩定、豐富和獨立的內容，和其他學科的重合不應該太多，可是如果我們現在寫這本書，很多內容都是抄數據挖掘、機器學習、網絡分析等領域的工作。

其次，現在很多研究還是以描述為主、無意或無法解釋「why」的問題。大部分的計算傳播學研究都發表在非社會科學的期刊上，因為他們不太糾結why問題。但是一旦拿到了傳統社會科學期刊，就繞不過這個問題。說到「傳統」，這就是社會科學的範式(paradigm)。能否改變這個範式？我倒覺得未必。這和傳統研究方法的慣性有關，同時也和技術手段有關。很多解決計算傳播的研究工具尚未誕生，很多傳統的統計方法需要改進——有的方法甚至還未誕生。

舉個例子：網絡抽樣問題。傳統抽樣方法是基於獨立及相等分佈(independent and identical distribution)假設，即個體的行為是互不相關的。然而網絡化行為都是互相影響的，這就違背了獨立分佈的假設，從根本上挑戰了網絡抽樣。假設你有總體數據，如全部Facebook或微信的用戶，那就不需要操心抽樣。但是絕大多數學者拿不到這種數據。就算拿到了，數據體量大得驚人，難以

對談計算傳播學：起源、理論、方法與研究問題

存儲和拷貝。所以抽樣是一種必要和理想的方案，但是由於現有各種抽樣方法都基於獨立分佈假設，所以全不適用。我們做過大量實驗，發現現在還沒有真正適合用於網絡抽樣的方法 (Xu & Zhu, 2016)。我也曾請教過做概率統計的學者，一開始他們覺得這是小菜一碟，但是後來發現正如我說的，網絡抽樣確實很難，但是他們沒有興趣花時間研究，因為這是一個「應用」問題，在他們學科裡價值不高。這類問題其實不少，很多目前都懸而未決。

HY：請對未來研究趨勢做些分析。

JZ：我們先分析現狀，再談談理想狀況。大部分計算傳播學研究還停留在文本挖掘上，方法上不夠多元，容易陷入過分語境化或具象化的泥坑。傳統學者看來，似乎就是so what的問題，即使你畫出更漂亮的圖、算出更準確的參數，但還是「新瓶舊酒」。理想狀態呢？其實未必需要挖空心思去琢磨。我們希望未來有更多對行為和效果的研究。傳播效果是有階梯的，從認知、態度到行為。最終目的還是行為的採納或改變。以前因為行為不易研究，所以退而求其次，研究動機和態度。現在有了行為數據，為何還要繞著走？

HY：談到研究需要抽象，您在八十年代譯介了麥奎爾 (Denis McQuail) 和溫德爾 (Sven Windahl) 的《大眾傳播模式論》(McQuail & Windahl, 1981)。其中您讚揚McQuail對理論的抽象化的表述。這些表述對計算傳播學有什麼啟發？

JZ：《大眾傳播模式論》確實是一本好書，其中最重要的一個特點就是用圖像的方式描述抽象的概念及其之間關係，可以叫做「理論可視化」。書中很多理論的原作者自己都沒有想到過這種表述方法。McQuail和Windahl將其概念和關係提取出來，比原先的文字描述更抽象了一步，但形式上看上去卻好像更具象 (其實只是圖像比文字更形象的錯覺而已)。其實我在第二版的譯者前言中提到，兩位作者所做的抽象還可以更進一步，如使用更抽象的二維座標圖，從而可以直接和數學公式相對應 (祝建華，2008)。Chaffee (1991) 在他的 *Explication* 一書中詳細討論了概念明示精準問題，一個研究需要闡釋 (explication) 其理論的增量是多少，多少是跟前人重複的，以及概念可否操作化等問題。

《傳播與社會學刊》，(總)第44期(2018)

HY：所以說，一個學科的抽象化的程度，顯示了一個學科的成熟程度。那麼計算傳播學有沒有屬於自己的獨特的學科邊界和知識發展體系？

JZ：很多人批評傳播學沒有獨特的學科邊界和知識體系，我不敢苟同。傳播學有自己獨特的學科體系，其核心就是5W。80%–90%的傳播學理論和研究回答的是5W中的某個問題。相反，從其它學科引入的理論和案例，不一定涉及5W問題。計算傳播學呢？同樣也是以5W為核心，那是傳統傳播學的知識財富，千萬不要把寶貝嬰兒當作洗澡水潑掉。

那麼，計算傳播學可以或者應該在哪些方面超越傳統定量研究呢？我認為大概有以下幾方面：

(1) 研究用戶行為而不是他們的動機或態度(前面已有所討論，這裡不再展開)；

(2) 研究傳播的動態過程而不是其靜態結局(以前也說傳播是個動態過程，但是經常因為缺乏時間信息而無法實現)；

(3) 研究網絡化條件下的用戶行為(以前也承認人與人之間的關聯，人跟環境之間的互動，但由於缺乏相關的數據和分析手段而停留在光說不做(lip service)階段)；

(4) 研究從個體(用戶)到團體(家庭、群體等)到總體(社會)等不同層次之間的交互影響(以前也有多層分析、但只限於少數幾個共用數據，現在的網絡數據，幾乎每個都涉及多層結構)；

(5) 研究具有地理空間位置信息的用戶行為和傳播過程(這在以前基本不可能)；

(6) 採用描述性(descriptive)或探索性(exploratory)研究(這在以前則是被鄙視的)。這一點在計算傳播學內外都有很多爭議。我認為描述性和探索性研究在任何新研究領域(如社交媒體、移動媒體、穿戴媒體等)的早期是必要和有益的，所以目前應該允許甚至鼓勵。當然，三五年後，如果我們還停留在探索和描述階段，那就有問題了。

計算傳播學的這些特點，是否能夠產生新的知識？現在還難說，要靠大家的努力。進一步問，是否能夠帶來新的範式轉移

對談計算傳播學：起源、理論、方法與研究問題

(paradigm shift)? 更是有待時日檢驗。1983年，傳播學刊 (*Journal of Communication*, JoC) 出了一期主題為「正在發酵中的傳播學」(Ferment in the Field) 的專輯，提出了傳播學範式轉移的必要性，引起很大反響和共鳴。大家一方面是對現狀範式的不滿，另一方面又覺得推陳出新似乎很容易。但是三十多年來，傳播學還在發酵之中，翻來覆去講的只是「推陳」的必要性，誰也沒有「出新」的解決方案。1993年JoC又出了一期發酵專輯(文章更多，所以分了兩期登完)，今天沒有幾個人記得這件事了。最近JoC在編第三次發酵專輯，還沒有問世，有甚麼解決方案嗎？大家拭目以待吧。我押注在計算傳播學上。它已是一種有別於現有範式的新範式，也許短期內還不會取代現有範式，但很有可能成為主流範式之一。當然計算傳播學面臨的問題和挑戰也很多。借用上海財經大學2017年末召開的一個數據新聞研討會的總結報導的標題「數據新聞一哄而上，問題一把」(楊顏菲、劉璐，2017)，計算傳播學也有點一哄而上，隨之出現一堆疑問。我們計算傳播學者要正視問題，避免曇花一現的結局。

HY: 如果博士研究生要進入這個領域，需要有什麼注意的地方？如何啟發他們的研究思路？

JZ: 計算傳播學方法需要有更長的訓練時間，大家首先對此要有充分的心理準備。理工科背景的需要學習經典的社會科學理論和方法；文科背景的需要學習傳統的定量方法。另外，不要放棄傳統的量化研究方法，如不熟悉則必須學會。如果是傳統方法做不好而來趕計算傳播學這個新潮的話，恐怕做不深，往往是新瓶舊酒。我能夠在計算傳播學裡有點發言權，並不是因為我的計算方法特別強，而是在傳統的理論、方法、尤其是研究設計 (research design) 上有比較多的經驗積累。微軟研究院的數據可視化專家在社會科學文獻中發現我20多年前的「零和遊戲」理論，很適合作為展示公眾輿論之間競爭關係的框架，於是來找我們合作，結果發表在計算機可視化的頂級期刊上 (Xu et al., 2013)。最近北京大學計算機學院請我在他們2018年暑期學校中給來自海內外的計算機學生開設一門計算社會科學課程，我將把重點放在社會科學的研究設計上。

《傳播與社會學刊》，(總)第44期(2018)

XZ：計算社會科學中有一種多主體建模(agent-based modelling, ABM)的研究方法。在傳播學領域，曾有學者用ABM來研究傳播行為(如Song & Boomgaarden, 2017)。您覺得這種研究方法，會在未來成為計算傳播學研究的一個分支嗎？

JZ：我必須先聲明一下，我對ABM的瞭解相當膚淺，流覽過一些涉及傳播問題的論文，沒有被打動過，所以以下評論，僅限於傳播研究中的ABM，也許並不能推廣到其它領域。我對這類研究的主要質疑是缺乏基準真相(ground truth)的驗證。什麼是基準真相？簡單說，真實世界的代表性數據。剛才我一直都在說，科學的原動力，來自於數據；因為有了數據，才有人去研究；有人研究了，才會發現現有的方法不足，才有方法的進步；有了方法的改進，才会有科學研究發現的突破。科學發展的三部曲應該是數據、方法、理論。ABM的邏輯過程是理論假設—仿真數據—驗證理論，當中跳過了實證數據。沒有後者，如何證偽理論模型？當然，有些ABM研究也用到了實證數據，放在最後，比較模擬數據與真實數據之間的相似性。其實這種比較並非嚴格的驗證或證偽，基本上都是在整體層面(aggregated level)上的類比，我還沒有看到過個體層面的證偽案例(當然也許存在，我並沒有做過系統的檢索)。哪怕兩者相似甚至高度相似，還不是證明的必要條件，並沒有排除產生這種相似的其它原因。舉個經典案例，人類行為的時間分佈，整體上呈冪律分佈(power-law distribution, 見Barabasi, 2005)。最初的發現者Albert-Lazsol Barabasi等並進一步假設這是由於人們根據各種行為不同重要性而做出的不同優秀排序而決定的(Vázquez et al., 2006)。如果用ABM來研究，可以設定各種行為的排序參數，然而隨機生成數據，再混在一起，果真是冪律分佈。這樣就證明了優先排序論嗎？沒有。因為還有很多其它可能性，如人的記憶等(Vázquez, 2007)。

對於ABM在傳播學中的必要性和可行性，我的基本看法是，在沒有數據或可靠數據的情況下，有些學者採用模擬的方法製造仿真數據，是可以理解的。但是我們必須知道，這種研究僅僅展示了有關理論假設的「表面可能性」(probable cause)而非該理

對談計算傳播學：起源、理論、方法與研究問題

論的最後證明，後者需要有真實數據的嚴格證偽。大數據時代還需要模擬仿真數據嗎？對於有志做ABM的學者，我的建議有兩點。第一，一定需要用真實數據來做驗證(validation)，才能提出有突破性的結果。第二，要仔細查看一下過去相似的研究，從中總結一下經驗和教訓。ABM在傳播學中的應用早已有之，最常見的就是沉默螺旋理論。我曾經看到過數十篇類似的文章，都號稱自己解決了一些理論問題。一開始我每看到這類文章都很激動，每篇都仔細看，結果發現這些文章的結論五花八門，每一篇都不一樣。儘管每一篇都內在自成系統，有些故事也很有趣，但是每個研究無法驗證、互相之間也無法對話。這種困局，好多年了，如何走出來？我認為出路還是在實證數據之中。

參考文獻

中文部分 (Chinese Section)

- 楊顏菲、劉璐 (2017年11月13日)。〈從一哄而上到問題一把，數據新聞的未來在哪裡？〉《上觀新聞》，取自<http://www.jfdaily.com/news/detail?id=70953>。
- 祝建華 (2001)。〈中文傳播研究之理論化與本土化：以受眾及媒介效果的整合理論為例〉。《新聞學研究》，第68期，頁1-21。
- 祝建華 (2008)。〈譯者的話〉。《大眾傳播模式論 (第二版)》，頁3-4 (麥奎爾和溫德爾著、祝建華譯)。上海：上海譯文出版社。
- 祝建華 (2013)。〈一個文科教授眼中的大數據：多、快、糙、耗〉。《大數據中國》，第1期，頁10-12。
- 祝建華、彭泰權、梁海、王成軍、秦潔、陳鶴鑫 (2014)。〈計算社會科學在新聞傳播研究中的應用〉。《科研信息化技術與應用》，第5卷第2期，頁3-13。

英文部分 (English Section)

- Barabasi, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), 207-211.
- Biocca, F. (1988). Opposing conceptions of the audience: The active and passive hemispheres of mass communication theory. *Annals of the International Communication Association*, 11(1), 51-80.
- Chaffee, S. H. (1991). *Explication*. Newbury Park, CA: Sage.

《傳播與社會學刊》·(總)第44期(2018)

- Eftekhari, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Computers in Human Behavior*, 37, 162–170.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., & Jebara, T. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915), 721.
- Liebes, T., & Katz, E. (1993) *The export of meaning: Cross cultural readings of Dallas*. Cambridge: Polity.
- McQuail, D., & Windahl, S. (1981). *Communication models for the study of mass communications*. Longman.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., & Pinker, S. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182.
- Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton University Press.
- Skocpol, T. (1979). *States and social revolutions: A comparative analysis of France, Russia and China*. Cambridge University Press.
- Song, H., & Boomgaarden, H. G. (2017). Dynamic spirals put to test: An agent based model of reinforcing spirals between selective exposure, interpersonal networks, and attitude polarization. *Journal of Communication*, 67(2), 256–281.
- Vázquez, A., Oliveira, J. G., Dezsö, Z., Goh, K. I., Kondor, I., & Barabási, A. L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(3), 036127.
- Webster, J. G., & Lichty, L. W. (1991). *Ratings analysis: Theory and practice*. Lawrence Erlbaum Associates.
- Webster, J. G., & Phalen, P. F. (2013). *The mass audience: Rediscovering the dominant model*. Routledge.
- Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., Rui, Y., & Ma, W. Y. (2017, February). Beyond the words: Predicting user personality from heterogeneous information. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 305–314). ACM.
- Xu, P. P., Wu, Y. C., Wei, E. X., Peng, T. Q., Liu, S. X., Zhu, J. J. H., & Qu, H. M. (2013). Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2012–2020.
- Xu, X. K., & Zhu, J. J. H. (2016). Flexible sampling large-scale social networks by self-adjustable random walk. *Physica A: Statistical Mechanics and its Applications*, 463, 356–365.
- Zhu, J. J. H., Guan, L., Peng, T. Q., & Liang, H. (2017). New bottle of old wine vs. old bottle of new wine: What contributions have computational methods made to social media research? *The 68th annual conference of International Communication Association*, San Diego, USA, May.

- Zhu, J. J. H., Chen, H. X., Peng, T. Q., Liu, X. F., & Dai, H. X. (2018). How to measure sessions of mobile device use? Quantification, evaluation, and applications. *Mobile Media & Communication*, 6.

祝建華著作選

- Zhu, J. H. (1992). Issue competition and attention distraction: A zero-sum theory of agenda-setting. *Journalism Quarterly*, 69(4), 825–836.
- Zhu, J. H., Watt, J. H., Snyder, L. B., Yan, J., & Jiang, Y. S. (1993). Public issue priority formation: Media agenda-setting and social interaction. *Journal of Communication*, 43(1), 8–29.
- Zhu, J. H., Milavsky, J. R., & Biswas, R. (1994). Do televised debates affect image perception more than issue knowledge? A study of the first 1992 presidential debate. *Human Communication Research*, 20(3), 302–333.
- McCombs, M., & Zhu, J. H. (1995). Capacity, diversity, and volatility of the public agenda: Trends from 1954 to 1994. *Public Opinion Quarterly*, 59(4), 495–525.
- Zhu, J. H., Weaver, D., Lo, V. H., Chen, C., & Wu, W. (1997). Individual, organizational, and societal influences on media role perceptions: A comparative study of journalists in China, Taiwan, and the United States. *Journalism & Mass Communication Quarterly*, 74(1), 84–96.
- Zhu, J. J. H., & He, Z. (2002). Perceived characteristics, perceived needs, and perceived popularity: Adoption and use of the Internet in China. *Communication Research*, 29(4), 466–495.
- Zhu, J. J. H., Mo, Q., Wang, F., & Lu, H. (2011). A random digit search (RDS) method for sampling of blogs and other web content. *Social Science Computer Review*, 29(3), 327–339.
- Peng, T. Q., Zhang, L., Zhong, Z. J., & Zhu, J. J. H. (2013). Mapping the landscape of Internet studies: Text mining of social science journal articles 2000–2009. *New Media & Society*, 15(5), 644–664.
- Zhang, L., & Zhu, J. J. H. (2014). Regularity and variability: Growth patterns of online friendships. *International Journal of Web Services Research*, 11(4), 24–36.
- Wang, Z. Z., & Zhu, J. J. H. (2014). Homophily versus preferential attachment: Evolutionary mechanisms of scientific collaboration networks. *International Journal of Modern Physics C*, 25(5), 1440014.
- Jiang, L. C., Wang, Z. Z., Peng, T. Q., & Zhu, J. J. H. (2015). The divided communities of shared concerns: Mapping the intellectual structure of e-health research in social science journals. *International Journal of Medical Informatics*, 84(1), 24–35.
- Ackland, R., & Zhu, J. J. H. (2015). Social network analysis. In P. Halfpenny & R. Procter (Eds.), *Innovations in digital research methods* (pp. 221–244). Sage Publications.

《傳播與社會學刊》，(總)第44期(2018)

Liang, H., & Zhu, J. J. H. (2018). Big data, collection of (social media, harvesting).
In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *International handbook of
communication methods*. Wiley & Sons.

本文引用格式

祝建華、黃煜、張昕之(2018)。
〈對談計算傳播學：起源、理論、方法與研究
問題〉。《傳播與社會學刊》，第44期，頁1-24。