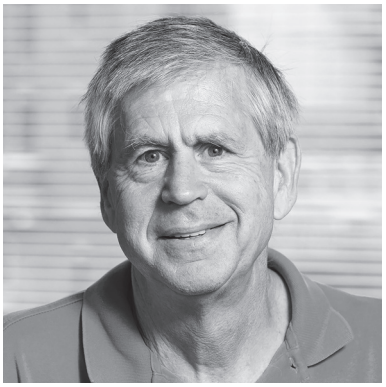
《傳播與社會學刊》，（總）第55期（2021）：1–18

學術對談

探索數位「社會之鏡」的未知疆域——計算社會科學先驅的洞見

對談人：邁克爾．梅西、宋韻雅、盧遠航
編　輯：宋韻雅
翻　譯：盧遠航



邁克爾．梅西教授
（Prof. Michael Macy）

「如果你問大多數計算社會科學領域的研究人員什麼最重要，大多數人會把機器學習列在榜單之首。具體來講，它包括詞嵌入、文本數據分析以及圖像處理。這些方法在計算機和信息科學的課堂上定期講授，卻很少見諸社會科學的教學內容。這使我想知道計算社會科學是否愈加可能從一種『社會科學技術派』的學科轉變成『計算科學社科派』的學科。我已經觀察到，一些社會學家對學科殖民化感到恐懼，表現出一定的防備心理，但這種防備卻可能導致社會科學錯過一次引領歷史進程的機會。」

宋韻雅，香港浸會大學傳理學院副教授。研究興趣：國際傳播、線上社會網絡、新聞學研究。電郵：yunyasong@hkbu.edu.hk
盧遠航，香港浸會大學傳理學院博士候選人。研究興趣：線上社會網絡、政治傳播、危機傳播。電郵：kevinlu@life.hkbu.edu.hk

Communication and Society, *55* (2021), 1–18

Dialogue

Exploring the Uncharted Terrain of the Digital "Socioscope": Insights from a Pioneer of Computational Social Science

Discussants: Michael MACY, Yunya SONG, Yuanhang LU
Editor: Yunya SONG
Translator: Yuanhang LU

Abstract

The continuously increasing availability of digital trace data has made the digital space a potent "socioscope" available to social scientists in their explorations of human behavior and social systems. A pioneer and founder of the field, Professor Michael Macy has witnessed the rapid growth of computational social science since the 1990s. What attracts and drives him to focus on the computational approach to social science? What does he think about this thriving field? What does he think about the future of the computational approach in social science research? In this dialogue, Professor Macy will discuss his academic career and research experience, his understanding of the computational approach and its methods, his views on cultural sensitivity in computational social science, the relationship between computational experiments and real-world phenomena, as well as his opinions about academic training in computational social science.

Yunya SONG (Associate Professor). School of Communication, Hong Kong Baptist University. Research interests: global communication, online social networks, journalism studies.
Yuanhang LU (Ph.D. candidate). School of Communication, Hong Kong Baptist University. Research interests: social network analysis, political communication, crisis communication.

Citation of this article: Macy, M., Song, Y., & Lu, Y. (2021). Exploring the uncharted terrain of the digital "socioscope": Insights from a pioneer of computational social science. *Communication and Society, 55*, 1–18.

致謝

本文承美國依阿華大學社會學系助理教授施永仁博士提出寶貴意見，特此致謝。

邁克爾．梅西教授簡介

作為世界計算社會學界的領軍人物，邁克爾．梅西現為美國康奈爾大學戈德溫．史密斯文理教授，兼任社會學系和信息科學系教授，以及社會動力學實驗室主任。梅西在哈佛大學取得博士學位後，於康奈爾大學開展其學術事業。他一直致力於計算社會學、社會網絡分析、政治極化以及社會影響等領域的研究，並因其開創性貢獻而享譽學林。他重點關注如何利用新穎的網絡數據源、計算模型和在線實驗來解決有關人類行為和社交互動的長期難題，尤其是考察人群聚合與離散的原因。他帶領一眾社會學家、信息科學家以及計算機科學家，使用計算模型、在線實驗室實驗以及媒介化互動的數字蹤跡來探索為人熟悉卻又難以釐清的社會互動模式。他當前的研究興趣包括：黨派差異在個人生活和興趣中的表現方式、意見級聯及其不可預測性，以及全球音樂流背後的跨文化情感節律。梅西的大量著述均見諸世界頂級學術期刊，包括*Science*、*Nature Human Behaviour*、*Science Advances*、*American Journal of Sociology*、*American Sociological Review*以及*Annual Review of Sociology*，橫跨社會學、計算科學以及自然科學等諸多領域，成就斐然。

MM：邁克爾．梅西
YS：宋韻雅

YS： 近二十年來，計算社會科學領域的學術研究與社會影響都經歷了前所未有的進步。從上世紀90年代起，您便投身該領域，並見證了它的發展和壯大。您之前是否預料到有這樣的巨大變化？您如何看待目前這個蓬勃發展的領域？除了互聯網技術的變革力量之外，還有什麼關鍵力量在推動這個領域的發展？

MM：這樣的增長主要得益於兩個層面，即數據和計算。人們在全球範圍內通過在線互動留下數字痕跡，這樣龐大的範圍是前所未有的。人們能夠在虛擬實驗室中進行隨機對照試驗，甚至能夠在Kickstarter等平台上進行田野實驗。除此之外，如今我們還擁有用於收集、管理和分析大量半結構化數據的計算資源。

YS： 您如何定義計算社會科學？是否有示範性的相關研究？

MM：由大衛．拉澤(David Lazer)牽頭，包括我在內的大約30位學者在《科學》雜誌上發表過一篇論文。這篇論文對該領域進行了定義。 我只想補充一點，我們應該真正認識到計算建模的重要性，而該領域不僅僅只是分析來自社交媒體或其他網站的數據。

YS： 您在研究中所倡導的「關係宣言」(relational manifesto)給我們留下了深刻的印象，該宣言為您對社會問題進行計算研究提供了理論基礎，您能否解釋一下它如何幫助研究者更好地觀察和理解由複雜網絡構成的當前社會？

MM：與研究社交網絡的「關係線」相比，研究社交網絡的「節點」要容易得多。社會科學家在很大程度上依賴於問卷調查來觀察和記錄個人的屬性，而行為科學家已經開發出強大的實驗方法來探究人們是如何做出決策的。但由於多種原因，研究人際互動還是困難重重。當使用問卷調查來研究網絡時，我們有可能將主觀感知到的網絡與客觀存在的社交網絡相混淆，比如問卷調查的措辭表述很可能影響報告的結果。而且很多非公開的內容是研究人員幾乎無法觀察和記錄的，人們通常更喜歡在公共領域之外進行互動，即使在公共領域，人際互動也可能會較短暫。此外，頻繁的互動會導致節點數和關係數之間的規模呈指數增加。更糟糕的是，大多數研究人員接受的統計方法的訓練都是基於研究獨立個體的，而這些方法所做出的假設通常在研究互動關係時並不成立。行為科學家已經使用隨機對照試驗來研究人與人之間的互動，但是直到最近，參與這些研究也大多是學習基礎心理學課程的大學生。簡而言之，網絡科學的最新進展以及來自全球社交網絡的數據，為研究節點之間的關係提供了前所未有的契機，包括社會影響力、同質性、傳染性、信息級聯、擴散和極化等過程。

YS： 當使用計算方法來考察人類在社會中的互動時，您會怎樣看待來自文化差異的潛在影響？例如，在東亞的集體主義社會中，人們可能具有不同的社會互動規則。對不同地區的研究人員來說，以計算方式來考察和解釋社會問題的時候，文化敏感性很重要嗎？

MM：通常來講，文化差異、性別和種族差異，以及世代效應等因素，都對社會科學中的所有方法和模型影響深遠。值得注意的是，雖

然性別、種族和世代可以由問卷調查的數據來測量，但是對於跨文化研究來說，問卷調查和傳統上基於實驗室的隨機對照試驗都比較難以實施。因此，基於問卷調查和實驗室的研究通常忽略了跨文化分析。然而，現在來自全球社交媒體和其他平台的在線數據使得研究跨文化的行為研究成為可能。之前，我和斯科特．戈德爾（Scott Golder）使用Twitter數據來追蹤跨文化的晝夜節律（diurnal rhythms）。通過與Spotify的研究人員合作，我和Minsu Park最近使用有關全球音樂下載的情感屬性的數據重新討論了這個主題。在這之前，我和Minsu曾使用YouTube數據來研究跨文化視頻消費。我們的實驗室還使用Twitter數據測試了亨廷頓頗有爭議的關於「文明的衝突」的跨文化理論。另外，帕特里克．帕克（Patrick Park）、喬什．布魯門斯托克（Josh Blumenstock）和我使用全球網絡數據來揭示「社交蟲洞」，即跨越全球社交網絡的高帶寬快捷連結方式。在我開啟學術生涯的1980年代，這些研究是不可想像的。

YS：您曾認為，在當前研究中，「廣泛共享的基於中層理論的分析工具包（模型和方法）」比「普適性的理論框架」更為重要。這是否意味著，在宏觀理論的構建之前，我們應該首先從海量數字化數據中挖掘出更多的新興理論並運用跨領域的分析工具？

MM：關於人類互動的宏大理論可以為社會科學和行為科學奠定更堅實的基礎，但前提是該理論在經驗上有效。之前構建普適理論的嘗試（例如，馬克思主義、弗洛伊德主義、行為主義和效用最大化）的嘗試並非總能奏效，儘管已證明某些理論比其他理論更有用。中層理論的建構過程更像是在拼接一個巨大的拼圖，即使拼圖零片的數量有限，也不能保證這其中只有唯一的拼組解法。相應地，大多數社會科學家認為，為了學科的發展，我們應該專注於提高測量能力和因果推理能力。因為當通過觀察到的模式來檢驗假設時，就像是在現有的知識背景下玩拼圖。這種方法傾向於產生漸進式的進步，其更多地取決於我們測量和建模能力的提高，而不是提升我們講述一個令人信服的故事的能力。而當拼圖遊戲

接近完成時，我們將開始看到整幅畫面。到目前為止，似乎每當我們弄清楚拼圖零片的放置位置時，我們都會發現，實際上拼圖中存在著比我們意識到的更多的零片。也許有一天會掉落更多的零片，但如果想根據這些假設得出宏大理論，似乎還遠遠不夠。

YS：近年來，計算社會科學領域中各種研究方法層出不窮，最讓您興奮的方法是什麼？您認為將來的研究中最具前景的方法是什麼？或者您認為研究人員仍迫切需要哪些方法？

MM：如果您問大多數計算社會科學領域的研究人員什麼最重要，大多數人會把機器學習列在榜單之首。具體來講，它包括詞嵌入、文本數據分析以及圖像處理。這些方法在計算機和信息科學的課堂上定期講授，卻很少見諸社會科學的教學內容。這使我想知道計算社會科學是否愈加可能從一種「社會科學技術派」(technical wing of the social sciences) 的學科轉變成「計算科學社科派」(social science wing of the computational sciences)的學科。我已經觀察到，一些社會學家對學科殖民化感到恐懼，表現出一定的防備心理，但這種防備卻可能導致社會科學錯過一次引領歷史進程的機會。

YS：您如何看待使用計算方法的社會模擬與觀察真實世界之間的關係？

MM：三十年前，基於行動者的建模改變了計算機模擬仿真，將「(找尋能概括)總體的模型」(model of the population)替換為「(概括)模型的總體」(population of models)。這些微觀仿真使人們有可能探索行為和結構假設的宏觀含義，而不是依靠理論上的直覺。依賴直覺的問題在於，複雜系統的許多結果是高度違反直覺的。但是，研究者(包括我在內)的早期努力確實是帶有「玩具協會」的性質。與問卷調查研究和政府統計領域的多元線性混合模型相比，它們更接近於實驗科學中的隨機對照試驗方法。一個經過精心設計的隨機對照試驗也是一個玩具，但它仍然可以成為一種強大的工具。如果實驗變得太過復雜，以致只能使用多元模型來分析結果，那麼單單進行因果推理的能力可能會受到嚴重損害。這一點通常也適用於仿真模型。我們不應將使模擬變得更「現實」作

為一種目的。除了基於行動者的模型，計算建模的另外兩種發展也給我留下了深刻的印象：經驗校準和雲計算技術。前者取得了長足的進步，後者在系統地探索廣闊參數空間上取得了許多突破。最令人印象深刻的例子是三角研究中心（Research Triangle Institute）的突破性研究，他們通過使用計算模型把多個數據集合併為「合成總體」（synthetic populations）。可以看到，他們在使用合成總體對流行病建模方面取得的成功，相當於將這個領域的發展推進了十年！鄧肯．沃茨（Duncan Watts）曾經認為來自社交媒體的數據將成為我們的「太空望遠鏡」——一種使我們能看到前所未見之事物的突破性工具。也許合成總體將會成為我們新的「太空望遠鏡」。

YS：計算社會科學的訓練必須同時發展計算素養和社會素養。具有社科背景的學生對社會問題具有較高敏感性和洞察力，但可能缺乏計算能力，而掌握計算能力的學生在面對社會議題時，卻可能提不出適當的問題。對有志於計算社會科學的各種學生來講，您認為怎樣的訓練能夠解決他們的這些困難？

MM：我同時在社會學和信息科學系任教，而社會動力學實驗室兼有這兩個系以及計算機科學和應用數學系的博士生。在研究項目的合作過程中，學生們互相學習研究所需的技能和理論框架。隨著時間的流逝，越來越難以分清哪個學生來自哪個學科背景。過去人們常常認為計算機科學家懂得方法技能但缺乏理論，而社會科學家理論紮實卻計算技能不足，但現在這些跨學科團隊表明，人們之前的看法太過武斷了。例如，計算機科學家其實在網絡擴散過程方面有很強的理論基礎，而社會科學家在對觀測數據進行因果推論方面也很擅長。因此，協同合作的機會是巨大的。

YS：回想過去，是什麼吸引並促使您專注於社會科學的計算方法，尤其是在這種方法受到較少關注的上世紀90年代？這是來自於理論或方法論層面上的考慮？還是源於您個人的研究興趣？

MM：我從1989年開始使用計算模型，是因為受到羅伯特．阿克塞爾羅德（Robert Axelrod）關於囚徒困境的進化模型的啟發。我對他

的模型進行了簡單的更改，將「模仿優勝者」(imitation of the fittest) 替換為「優勝劣換」(win stay, lose switch)，也就是強化學習。這些早期模型是非常簡單的玩具模型，但也得益於桌上型電腦運算的進步。它使得計算模型可以廣泛使用，並展示了如何使用計算來探索關於人類行為和社會互動的一系列假設背後的邏輯含義。

YS：這好像是第一次中國學界跟您這樣的計算社會學大家展開深入對話。您的文章〈數字足跡：在線社交研究的機遇與挑戰〉在中國學界被視為經典之作。您有什麼特別要與他們分享的嗎？

MM：感謝您的關注並培訓了新一代的計算社會科學家！國際合作的機會與建立跨學科邊界的橋梁一樣重要。

YS：對於如何在計算社會科學方面取得成功，您會給研究生們什麼建議？

MM：不要以「前人未研究」這樣的理由來選擇問題，「前人未研究」可能是因為覺得回答該問題價值不高，或者該問題難以被解答。實際上恰恰相反，研究者只管提出有趣而重要的問題，不要管它已經被研究過多少次。這其中總是會有新的東西需要探索——這就是為什麼它被稱為「科學」而不是「宗教」。宗教源自信仰，而科學需要質疑。人們能夠在信仰裡得到慰藉，卻必須在質疑中而尋找答案。

邁克爾．梅西教授著作選

Macy, M. W. (1991). Chains of cooperation: Threshold effects in collective action. *American Sociological Review*, *56*(6), 730–747.

Macy, M. W., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review*, 638–660.

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*(1), 143–166.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*(3), 702–734.

Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science*, *328*(5981), 1029–1031.

Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878–1881.

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*(1), 129–152.

Shi, F., Shi, Y., Dokshin, F. A., Evans, J. A., & Macy, M. W. (2017). Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nature Human Behaviour*, *1*(4), 1–9.

Park, P. S., Blumenstock, J. E., & Macy, M. W. (2018). The strength of long-range ties in population-scale social networks. *Science*, *362*(6421), 1410–1413.

Macy, M., Deri, S., Ruch, A., & Tong, N. (2019). Opinion cascades and the unpredictability of partisan polarization. *Science Advances*, *5*(8), eaax0754.

本文引用格式

邁克爾．梅西、宋韻雅、盧遠航（2021）。〈探索數位「社會之鏡」的未知疆域——計算社會科學先驅的洞見〉。《傳播與社會學刊》，第55期，頁1–18。

Academic Dialogue with **Michael MACY**

Exploring the Uncharted Terrain of the Digital "Socioscope": Insights from a Pioneer of Computational Social Science

MM: Michael MACY
YS: Yunya SONG

YS: Over two recent decades the field of computational social science has experienced rapid growth in both its academic research and in its social influence. From the 1990s you began conducting research in this field and you have witnessed its growth and development. Did you anticipate the dramatic changes we have seen? Apart from the transformative power of internet technologies, were there other key forces that drove the field's growth?

MM: Two developments led to this growth, data and computation. Online interaction leaves digital traces of human behavior at an unprecedented word embedding global scale. This includes the ability to conduct RCT experiments in virtual labs and even to conduct field experiments on platforms like Kickstarter. We also now have the computational resources to collect, manage, and analyze massive semi-structured data.

YS: How do you define computational social science? And how do you regard this thriving field currently? Does the field have any exemplary, or even canonical studies?

MM: There is the Science paper by David Lazer and about 30 others (including me). That paper has a definition of the field as well. I would only add that we should be sure to recognize the importance of computational modeling—the field is not just about analyzing data from social media or other web sites.

YS: Your remarkable "relational manifesto" provides a theoretical basis for computational enquiries into social issues. Can you

explain how it can help researchers better observe and better understand today's complex social networks?

MM: It is much easier to study the nodes of a social network compared to the edges. Social scientists rely heavily on survey methods to observe and record the attributes of individuals, and behavioral scientists have developed powerful experimental methods to probe how people make decisions. It is much harder to study human interaction, for many reasons. When we use surveys to study networks, we risk conflating the perceived network with the objective social network, and the relations people report are likely to be sensitive to how the survey items are worded. Moreover, people generally prefer to interact outside the public sphere, and even when in public, the interactions can be fleeting, which makes it nearly impossible for researchers to observe and record what takes place. And then there is the exponential increase in scale between the number of nodes and the number of edges. To make matters worse, most researchers are only trained in the statistical methods used to study individuals, and these methods make assumptions that generally do not hold when studying relations among individuals. Behavioral scientists have used RCT experiments to study interpersonal interaction, but until recently these were often limited to small numbers of college sophomores taking introductory psych courses. In short, recent advances in network science, along with data from global social networks, have opened up unprecedented opportunities to study the relations among the nodes, including processes of social influence, homophily, contagion, cascades, diffusion, and polarization, to name a few.

YS: In recent years we have witnessed increasing ideological and cultural polarization, especially in the U.S., but elsewhere around the world as well. Nationalism and even isolationism seem to be on the rise. Your relational explanation of polarization is intuitive and enlightening. Based on your model, is there a solution that could trigger a "depolarizing" process?

MM: There may be a tipping point in polarization, at which the process cannot be reversed. The Civil War in the US only ended with the defeat of one side and unprecedented loss of life and human suffering. I don't believe we have reached the tipping point in any of the countries

currently experiencing increasingly wide political and cultural divisions. However, there is an important danger sign. Historically, a common threat such as a foreign invasion or an economic collapse would mobilize a unified response. Yet in the U.S. and to a lesser extent in Europe, the response to the current pandemic has instead opened up yet another front in the culture war. That is a very bad sign of things to come. If the upcoming U.S. election is not decided by the voters because of allegations of voting fraud, that could be the tipping point, even if the allegations are baseless.

YS: When using a computational approach to examine human social interaction, to what extent do you consider the influence of cultural difference? For example, people in the collectivist societies of East Asia may have different rules of social interaction. To what extent does cultural sensitivity matter when researchers from different regions are examining and explaining a social issue in a computational way?

MM: Cultural differences usually matter for all methods and models in social science, along with gender and ethnic differences, cohort effects, and so on. An important caveat is that survey data usually allow us to measure gender, ethnicity, and cohort, but it is often impractical to implement a survey across cultures. This is also true of traditional lab-based RCT experiments as well. As a result, survey-based and lab-based research has often neglected cross-cultural analyses. Online data from global social media and other platforms now make it possible to study behavior across cultures. Early on, Scott Golder and I used Twitter data to track diurnal rhythms across cultures. Working with Spotify researchers, Minsu Park and I recently revisited that question using data on emotional attributes of global music downloads. Minsu and I had previously used YouTube data to study cross-cultural video consumption. Our lab also used Twitter data to test Huntington's controversial cross-cultural theory of the "clash of civilizations." And Patrick Park, Josh Blumenstock, and I used global network data to reveal "social wormholes"—high-bandwidth shortcuts that span global social networks. These studies would have been unthinkable back in the 1980s when I began my academic career.

YS: Since online interaction is already deeply woven into the people's daily lives, social science research is to some extent shifting from the offline to the online world. How do you see the relationship between online and offline interaction? How can we build a bridge between online and offline interaction in theory and in observation?

MM: Increasingly we are seeing studies that integrate data from online and offline sources. For example, David Lazer's team has linked social media data with voting records. There is also an explosion in the use of crowdsourcing sites to recruit participants for RCT experiments involving behavior that does not easily fit into either the on-line or off-line categories. Our study with Spotify tracked the music people stream. Is listening to music on-line or off-line behavior? Our lab has also used phone logs to study the structure of social networks. Are phone calls online or offline? Going back to the 19th century, behavioral scientists have long distinguished between face-to-face and device-mediated interaction, whether the device is Facebook or a train or ship carrying mail. Of course, there are vast differences in these technologies, not just the scale but also lag times in the interaction, whether the interactions are dyadic, whether the interactants know each other independently, and so on. Nevertheless, I find it amusing when people make a distinction between online interaction and "the real world," as if people experience a metaphysical transformation when they pull their phone out of their pocket.

YS: You have asserted that "a widely shared analytical toolkit of models and methods for middle-range theorizing" is more important than "a universally applicable theoretical framework" in current research. Does this mean we should be digging out more emergent theories from the masses of digital data now available and applying cross-cutting analytical tools before turning to the construction of macro-level theories?

MM: A grand theory of human interaction could put social and behavioral science on a much stronger foundation, but only if the theory is empirically valid. Previous attempts to construct general theories (e.g. Marxist, Freudian, behaviorist, and utility maximization) have not always turned out well, although some have proved more useful than

others. Middle-range theorizing is more like putting together a vast jigsaw puzzle, but with no guarantee that there is a unique solution or even a finite number of puzzle pieces. Accordingly most social scientists have concluded that progress is more likely if we focus on improving our ability to measure and our ability to make causal inferences as we test hypotheses motivated by observed patterns that seem puzzling against the backdrop of current knowledge. That approach tends to produce mostly incremental progress that depends more on advances in our abilities to measure and model than on our ability to tell a compelling story about the picture we will begin to see when the jigsaw puzzle is closer to completion. So far, it appears that each time we figure out where to put a piece in the puzzle, we discover that there are actually more pieces to the puzzle than we realized. Maybe the pieces will start to fall into place one day, but it does not appear that we are far enough along to adopt a grand theoretical approach based on that assumption.

YS: In recent years, as various computational social science research methods have emerged, which of the new methods most excite you? What are the methods you regard as the most promising for future research? In your opinion, what methods do researchers still badly need?

MM: If you ask most researchers in computational social science, machine learning will be near the top of most lists. Also word embedding, along with other tools for analyzing text as data. And image processing. These methods are regularly taught in computer and information science, and rarely in the social sciences, and that leads me to wonder if computational social science might increasingly shift location from the technical wing of the social sciences to the social science wing of the computational sciences. I have observed defensiveness among sociologists fearful of disciplinary colonization, but that defensiveness may mean that sociology misses an historic opportunity to lead the way forward.

YS: How do you see the relationship between social simulation using computational methods and real-world observation?

MM: Thirty years ago, agent based modeling transformed computer simulation by replacing a "model of the population" with a "population of models." These micro-level simulations made it possible to explore the macro-level implications of behavioral and structural assumptions, rather than relying on theoretically informed intuition. The problem with relying on intuition is that many of the outcomes of complex systems are highly counter-intuitive. However, the early efforts, including my own, were really toy societies. They were much closer to the RCT approach in experimental science than to multivariate linear mixed models in the world of survey research and government statistics. A well-designed RCT experiment is also a toy, but it can be a powerful tool nonetheless. If an experiment becomes so complicated that the results can only be analyzed using multivariate models, the ability to make causal inference is likely to be badly compromised. That often applies as well to simulation models. We should not try to make simulations more "realistic" as an end in itself. Having said that, I am also very impressed by two developments in computational modeling: enormous strides in empirical calibration, and breakthroughs in the ability to harness the power of cloud computing to systematically explore a vast parameter space. Some of the most impressive examples are the breakthroughs at the Research Triangle Institute in using computational models to create synthetic populations by linking multiple datasets. Look at their successes in modeling epidemics on synthetic populations and then fast forward another ten years! Duncan Watts once thought data from social media would be our "space telescope"—the breakthrough tool that lets us see things we could never see before—an analogy that Duncan has since walked back. Maybe the space telescope will instead turn out to be synthetic populations.

YS: Training in computational social science must develop both computational literacy and social literacy. Students drawn to the social sciences tend to have good sensitivity and insight toward social issues but may lack computational aptitude. Students who can easily master computational skills may not ask appropriate questions about social issues. How do you see training resolving these difficulties for the various students interested in computational social science?

MM: I have a joint appointment in the departments of Sociology and Information Science, and the Social Dynamics Lab attracts PhD students from both departments, as well as from computer science and applied math. In the course of working together on research projects, the students teach each other the skills and theoretical frameworks needed for the research. Over time, it becomes increasingly difficult to know which student came from which disciplinary background. These cross-disciplinary teams also show that it is too simple to say that computer scientists have methodological skills but no theory and social scientists have theory but lack computational skills. For example, computer scientists often have strong theoretical backgrounds in processes of network diffusion, and social scientists have skills in causal inference with observational data. So the opportunities for synergy are enormous.

YS: Looking back, what attracted you and drove you to focus on the computational approach to the social sciences, especially when this approach was receiving little attention?

MM: I began using computational models in 1989, inspired by Robert Axelrod's evolutionary model of the prisoner's dilemma. I made a simple change to his model, replacing "imitation of the fittest" with "win stay, lose switch," also known as reinforcement learning. These early models were very simple toy models, yet they benefited from breakthroughs in desktop computing that made computational modeling widely accessible and showed how computation can be used to explore the logical implications of a set of assumptions about human behavior and social interaction.

YS: This is the first time members of the Chinese research community have had an opportunity to hear from such a giant in the field of computational social science and sociology. Your article "Digital Footprints: Opportunities and Challenges for Online Social Research" is viewed as a classic. Do you have anything in particular to share with them?

MM: Thank you for your interest and for training a new generation of computational social scientists! The opportunity for international collaboration is as important as bridges across disciplinary boundaries.

YS: What one piece of advice would you give them about how to succeed in computational social science?

MM: Don't pick questions based on "no one has looked at this before" because it is likely that no one looked before either because the question is not useful to answer or it is impossible to answer. Instead, ask interesting and important questions, no matter how many others have already taken a shot. There will always be something new to explore—that is why it is called "science" and not "religion." (Religion is about faith, science is about skepticism. Answers come through skepticism, not through faith. Solace comes through faith, not through skepticism.)

Selected Works by Michael Macy

Please refer to the end of the Chinese version of the dialogue for Michael Macy's selected works.