Global Village or Virtual Balkans? Evolution and Performance of Scientific Collaboration in the Information Age

Xinlin Yao
School of Economics and Management, Nanjing University of Science and Technology, 200 Xiaoling Wei Street, Nanjing, 210094, China

Cheng Zhang
School of Management, Fudan University, 670 Guoshun Road, Shanghai, 200433, China.
E-mail: zhangche@fudan.edu.cn

Zhe Qu
School of Management, Fudan University, 670 Guoshun Road, Shanghai, 200433, China

Bernard C.Y. Tan
Department of Information Systems and Analytics, National University of Singapore, Computing 2, 15 Computing Drive, Singapore, 117418, Singapore

Scientific collaboration is essential and almost imperative in modern science. However, collaboration may be difficult to achieve because of 2 major barriers: geographic distance and social divides. It is predicted that the advancement of information communication technologies (ICTs) will bring a puzzled conclusion for collaboration in the scientific community: the "Global Village" trend with significantly increased physical distance among collaborated scientists and the "Virtual Balkans" trend with significantly increased social stratification among collaborated scientists. The results of this study reveal that the scientific community evolves towards the Global Village generally on both the geographic and social dimension, but with variations in term of collaboration patterns. The influence of such collaboration patterns on research performance (that is, productivity and impact), however, is asymmetric to each side of collaborators. When researchers from top-tier and general-tier institutions collaborate, researchers from top-tier institutions face a decrease in research productivity and impact, whereas researchers from general-tier institutions increase in research productivity and impact. Furthermore, the development of ICTs plays an important role in shaping the evolving trends and moderating effects of collaboration patterns. Our findings provide a comprehensive understanding of scientific collaboration in the geographic, social, and technological aspect.

Additional Supporting Information may be found in the online version of this article.

Received January 29, 2018; revised April 5, 2019; accepted April 14, 2019

Introduction

Scientific collaboration has become imperative in many fields. Over 90% of publications in science, technology, and engineering are collaborative (Bozeman & Boardman, 2014). Scientific collaboration helps researchers to access and integrate knowledge, skills, materials, and other necessary resources for tacking scientific problems, which also enhances research productivity, breeds innovation, and carves successful career paths for researchers (Gazni & Didegah, 2011; Gazni, Sugimoto, & Didegah, 2012; Zhang, Bu, Ding, & Xu, 2018). For instance, Larivière et al. (2016) observed that collaboration in science has been increasing over the past century, and that collaboration was positively correlated with research quality.

Apart from these benefits of collaboration, it is worth noting that scientific collaboration is not easy to achieve because researchers need to transcend various barriers that hinder communication and collaboration, such as geographic distance and social divides. The development and spread of information communication technologies (ICTs) seem to provide an effective solution through lowering communication cost, enhancing communication quality, and improving information transparence, leading the scientific community moves towards a "Global Village" (Boudreau et al., 2017; Ding, Levin, Stephan, & Winkler, 2010; Srinivasan, 2017). However, some scholars made the opposite prediction about the effects of ICTs in science. They believed that, empowered by search and filtering ICTs, researchers are more able to select and optimize collaborators for perceived relevance and reputation, leading to more fragmented intellectual and social communications, thus forming a "Virtual Balkan" in the scientific community (Van Alstyne & Brynjolfsson, 1996, Van Alstyne & Brynjolfsson, 2005; Rosenblat & Mobius, 2004). For instance, scientists in a lower-rank institution may find it difficult to collaborate with researchers in institutions with higher rank even if ICTs lowered the communication cost, because researchers may prefer scholars with a similar background to work together due to their homophily preferences (Zhang et al., 2018).

Changes in scientific collaboration may determine research performance, such as productivity and impact. Understanding the evolution of scientific collaboration is therefore critical, especially when determining how to enhance research performance and whom to collaborate with. Scientific collaboration is characterized by being influenced by spatial, cognitive, social, and technological factors (Hara, Solomon, Kim, & Sonnenwald, 2003). Although previous studies have provided some insights on scientific collaboration, they mainly focused on one factor or limited disciplines. Therefore, this study aimed to provide a deeper understanding of the intertwined nature of scientific collaboration on geographic, social, and technological dimension through addressing the following three questions:

1. What are the evolving trends of scientific collaboration in terms of social stratification and geographic separation?
2. What are the effects of cross-tier collaboration on research productivity and impact?
3. How does the development of ICTs influence the evolving trend and the effects of cross-tier collaboration?

This study uses the publications of researchers affiliated with 607 institutions in the United States and citations from the Web of Science (WoS) database to answer these questions. The rest of this article is organized as follows. The Theoretical Background section summarizes related works on scientific collaboration and the effects of ICTs in science, which lay the theoretical foundations for our hypotheses. In Methods, we describe the data and measures in this study. The Results section then reports the results and robustness analysis. Finally, we draw the conclusions and discuss some limitations and future research directions.

Theoretical Background and Hypotheses

Scientific Collaboration: Consequences and Barriers

Scientific collaboration, the process where two or more individuals/institutions interact to share knowledge, experience, and resources for creating and sharing understanding and achieving common goals in scientific activities, has been a major form in knowledge production (Gazni et al., 2012; Zhang et al., 2018). Three major trends have made scientific collaboration dominant and imperative. One is the growing specialization in science that calls for integrating skills, knowledge, and abilities from various disciplines to innovate (Hara et al., 2003; Stevens & Campion, 1994). The other is the increasing scale and complexity of scientific problems, which makes collaboration crucial in the context of "Big Science" (Galison, Hevly, & Weinberg, 1992; Pennington, Simpson, McConnell, Fair, & Baker, 2013). The third trend is the increasing need for sharing equipment, facilities, or other resources from collaboration to tackle complex research problems (Kling & McKim, 2000).

Apart from these trends, many previous studies support that scientific collaboration has positive impacts on individuals, institutions, and countries (Bu et al., 2018; Li & Li, 2015). The fundamental need for scientific collaboration is accessing skills, knowledge, and other necessary resources for research (Beaver, 2001; Birnholtz, 2007). Although there are findings against whether collaboration has a positive affect on research productivity (Abramo, D'Angelo, & Di Costa, 2009; Lee & Bozeman, 2005), many studies support that scientific collaboration could enhance research impact and breed innovation (Wuchty, Jones, & Uzzi, 2007; Ding, 2011; Bozeman & Boardman, 2014; Larivie`re et al., 2015). Furthermore, many scholars believe that scientific collaboration is helpful in building reputation and establishing a successful career path, which significantly motivates scholars who are seeking being visible and recognized (Amjad et al., 2017; Bu et al., 2018; Zhang et al., 2018).

However, the benefits of collaboration may not be easily obtained because of various barriers that hinder scientific collaboration. Besides the difficulties of integrating diverse expertise and knowledge in research (Schaltegger, Beckmann, & Hansen, 2013), geographic barriers and social barriers are two major forces that strain and end scientific collaboration.

The role of geography in scientific research has attracted considerable attention in previous studies. Many of them focused on investigating the spatial patterns of scientific collaboration and the effects of geographic distance on collaborating (Frenken, Hardeman, & Hoekman, 2009). For instance, Ponds, Van Oort, and Frenken (2007) found that geographic proximity was important for collaboration between institutions with different backgrounds. Along with the advancement of ICTs, some have declared an end to "the tyranny of distance" (Cairncross, 1997). However, some studies suggested that geographic distances still mattered. For instance, Hoekman, Frenken, and Tijssen (2010) examined the changing spatial patterns of collaboration within Europe and found that geographic distance was still related to collaboration. Similarly, Sidone, Haddad, and Mena-Chalco (2017) found that, although collaboration showed a trend of geographic deconcentration in Brazil, distance still significantly determined the intensity of knowledge flow in collaboration networks. In terms of geographic barriers, the migration and mobility of scientists could be another way to transcend establishing ties. Previous studies have shown that the international migration and mobility of researchers has

DOI: 10.1002/asi

significant influence on their productivity and scientific collaboration (Jonkers & Tijssen, 2008; Marmolejo-Leyva, Perez-Angon, & Russell, 2015). Nevertheless, a recent study suggested that the mobility of scientists was vulnerable to the immigration policies and the political environment, and investigated the collaboration and mobility of countries in the "travel bans" implemented by U.S. President Trump in 2017 (Chinchilla-Rodríguez, Bu, Robinson-García, Costas, & Sugimoto, 2018). Therefore, we expect that geographic barriers might still play important roles in scientific collaboration and it is necessary to know the geographic pattern of scientific collaboration for scholars and policy makers.

Despite the institutional environment, social factors that influence scientific collaboration are growingly emphasized, such as personal differences (Birnholtz, 2007), linguistic or cultural divides (Freeman & Huang, 2015), and academic rank or institutional barriers (Jones, Wuchty, & Uzzi, 2008; Li, Li, Liu, & Liu, 2018; Thijs & Glänzel, 2010) (see Hall et al., 2018, for a detailed review). It is worth noting that institutions are stratified and are not weighted equally in the scientific collaboration network (Chang & Huang, 2013); thus, choosing whom to collaborate with is far from a random selection and would profoundly determine the different performances of scientific collaboration. From the perspective of social networks, homophily, transitivity, and preferential attachment are three major mechanisms that influence the generation of scientific collaboration (Zhang et al., 2018), which also suggests the underlying rationales of social divides that would hinder scientific collaborations. For instance, homophily refers to people who tend to build connections with similar others, which would make researchers form connections with those who have similar educational backgrounds, research interests, or even academic ranking as themselves (Freeman & Huang, 2014; Sie, Drachsler, Bitter-Rijpkema, & Sloep, 2012), thus leading to more homogenous collaborations. Transitivity, also known as the path-dependent effect, offers researchers a direction to select potential collaborators rather than random selection, which lowers the searching and matching cost but also creates barriers to other potential scholars (Schilling & Phelps, 2007; Zhang et al., 2018). Furthermore, preferential attachment reflects an important motivation for collaborating with notable scholars to lift up one's own productivity and visibility (Milojević, 2010; Zhang et al., 2018), which is also known as the Matthew Effect, where notable scholars are more attractive and more likely to form new collaborations. Taken together, we expect that those social mechanisms would play significant roles in scientific collaborations, but few studies presented the evolving trend of the extent of social stratification in scientific collaborations.

Effects of ICTs on Scientific Collaboration

As Hara et al. (2003) said, scientific collaboration was complex and characterized by being shaped by spatial, cognitive, social, and technological factors. The role of ICTs in scientific collaboration is at least as important as the geography and social factors.

However, the effects of ICTs in science seem to be inconclusive, because two opposing perspectives exist. One is that the advancement of ICTs inevitably removes the geographic barriers in scientific collaboration and widens access to distant scholars, thus leading to a "Global Village" in science discipline (Srinivasan, 2017). Besides lowering the cost of communication, Ding et al. (2010) suggested that ICTs also provided access to knowledge and materials as research input, which could enhance productivity, collaboration, and show a demarcating effect for scholars at the margins of a profession. On the contrary, some scholars argued that the development of ICTs could lead to the fragmented intellectual and social interactions, implying a "Virtual Balkan" in science (Rosenblat & Mobius, 2004; Van Alstyne & Brynjolfsson, 1996, 2005). Two reasons may contribute to fragmented academic communications: one is that the explosion of online information makes explicit searching and filtering necessary (Van Alstyne & Brynjolfsson, 1996). The other is that the searching and filtering process is dependent on preferences, resulting in the selective exposure to information and potential collaborators based on common interests, status, academic reputations, and other nongeographic factors (Van Alstyne & Brynjolfsson, 1996, 2005).

Actually, these two perspectives predict the effects of ICTs in terms of geographic barriers and social divides in scientific collaborations, respectively. The advancement of ICTs would help scholars transcend geographic barriers due to the lowered communication cost and increased accessibility. However, ICTs also afford scholars the ability to effectively and efficiently filter and search for their interested potential collaborators, which may significantly strengthen the effects of homophily, transitivity, and preferential attachment (Zhang et al., 2018) in building connections, thus motivating fragmented interactions among scholars. Although several studies have focused on one aspect, few studies have simultaneously discussed the effects of ICTs on both geographic and social factors in scientific collaborations.

To address these concerns, we aim to provide a deeper understanding of the intertwined nature of geographic, social, and technological factors in scientific collaboration in terms of both the natural and social sciences. More specifically, based on the existing literature, we propose and test the following hypotheses in this study:

> Hypothesis 1 (H1): The development of ICTs plays a significant role in influencing the geographic separation and social stratification in scientific collaboration.

> Hypothesis 2 (H2): For general-cited tier institutions, collaborating with top-cited tier institutions has a positive effect on their research performance.

> Hypothesis 3 (H3): The development of ICTs plays a significant moderating role in influencing the effects of cross-tier collaboration on research performance.

Methods

Data and Preprocessing

Our data set covered 4,246,653 articles' publication information in 14.3 thousand journals in the natural and social sciences from 1980 to 2009 provided by WoS, including 4,079,829 articles in the natural sciences and 166,824 articles in the social sciences (see S6 in Supplementary Materials for more details). The main field of each article was determined by the major subject category assigned by WoS. Among 260 subject categories, 175 subjects were coded in the natural sciences, 55 subjects were coded in the social sciences, and the rest of 30 subjects in Arts and Humanities were not included in this study (see S7 in for Supplementary Materials for the detailed list).

The raw data were stored in ASCII stream text files, using tag characters to indicate different fields accordingly. In general, every file contained three basic types of information: file information, source issue information, and source item information. However, different from databases organized by entity relationship, the records in WoS require great manual work to distinguish and reconstruct the relationships between data. For example, the authors' names of an article are stored in a field that needs further processing to split them. Additionally, some fields in the database have ambiguous information that may affect the results. For example, in the institutions' location field, both USA and America refer to the USA. We conducted further manual processing to recognize them as one country. The list of variants for institutions provided by WoS[1] was also used to help identify different name variants for institutions. Similar things happened with the authors' name and institutions' name. In our processing, we use other fields' information like e-mail address, zip code, and other recorded information to correct them. For instance, if two records only differed in the middle initials and other fields were consistent, they were coded as being the same person because the same author might use slightly different names like "David Robert" or "David J. Robert."

Measures

Scientific collaboration. Coauthorship is a standard approach of measuring scientific collaboration, although some scholars argued that it did not depict the whole picture of the process of scientific collaboration (Laudel, 2002; Lundberg, Tomson, Lundkvist, & Brommels, 2006). In this study, we used coauthorship in the publication to measure scientific collaboration, as Franceschet and Costantini (2010) noted that "coauthorship in publications is widely considered as a reliable proxy for scientific collaboration,", (p. 541) which is consistent and comparable with previous studies (Gazni et al., 2012; Larivière, Gingras, Sugimoto, & Tsou, 2015). Furthermore, we merely focus on the interinstitution collaboration in the United States to control variances from other social factors that influence scientific collaboration, such as differences in culture, languages, and national policies.

[1] http://ips.clarivate.com//m/pdfs/UnifyingNameVariants.pdf

Institution tier of total citations. To investigate the patterns of scientific collaboration, we first ranked the 607 U.S. institutions with the total citation counts of all their publications to classify them into the top-cited tier and the general-cited tier. In this ranking approach, we merely considered within-institution publications so that the rankings are independent of each institution's performance in interinstitution collaborations. In the main analysis, the top 30 U.S. institutions in our list are classified as the top-cited tier, and the rest of the institutions are classified as the general-cited tier. We also employed an alternative classification that regards the top 50 U.S. institutions as the top-cited tier in the robustness checks (see S1 in Supplementary Materials).

Social stratification and collaboration patterns. We then apply the index of balkanization in Van Alstyne and Brynjolfsson (1996) (Van Alstyne & Brynjolfsson, 2005) as below to measure the social stratification of scientific collaborations,

$$\text{Index of Balkanization} = 1-\mu_a = 1-\frac{\sum_{t\in T}\sum_{s\neq t}\frac{\left\|M(t)\bigcap M(S)\right\|^2}{\|M(t)\|\|M(s)\|}}{T(T-1)}$$

in which authors of collaboration are divided into T subsets. $M(t)$ is the t-th subset and $\|M(t)\|$ is the size of the t-th subset, that is, the number of scholars in it. It is worth noting that μ_a measures the overlap of contact between groups, implying that the more diverse an agents' interactions, the more he or she was integrated with different groups (Donath, 1995). Thus, per interpreting balkanization as the fragmented intellectual and social interactions in science, an index of balkanization could be naturally defined as 1-μ_a (Van Alstyne & Brynjolfsson, 2005). The index of balkanization ranges from 0 to 1, with a lower value indicating more collaboration between researchers from institutions in different tiers.

We obtained three types of collaboration patterns based on the composition of the coauthors' institution tiers (that is, top or general): top-tier collaboration, cross-tier collaboration, and general-tier collaboration. Top (or general) -tier collaboration is determined from coauthored articles involving only researchers from the top (or general) -tier institutions. Cross-tier collaboration is determined from coauthored articles involving researchers from both top-tier and general-tier institutions.

We used affiliation information of authors to count the number of authors in each institution every year, then obtained the number of authors in each tier every year, and the number of collaborators within/cross-tier. Therefore, we calculate the index of balkanization every year by treating each tier as a subset to measure the social stratification extent.

Geographic separation. Using authors' affiliation information, we calculated the geographic distance in collaborations to reflect the extent of geographic separation. The location (latitude and longitude) of each author (researcher) was determined based on the zip code of her/his institution. Using the Haversine formula:

$$2r.\mathrm{atan2}\left(\sqrt{\sin^2\left(\frac{\Delta\varphi}{2}\right)+\cos(\varphi 1).\cos(\varphi 2).\sin^2\left(\frac{\Delta\lambda}{2}\right)},\sqrt{1-\left[\sin^2\left(\frac{\Delta\varphi}{2}\right)+\cos(\varphi 1).\cos(\varphi 2).\sin^2\left(\frac{\Delta\lambda}{2}\right)\right]}\right)$$

where φ is latitude, λ is longitude, r is the earth's radius (mean radius = 6,371 km), $\Delta\varphi$ is the difference between two latitude points, atan2 is the arctangent of the quotient of its arguments (x, y); we calculated the geographic distance between every pair of collaborators. Two coauthors who are from the same institution would have a collaboration distance of zero. If an article has more than two coauthors, their dyadic distances were averaged to get the collaboration distance.

Research performance. We employed two metrics to measure research performance: research productivity and research impact. Research productivity of researchers is measured by their number of publications. Publications with multiple authors are attributed in equal fractions to all coauthors. For example, if an article has three authors, each author has a 1/3 count of this publication. Research productivity of institutions is computed by summing the research productivity of all researchers within the same institution.

Research impact is measured by the number of citations of the publication in the following years. It is important to choose an appropriate time window to balance the accuracy of measurement and the loss of data. Through analyzing all WoS journal publications, Wang (2013) found that the cumulative citation number is positively correlated with the total citations, and the correlation increased rapidly in the early years and then slowly until reaching one. Wang (2013) showed that the inflection point of the increasing rate of correlation was around a 5-year citation number, in which the Spearman correlation between the 5-year citation number and total citations was 0.871. Therefore, we employed a 5-year citation count to measure the research impact and we can determine the total citations for all articles by the end of 2004 in our data set. Likewise, research impact of institutions was computed by summing the research impact of all researchers within the same institution.

Development of ICTs. We developed three stages of the Information Age to investigate how the evolving trends change along with the development of ICTs from 1980 to 2009. The stages were developed based on the development of ICTs that especially affect communication and collaboration among researchers, specifically: (i) Pre-Information Age (1980–1990): In this stage, the Internet was preliminary but many early applications were introduced in academia, such as e-mail, file transfer, and remote login (Leiner et al., 1997), which laid a crucial foundation for the following development. (ii) Web 1.0 (1991–1999): This stage was characterized by the spread of the World Wide Web (WWW) and e-mail, which allows Internet users to search and browse more online information and more easily access others. The development of digital libraries and online catalog in this stage were important for scientific research as well (Liang, 2015). (iii) Web 2.0 (2000–2009): This stage was featured by the prevalence of blogs, wikis, instant messaging tools, and other social network services, which support not only communication but also collaboration (Yao, Li, Zhang, & Ling, 2017).

In the robustness check, we employed the growth of the Internet hosts, an important component of Internet infrastructure, as an alternative measure of ICT's development.

Results

Trends of Social Stratification in Scientific Collaboration

The extent of social stratification in scientific collaboration was measured with the index of balkanization. A lower balkanization index indicated a higher extent of cross-tier collaboration. As presented in Figure 1, the balkanization index decreased from 0.93 in 1980 to 0.77 in 2009 in the natural sciences and decreased from 0.90 in 1980 to 0.79 in 2009 in the social sciences.

Correlation analysis results showed that this decreasing trend was significant along with years in both the natural sciences ($r_{Pearson} = -0.994$, $p < .000$; $r_{Spearman} = -0.997$, $p < .000$) and the social sciences ($r_{Pearson} = -0.966$, $p < .000$; $r_{Spearman} = -0.975$, $p < .000$), indicating an increasing trend of cross-tier collaborations. In contrast to the earlier prediction, the research communities in the natural sciences and the social sciences appeared to be more integrated rather than being fragmented or moving towards the Virtual Balkans.

We employed the Kruskal–Wallis Test to examine how balkanization indices changed across three stages of the Information Age. As presented in Panel A of Table 1, the decreasing rank means from the pre-Information Age to Web 2.0 in both the natural and social sciences were consistent with Figure 1, and the Kruskal–Wallis Test showed that the balkanization indices significantly differed in three stages of Information Age in both natural and social sciences.

We further reported the results of pairwise comparison in Panel B. For both natural and social sciences, the results showed that balkanization indices in Web 2.0 were significantly lower than that in the pre-Information Age ($p < .001$); however, the difference between Web 1.0 and the pre-Information Age was only significant ($p < .05$) in natural sciences, not in social sciences. Generally, these results indicated that a higher extent of cross-tier collaboration along with advancement of ICTs, especially in Web 2.0.

Trends of Geographic Separation in Scientific Collaboration

We then examined the evolution of geographic separation in scientific collaboration from 1980 to 2009, which is measured as the geographical distance between the institutions

DOI: 10.1002/asi

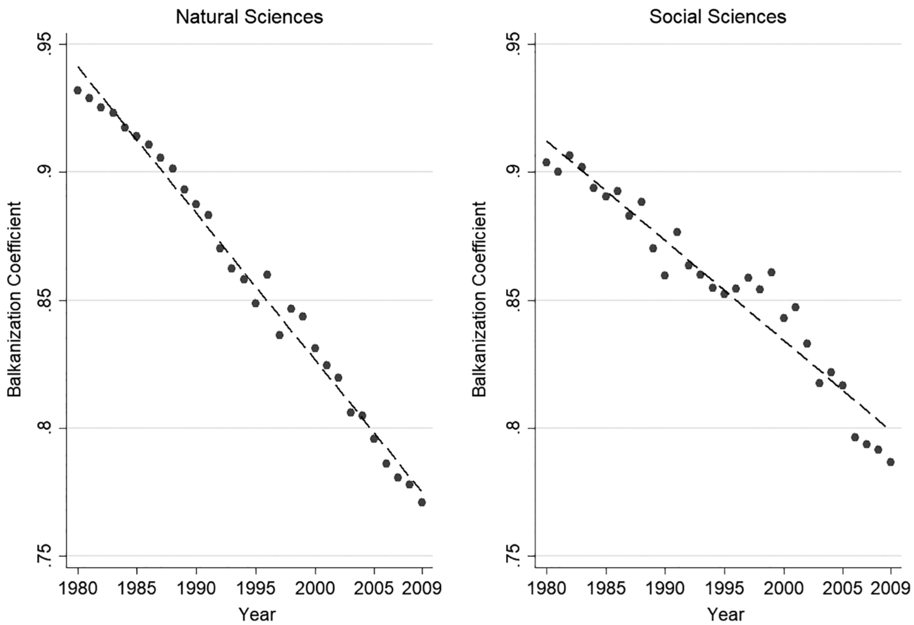

FIG. 1. Changes in balkanization coefficient of scientific collaboration from 1980 to 2009.

TABLE 1. Comparison of balkanization coefficient across stages of the information age.

	Natural sciences		Social sciences	
Panel A	Rank mean	Kruskal–Wallis test	Rank mean	Kruskal–Wallis test
Pre-Infor Age (1980–1990)	25.00	25.742***	24.55	24.517***
Web 1.0 (1991–1999)	15.00		15.56	
Web 2.0 (2000–2008)	5.50		5.50	
Panel B	Rank mean difference	Kruskal–Wallis test	Rank mean difference	Kruskal–Wallis test
Pre-Infor Age vs. Web 1.0	−10.00	**	−8.99	n.s.
Pre-Infor Age vs. Web 2.0	−19.50	***	−19.05	***
Web 1.0 vs. Web 2.0	−9.50	n.s.	−10.06	*

*$p < .05$.
**$p < .01$.
***$p < .001$, two-tailed.

of coauthors of publications. We found that collaboration distance has increased from 43.8 km in 1980 to 76.5 km in 2009 in natural sciences and from 68.7 km in 1980 to 119.2 km in 2009 in social sciences. However, as presented in Figure 2, changes of geographic distance in collaboration showed different trends among different collaboration patterns, that is, top-tier collaboration, general-tier collaboration, and cross-tier collaboration.

We conducted the correlation analysis to examine the statistical significance of these evolving trends. As reported in Table 2, the overall increasing trend of geographic distance was significant because of the positive correlations in both natural ($r_{Pearson} = 0.447$, $p < .05$; $r_{Spearman} = 0.477$, $p < .01$) and social sciences ($r_{Pearson} = 0.441$, $p < .05$; $r_{Spearman} = 0.501$, $p < .01$).

However, the evolving trend differed among different collaboration patterns. In social sciences, geographic distance of top-tier collaboration showed an increasing trend, while a significantly decreasing trend ($r_{Pearson} = -0.669$, $p < .001$; $r_{Spearman} = -0.738$, $p < .001$) was observed in natural sciences, indicating researchers were seeking collaborators who are in the similar top-tier institutions and nearby. The evolving trend of cross-tier collaboration also showed opposite direction in natural sciences (increasing) and social sciences (decreasing). Geographic distance of general-tier collaboration showed an increasing trend in both natural and social sciences.

With the advancement of ICTs, it was expected that geographic distance shall no longer be a barrier. However, the evolving trend of collaboration distance seemed to be inconclusive because of the opposite evolving directions presenting in different collaboration patterns. Therefore, we further examined how the change of collaboration distance would differ between three stages of the Information Age. We conducted a Kruskal–Wallis Test and pairwise comparisons and reported the results in Table 3.

From Table 3, per the Kruskal–Wallis Test, we first observed that differences of geographic distance significantly differ in three stages of the Information Age in all collaboration patterns of both the natural and social sciences. The pairwise comparison shows that differences of geographic distance were not significant between the pre-Information Age and Web 1.0 in all collaboration patterns, and in both the natural and social sciences, which might indicate that

DOI: 10.1002/asi

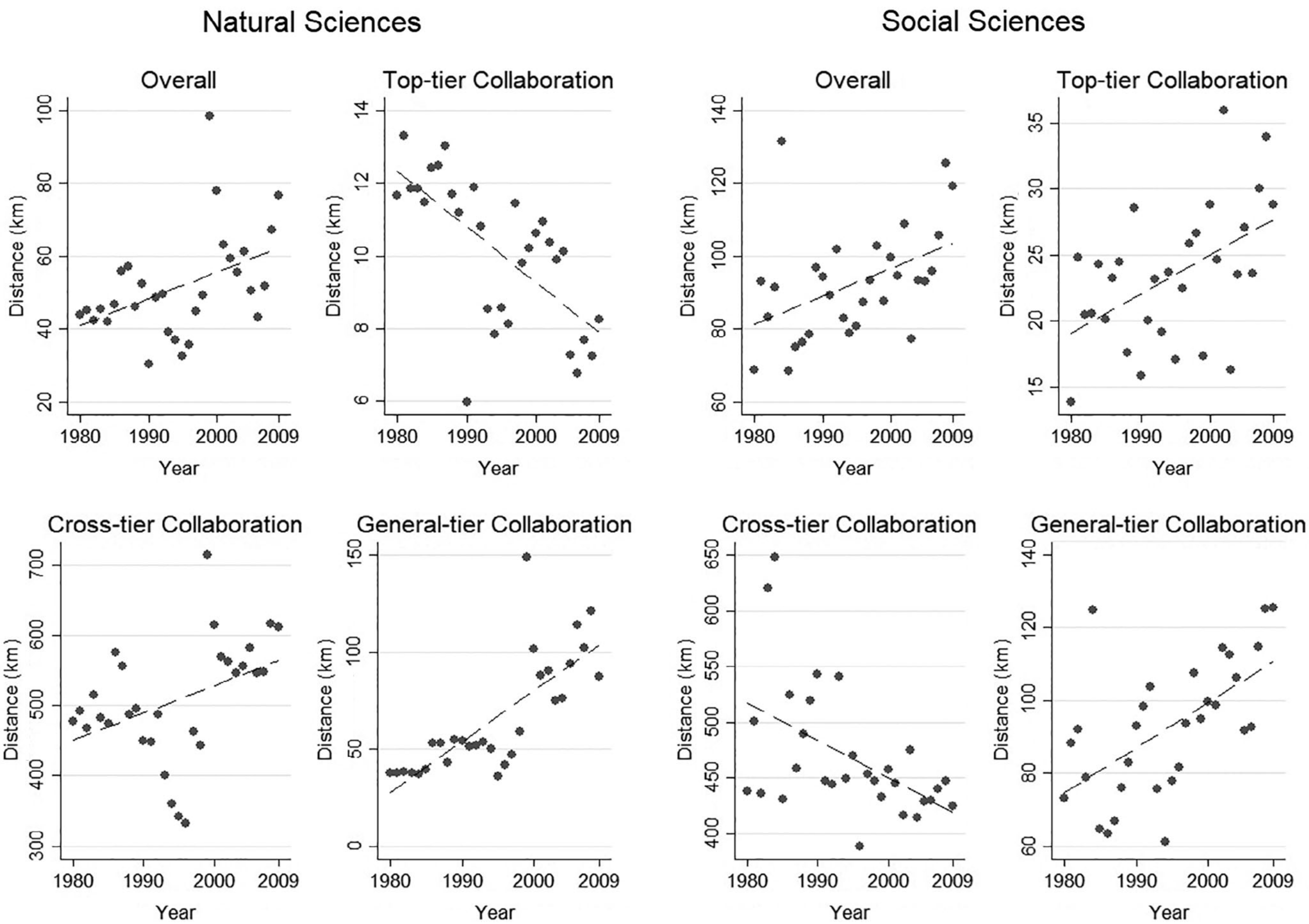

FIG. 2. Changes in geographic distance of scientific collaboration from 1980 to 2009.

TABLE 2. Correlation analysis of geographic distance of scientific collaboration and years.

	Natural sciences				Social sciences			
	Pearson	*p*	Spearman	*p*	Pearson	*p*	Spearman	*p*
Overall	0.447	*	0.477	**	0.441	*	0.501	**
Top-tier Collaboration	−0.669	***	−0.738	***	0.498	**	0.469	**
Cross-tier Collaboration	0.401	*	0.449	*	−0.507	**	−0.495	**
General-tier collaboration	0.775	***	0.811	***	0.588	**	0.614	***

*$p < .05$.
**$p < .01$.
***$p < .001$, two-tailed.

ICTs in Web 1.0 have not significantly impacted the geographic separation in scientific collaboration.

It was worth noting that Web 2.0 made a difference. Overall, we found that collaboration distance significantly increased in Web 2.0 compared with the pre-Information Age in both natural (*RankMean Diff.* = 9.95, $p < .05$) and social sciences (*RankMean Diff.* = 9.75, $p < .05$). General-tier collaboration in both the natural and social sciences showed a consistent increasing trend, suggesting that ICTs in Web 2.0 help those researchers collaborate with distant counterparts in similar general-cited institutions. However, we also found some decreasing trend of geographic distance between Web 2.0 and the pre-Information Age. For instance, the decreasing distances of top-tier collaboration in the natural sciences (*RankMean Diff.* = −13.04, $p < .01$) indicated that researchers in top-tier institution seek to collaborate with nearby researchers who are in top-tier institutions rather than the distant ones. Cross-tier collaboration in the social sciences showed a similar decreasing trend of collaboration distances. These decreasing trends might result from using ICTs to select and filter potential collaborators rather than bridging the geographic gaps.

DOI: 10.1002/asi

TABLE 3. Comparison of geographic distance in scientific collaboration across stages of the information age.

Panel A Natural sciences	Kruskal–Wallis test	Pre-Infor Age vs. Web 1.0		Pre-Infor Age vs. Web 2.0		Web 1.0 vs. Web 2.0	
		Rank mean difference	p	Rank mean difference	p	Rank mean difference	p
Overall	9.578**	−1.21	n.s.	+9.95	*	+11.17	*
Top-tier Collaboration	12.625**	−9.30	n.s.	−13.04	**	−3.73	n.s.
Cross-tier Collaboration	15.212**	−6.27	n.s.	+9.33	*	+15.60	***
General-tier collaboration	16.827***	+4.24	n.s.	+15.41	***	+11.71	*
Panel B Social sciences	**Kruskal–Wallis test**	**Pre-Infor Age vs. Web 1.0**		**Pre-Infor Age vs. Web 2.0**		**Web 1.0 vs. Web 2.0**	
		Rank mean difference	p	Rank mean difference	p	Rank mean difference	p
Overall	6.739*	+2.66	n.s.	+9.75	*	+7.09.	n.s.
Top-tier collaboration	7.682*	−0.03	n.s.	+9.44	*	+9.47	n.s.
Cross-tier Collaboration	8.117*	−6.22	n.s.	−10.90	*	−4.68	n.s.
General-tier collaboration	11.520**	+3.80	n.s	+12.81	**	+9.01	n.s.

*$p < .05$.
**$p < .01$.
***$p < .001$, two-tailed. +: distances increase, −: distances decrease.

Taken together, these results suggested that researchers from general-tier institutions were leveraging advances in ICTs to collaborate with other distant researchers from general-tier institutions (in both the natural and social sciences) or from top-tier institutions (only in the natural sciences). However, top-tier collaboration in the natural sciences and cross-tier collaboration in the social sciences were becoming more localized.

Effects of Cross-Tier Collaborations on Research Performance

We further examined the effects of collaboration patterns on research performance to answer whom to collaborate with. No matter researchers in top-tier institutions or in general-tier institutions, they can only choose to collaborate with researchers within the same tier (within-tier collaborations) or from the other tier (cross-tier collaborations). Therefore, we focus on the effects of cross-tier collaborations here, and the effects of within-tier collaborations will naturally be the reverse.

In the following sections, we examine the effects of cross-tier collaborations on research productivity and research impact, respectively. We also controlled the number of authors (*AuthorNum*) and number of collaborating institutions (*CollaboratorNum*) in each institution. The effects of ICT's development are captured through two dummy variables first, and further analyzed with the growth of Internet Host (*HostNum*) as a proxy in robustness checks. Table 4 summarizes the studied variables.

TABLE 4. Studied variables and descriptions.

Variable name	Description
Cross-tier CollaborationRatio	The proportion of cross-tier collaboration in total collaboration for each institution
AuthorNum	Number of authors in each institution
CollaboratorNum	Number of collaborating institutions for each institution
Dum_Web1.0	1 = during the Web 1.0 (from 1991 to 1999); 0 = otherwise
Dum_Web2.0	1 = during the Web 2.0 (from 2000 to 2009); 0 = otherwise
Year	Calendar year
HostNum	Number of Internet host in each year

Research productivity. As research productivity is measured by the number of publications, we estimated fixed-effects negative binomial models and they are reported in Table 5. The fixed-effect negative binomial model controlled other unobservable heterogeneities of institutions and the bias due to dispersion of the dependent variable (that is, research productivity of institutions). Independent variables are standardized before generating quadratic and interaction items.

As reported in Table 5, for top-tier institutions, cross-tier collaboration showed a negative effect on research productivity in both the natural sciences ($\beta = -0.271$, $p < .001$) and social sciences ($\beta = -0.338$, $p < .001$). On the contrary, for general-tier institutions, cross-tier collaboration showed a positive effect on research productivity in both the natural sciences ($\beta = 0.210$, $p < .001$) and social sciences ($\beta = 0.329$, $p < .001$). Therefore, researchers from general-tier institutions seemed to be better off by collaborating with researchers from top-tier institutions but worse off by collaborating with their counterparts from general-tier institutions.

We further examined whether these effects would be differential across three stages of ICT's development. We found that the coefficient of the interaction term *Cross-tierCollaborationRatio*Dum_Web1.0* was not significant for both top- and general-tier institutions and in both the natural and social sciences. This was consistent with the collaborating distance not differing between the pre-Information Age and Web 1.0, indicating the structure of collaboration had not significantly changed till the stage of Web 1.0.

From (2) and (6) in Table 5, for top-tier institutions we found that the coefficient of *Cross-tierCollaboration*

TABLE 5. Effects of cross-tier collaboration on research productivity and the moderating role of stages of the information age.

Discipline	Natural sciences				Social sciences			
Institution tier	Top		General		Top		General	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cross-tierCollaborationRatio	−0.271***	−0.217***	0.210***	0.142***	−0.338***	−0.402***	0.329***	0.295***
	(0.034)	(0.026)	(0.006)	(0.009)	(0.077)	(0.077)	(0.009)	(0.015)
AuthorNum	0.105***	0.378***	0.517***	1.182***	0.177***	0.466***	0.602***	1.538***
	(0.005)	(0.011)	(0.013)	(0.029)	(0.009)	(0.022)	(0.017)	(0.054)
CollaboratorNum	0.974***	1.074***	0.959***	1.369***	1.330***	2.148***	1.410***	1.445***
	(0.033)	(0.070)	(0.013)	(0.022)	(0.061)	(0.197)	(0.016)	(0.039)
CollaboratorNum^2	−0.117***	−0.287***	−0.255***	−0.708***	−0.144***	−0.495***	−0.245***	−0.714***
	(0.005)	(0.021)	(0.003)	(0.011)	(0.008)	(0.055)	(0.004)	(0.020)
Dum_Web1.0		−0.033		−0.114***		0.010		−0.377***
		(0.100)		(0.014)		(0.262)		(0.027)
Dum_Web2.0		−0.037		−0.058**		0.352		−0.560***
		(0.091)		(0.020)		(0.221)		(0.037)
Cross-tierCollaborationRatio Dum_Web1.0*		−0.027		0.000		−0.051		−0.022
		(0.018)		(0.011)		(0.058)		(0.020)
Cross-tierCollaborationRatio Dum_Web2.0*		−0.043*		0.058***		−0.175**		−0.013
		(0.017)		(0.012)		(0.055)		(0.019)
*AuthorNum*Dum_Web1.0*		−0.140***		−0.445***		−0.117***		−0.610***
		(0.008)		(0.028)		(0.018)		(0.054)
*AuthorNum*Dum_Web2.0*		−0.222***		−0.667***		−0.230***		−0.890***
		(0.009)		(0.027)		(0.018)		(0.051)
*CollaboratorNum*Dum_Web1.0*		0.064		−0.120***		−0.250		0.108**
		(0.096)		(0.021)		(0.251)		(0.040)
*CollaboratorNum*Dum_Web2.0*		0.005		−0.401***		−0.667**		0.051
		(0.077)		(0.020)		(0.204)		(0.038)
*CollaboratorNum^2*Dum_Web1.0*		0.074**		0.342***		0.146*		0.289***
		(0.025)		(0.011)		(0.063)		(0.021)
*CollaboratorNum^2*Dum_Web2.0*		0.146***		0.528***		0.315***		0.454***
		(0.021)		(0.011)		(0.055)		(0.019)
Year	−0.101***	−0.126***	0.087***	−0.050***	0.006	0.057*	0.080***	0.191***
	(0.013)	(0.013)	(0.006)	(0.008)	(0.024)	(0.028)	(0.008)	(0.014)
Intercept	2.451***	2.860***	2.334***	2.919***	0.206	−0.319	0.560***	1.164***
	(0.080)	(0.088)	(0.017)	(0.022)	(0.115)	(0.198)	(0.014)	(0.027)
Wald chi-squared	9071.9	17160.1	31012.2	47597.3	12331.7	19979.5	40891.9	51855.9
Log likelihood	−5785.5	−5523.8	−48948.6	−47102.3	−11110.9	−10956.0	−78934.6	−78338.8
Observations	900	900	14,905	14,905	900	900	14,604	14,604
Model significance	***	***	***	***	***	***	***	***

*$p < .05$.

**$p < .01$.

***$p < .001$, two-tailed; Standard errors in parentheses; independent variables are standardized.

*Ratio*Dum_Web2.0* was negative in both the natural sciences ($\beta = -0.043$, $p < .05$) and social sciences ($\beta = -0.175$, $p < .01$), suggesting that negative effects of cross-tier collaboration on research productivity would be escalated in Web 2.0, meaning that researchers from top-tier institutions would benefit more through collaborating with their counterparts in top-tier institutions in Web 2.0. For general-tier institutions, we found a positive coefficient of *Cross-tierCollaborationRatio*Dum_Web2.0* in the natural sciences ($\beta = 0.058$, $p < .001$), suggesting that researchers from the general-tier would benefit more through collaborating with researchers in top-tier institutions in Web 2.0.

Research impact. We further examined the effects of cross-tier collaborations on research impact. Similarly, we also estimated fixed-effects negative binomial models because of the dispersion of research impact. We report the results in Table 6.

As shown in Table 6, for top-tier institutions, cross-tier collaboration showed a negative effect on research impact in both the natural ($\beta = -0.172$, $p < .001$) and social sciences ($\beta = -0.141$, $p < .05$). Contrarily, for general-tier institutions, cross-tier collaboration showed a positive effect on research impact in both the natural sciences ($\beta = 0.260$, $p < .001$) and social sciences ($\beta = 0.306$, $p < .001$). Hence, researchers from general-tier institutions appeared to be better off by collaborating with researchers from top-tier institutions.

We also examined whether these effects differed among three stages of the Information Age. For top-tier institutions, the coefficients of *Cross-tierCollaboration Ratio*Dum_Web2.0* were insignificant in both the natural and social sciences. For general institutions, we found a positive coefficient of *Cross-tierCollaborationRatio*Dum_Web2.0* in the natural sciences ($\beta = 0.051$, $p < .01$), suggesting the benefits from collaborating with researchers

DOI: 10.1002/asi

from top-tier institutions would be strengthened in Web 2.0. However, the coefficient turned out to be negative in the social sciences ($\beta = -0.051$, $p < .05$), suggesting the benefits from collaborating with researchers in top-tier institutions would be diminished in Web 2.0. The opposite effects could partially be attributed to the different geographic structures of the cross-tier collaboration: in the natural sciences, distances of cross-tier collaborations were increased, while in the social sciences, cross-tier collaborations became more localized.

Robustness Checks

We conducted several alternative estimations to examine the robustness of our results. We accounted for alternative classification of institution tiers and an alternative measure of ICT's development.

Alternative classification of top-tier institutions. As a robustness check, we classified the top 50 U.S. institutions as the top-tier because empirical studies typically did not regard U.S. institutions outside the top 50 as top institutions. Using the top 50 U.S. institutions as the top tier, we also examined trends of social stratification, trends of geographic separation, and the effects of cross-tier collaborations on research performance. The details are reported in the Supplementary Materials.

Trends of social stratification. When using the top 50 U.S. institutions as the top tier, the balkanization index decreased from 0.94 in 1980 to 0.78 in 2009 in the natural sciences and decreased from 0.90 in 1980 to 0.77 in 2009 in the social sciences (see Figure S2 in the Supplementary Materials). The correlation analysis results showed that this decreasing trend was significant along with years in both the natural sciences

TABLE 6. Effects of cross-tier collaboration on research impact and the moderating role of stages of the information age.

Discipline	Natural sciences				Social sciences			
Institution tier	Top		General		Top		General	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cross-tierCollaborationRatio	−0.172***	−0.118*	0.260***	0.208***	−0.141*	−0.098	0.306***	0.308***
	(0.049)	(0.047)	(0.008)	(0.012)	(0.070)	(0.076)	(0.010)	(0.016)
AuthorNum	0.094***	0.310***	0.710***	1.123***	0.156***	0.258***	0.716***	0.945***
	(0.011)	(0.022)	(0.021)	(0.047)	(0.013)	(0.029)	(0.029)	(0.071)
CollaboratorNum	0.806***	1.107***	1.338***	1.595***	0.704***	1.295***	1.134***	1.297***
	(0.058)	(0.115)	(0.018)	(0.033)	(0.074)	(0.179)	(0.022)	(0.046)
CollaboratorNum^2	−0.127***	−0.325***	−0.385***	−0.732***	−0.101***	−0.321***	−0.295***	−0.549***
	(0.010)	(0.037)	(0.006)	(0.019)	(0.012)	(0.055)	(0.007)	(0.026)
Dum_Web1.0		0.696***		−0.085***		0.545*		−0.088**
		(0.160)		(0.023)		(0.230)		(0.033)
Dum_Web2.0		0.388		−0.127***		0.953***		−0.187***
		(0.218)		(0.034)		(0.287)		(0.047)
Cross-tierCollaborationRatio Dum_Web1.0*		−0.036		0.009		0.051		−0.037
		(0.028)		(0.015)		(0.043)		(0.021)
Cross-tierCollaborationRatio Dum_Web2.0*		−0.014		0.051**		0.060		−0.051*
		(0.033)		(0.018)		(0.050)		(0.024)
*AuthorNum*Dum_Web1.0*		−0.118***		−0.260***		−0.050*		−0.203**
		(0.015)		(0.045)		(0.021)		(0.067)
*AuthorNum*Dum_Web2.0*		−0.189***		−0.496***		−0.088***		−0.255***
		(0.017)		(0.045)		(0.022)		(0.067)
*CollaboratorNum*Dum_Web1.0*		−0.551***		−0.147***		−0.539*		−0.087
		(0.156)		(0.030)		(0.227)		(0.045)
*CollaboratorNum*Dum_Web2.0*		−0.375*		−0.339***		−0.894***		−0.135**
		(0.166)		(0.031)		(0.226)		(0.046)
*CollaboratorNum^2*Dum_Web1.0*		0.195***		0.276***		0.165**		0.188***
		(0.041)		(0.017)		(0.061)		(0.025)
*CollaboratorNum^2*Dum_Web2.0*		0.220***		0.465***		0.271***		0.278***
		(0.040)		(0.017)		(0.056)		(0.025)
Year	0.189***	0.144***	0.167***	0.102***	0.215***	0.230***	0.177***	0.205***
	(0.021)	(0.026)	(0.009)	(0.015)	(0.028)	(0.035)	(0.010)	(0.020)
Intercept	2.353***	2.207***	1.522***	1.804***	1.883***	1.474***	0.929***	1.133***
	(0.114)	(0.129)	(0.015)	(0.027)	(0.139)	(0.185)	(0.018)	(0.036)
Wald chi-squared	4767.6	6188.4	27149.6	30991.5	4885.7	5509.7	15989.5	16527.4
Log likelihood	−6923.5	−6830.7	−60358.2	−59820.6	−5113.6	−5074.3	−41534.6	−41385.9
Observations	750	750	12,134	12,134	750	750	11,939	11,939
Model significance	***	***	***	***	***	***	***	***

*$p < .05$.

**$p < .01$.

***$p < .001$, two-tailed; Standard errors in parentheses; Independent variables are standardized.

DOI: 10.1002/asi

($r_{Pearson} = -0.992$, $p < .000$; $r_{Spearman} = -0.998$, $p < .000$) and social sciences ($r_{Pearson} = -0.965$, $p < .000$; $r_{Spearman} = -0.985$, $p < .000$).

The results of the Kruskal–Wallis Test showed that the balkanization index significantly differed among three stages of the Information Age. In addition, the pairwise comparison indicated a significantly decrease in Web 2.0 compared with the pre-Information Age (see Table S2 in the Supplementary Materials). These results were consistent with results using top 30 as the top tier.

Trends of geographic separation. When using the top 50 U.S. institutions as top-tier institutions, we also observed the opposite evolving trends of cross-tier collaborations (see Table S3a in the Supplementary Materials): in the natural sciences, there was an increasing trend of geographic distance ($r_{Pearson} = 0.611$, $p < .001$); however, a decreasing trend of geographic distance presented in the social sciences ($r_{Pearson} = -0.396$, $p < .05$).

The results of the Kruskal–Wallis Test and pairwise comparisons were generally consistent with our former analysis (see Table S3a). For instance, we observed a consistent increasing trend in general-tier collaborations in both the natural and social sciences. But we did not observe significant differences of top-tier collaboration in the natural sciences and of cross-tier collaboration in the social sciences.

Effects of cross-tier collaborations on research productivity. Using the top 50 U.S. institutions as the top tier, we also estimated the fixed-effects negative binomial models to examine the effects of cross-tier collaborations on research productivity (see Table S4 in the Supplementary Materials).

For top-tier institutions, we only observed the negative effects of cross-tier collaborations in the natural sciences ($\beta = -0.076$, $p < .001$) but not in the social sciences. For general-tier institutions, we consistently observed the positive effects of cross-tier collaborations in both the natural ($\beta = 0.204$, $p < .001$) and social sciences ($\beta = 0.140$, $p < .001$).

As with the moderating effects of different stages of the Information Age, for top-tier institutions we observed the negative moderating effects of Web 2.0 in both the natural ($\beta = -0.064$, $p < .001$) and social sciences ($\beta = -0.104$, $p < .001$) as well. For general-tier institutions, a positive moderating effect of Web 2.0 in the natural sciences ($\beta = 0.040$, $p < .01$) was also observed. These results were consistent with our previous analysis.

Effects of cross-tier collaborations on research impact. For general-tier institutions, we still found that cross-tier collaborations showed positive effects in both the natural ($\beta = 0.251$, $p < .001$) and social sciences ($\beta = 0.302$, $p < .001$). In addition, the moderating effects of Web 2.0 also showed contrarily in the natural sciences ($\beta = 0.064$, $p < .01$) and social sciences ($\beta = -0.066$, $p < .01$). These results were consistent with our previous findings. For top-tier institutions, the negative effects of cross-tier collaborations only existed in the social sciences ($\beta = -0.079$, $p < .05$), but not in the natural sciences. More details are in Table S5 in the Supplementary Materials.

Alternative measure of ICT's development. Although we developed three stages of the Information Age to capture the effects of ICT's development, the analysis might be overly simplified with dummy variables. Therefore, in this part we employed alternative measures of ICT's development—the growth of the Internet hosts to examine the robustness of previous findings.

To communicate in the Internet, one must use an IP address, which maps to one host. Therefore, the number of Internet hosts reflects the scale and the growth of the network, indicating the changing ability of ICTs. Figure 3 showed the increasing number of Internet hosts from 1981 to 2009.[2] The data before 1981 are unavailable. We could see that the boost in the number of Internet hosts is consistent with the stage of Web 1.0 and Web 2.0.

We also estimated the fixed-effects negative binominal models to examine how the growth of the Internet hosts moderates the effects of cross-tier collaborations on research productivity and research impact. We report the results in Table 7.

In terms of research productivity, for top-tier institutions the negative coefficient of *Cross-tierCollaborationRatio*HostNum* was only significant in the social sciences ($\beta = -0.070$, $p < .01$) but not in the natural sciences. For general-tier institutions, we only observed a significantly positive moderating effect in the natural sciences ($\beta = 0.015$, $p < .01$), which was consistent with our previous findings. In terms of research impact, we did not observe the significant moderating effects of growth of the Internet host in either the natural or social sciences, for either top- or general-tier institutions.

Conclusion and Discussion

Scientific collaboration has dominated in almost all disciplines, but successful and efficient collaboration among researchers was still challengeable because it was intertwined with geographic, social, and technological factors. The effects of ICTs in science were inconclusive due to two perspectives contrarily predicted the academic community that would evolve into a Global Village and a Virtual Balkan, respectively. Therefore, it is necessary to examine the evolving trends of scientific collaboration in both the geographic and social dimensions and how different collaboration patterns affect research performance. Furthermore, it is important to examine the role of ICT's development to provide a deeper understanding of the intertwined nature of scientific collaboration.

Our findings showed that social stratification in scientific collaboration, measured with the index of balkanization,

[2] Data come from the ISC Domain Survey conducted by the Internet Systems Consortium. The survey typically works by querying the domain system to count hosts. See https://www.isc.org/network/survey/about/ for more details.

DOI: 10.1002/asi

significantly decreased from 1980 to 2009 in both the natural and social sciences. The lower balkanization index indicated a higher extent of cross-tier collaboration, suggesting that the scientific community became more integrated rather than fragmented or a Virtual Balkan. These findings were consistent in respectively treating top 30 institutions and top 50 institutions as the top-cited tiers.

Understanding the geographic separation in scientific collaboration from 1980 to 2009 revealed some interesting facts. Generally, the collaboration distance significantly increased in both the natural and social sciences, indicating an evolving trend towards a Global Village where researchers were seeking to collaborate with distant scholars. However, the evolving trends differ among different collaboration patterns. For general-tier collaboration, collaboration distance showed consistent increasing trends in both the natural sciences and social sciences. While for top-tier collaboration, the distance decreased in the natural sciences but increased in the social sciences, suggesting that collaboration among researchers from top-tier institutions became more localized. For cross-tier collaboration, distance also showed opposite trends in the natural and social sciences. Considering the role played by ICT's development, we found the collaboration distance significantly changed in Web 2.0 compared with the pre-Information Age. Therefore, we believed that an increasing trend of collaboration distance indicated that researchers leveraged the development of ICTs to access distant partners, while the decreasing collaboration distance may be

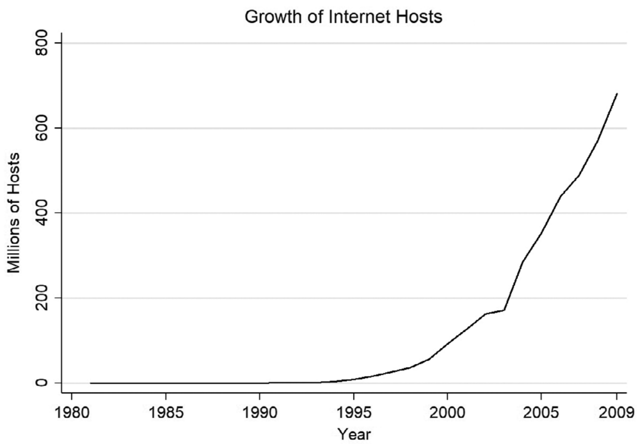

FIG. 3. Growth of Internet hosts from 1981 to 2009.

TABLE 7. Effects of cross-tier collaboration on research performance and the moderating role of Internet host number.

	Research productivity				Research impact			
Discipline	Natural sciences		Social sciences		Natural sciences		Social sciences	
Institution tier	Top	General	Top	General	Top	General	Top	General
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cross-tierCollaborationRatio	−0.249***	0.190***	−0.562***	0.305***	−0.122*	0.247***	−0.016	0.277***
	(0.032)	(0.006)	(0.084)	(0.009)	(0.055)	(0.012)	(0.083)	(0.014)
AuthorNum	0.161***	0.677***	0.278***	0.858***	0.121***	0.709***	0.166***	0.712***
	(0.006)	(0.014)	(0.011)	(0.021)	(0.011)	(0.020)	(0.015)	(0.029)
CollaboratorNum	1.019***	1.070***	1.423***	1.435***	1.160***	1.284***	0.675***	1.172***
	(0.038)	(0.013)	(0.066)	(0.018)	(0.133)	(0.018)	(0.161)	(0.026)
CollaboratorNum^2	−0.151***	−0.288***	−0.203***	−0.324***	−0.193***	−0.329***	−0.105***	−0.312***
	(0.007)	(0.004)	(0.011)	(0.006)	(0.023)	(0.007)	(0.027)	(0.010)
HostNum	−0.305***	0.052***	0.132	0.105***	−0.720**	0.042	0.506	−0.008
	(0.053)	(0.006)	(0.085)	(0.011)	(0.273)	(0.022)	(0.309)	(0.030)
Cross-tierCollaborationRatio HostNum*	0.001	0.015**	−0.070**	−0.004	0.039	0.019	0.067	−0.035
	(0.008)	(0.005)	(0.023)	(0.008)	(0.032)	(0.018)	(0.044)	(0.023)
*AuthorNum*HostNum*	−0.026***	−0.107***	−0.035***	−0.113***	−0.075***	−0.314***	−0.043***	−0.117***
	(0.002)	(0.005)	(0.003)	(0.007)	(0.010)	(0.024)	(0.012)	(0.034)
*CollaboratorNum*HostNum*	0.168***	−0.143***	−0.021	−0.023**	0.334	−0.252***	−0.348	−0.059
	(0.030)	(0.005)	(0.048)	(0.008)	(0.176)	(0.020)	(0.197)	(0.031)
*CollaboratorNum^2*HostNum*	−0.009*	0.075***	0.018*	0.042***	0.015	0.213***	0.099**	0.103***
	(0.005)	(0.002)	(0.007)	(0.002)	(0.029)	(0.008)	(0.033)	(0.012)
Year	−0.067***	−0.020**	−0.102***	−0.029*	0.226***	0.100***	0.186***	0.163***
	(0.015)	(0.007)	(0.028)	(0.011)	(0.024)	(0.011)	(0.032)	(0.014)
Intercept	2.611***	2.611***	0.138	0.729***	1.916***	1.653***	2.092***	0.989***
	(0.084)	(0.018)	(0.120)	(0.015)	(0.211)	(0.017)	(0.253)	(0.021)
Wald chi-squared	10781.7	37889.6	16756.4	46609.8	5017.3	28632.4	4867.1	15633.8
Log likelihood	−5486.1	−46531.5	−10638.9	−76329.7	−6595.4	−57903.4	−4889.8	40068.5
Observations	870	14,486	870	14,192	720	11,710	720	11,516
Model significance	***	***	***	***	***	***	***	***

*$p < .05$.

**$p < .01$.

***$p < .001$, two-tailed; Standard errors in parentheses; independent variables are standardized.

DOI: 10.1002/asi

partially attributed to using ICTs to search and filter interested scholars nearby.

We further discovered the effects of collaboration patterns on research performance to help researchers choose whom to collaborate with. Our findings showed that cross-tier collaboration would benefit researchers from general-tier institutions in terms of productivity and impact at the expense of researchers from top-tier institutions. These findings were consistent with previous studies that revealed that researchers from nonelite institutions tend to gain from collaborating with researchers from elite institutions (Agrawal & Goldfarb, 2008). For general-tier institutions, the cross-tier collaboration could be achieved through hiring graduates of top-tier institutions, appointing researchers from top-tier institutions to honorary and supervisory roles, and sharing valuable research resources (for example, unique data sets and field research opportunities) with researchers from top-tier institutions, and so on. Nevertheless, for top-tier institutions, our results suggested that collaboration with researchers from general-tier institutions did not bring a gain in productivity (and even impact), as some previous studies suggested (Knorr, Mittermeir, Aichholzer, & Waller, 1976; Pao, 1992). These findings revealed two important implications: One was that not all collaborations were equal; some collaborations could enhance productivity (or impact), while others diminished. The other one was that, although productivity and impact were important indicators of research performance, they were not the only indicators. Besides enhancing productivity and impact, scientific collaboration seeks an optimal fit among resources in many cases and building complementary relationships (Lee & Bozeman, 2005).

It is worth noting that the development of ICTs plays an important role in scientific collaboration through influencing both geographic barriers and social stratification. Generally, the decreasing trend of social stratification and the increasing trend of collaboration distance suggested that the scientific community move towards a flatter world along with the development of ICTs. However, different evolving trends of collaboration distance among different collaboration patterns suggested that researchers may leverage ICTs not only to bridge geographic gaps to reach distant partners, but also in helping to form more localized collaboration relationships. Moreover, general-tier institutions took more advantages from leveraging ICTs compared to top-tier institutions, which helped to enhance productivity and impact.

Therefore, our findings have important implications for scholars, institution administrators, and national policy makers. For scholars, especially individuals from general-tier institutions, our findings support that the development of ICTs benefits them more, while they are often lacking in research resources and facing greater mobility constraints due to the internal or external political environment (Chinchilla-Rodríguez et al., 2018). For administrators of general-tier institutions, our findings suggest that ICTs could be employed to reach distant potential collaborators and connect to top-tier institutions, which helps them to narrow the physical and nonphysical gaps among other scientists. In addition, the results of this study suggest that it would be worthwhile for national funding agencies to promote cross-tier collaboration to develop more institutions producing high-impact research, given that a considerable amount of research resources is allocated to scholars in top-tier institutions (Bohlmann, Calantone, & Zhao, 2010).

One of the limitations of this study is that we only investigated the scientific collaboration in the United States and we did not consider the potential influences of the mobility of researchers within the United States. Future analysis could incorporate this factor to provide more insight. In addition, similar analysis could be extended to reveal evolving trends of scientific collaboration in other countries and international collaborations, which could also consider the potential influences of global mobility. Furthermore, although we considered both natural and social sciences, the results could be too encapsulated to reflect characteristics of certain fields. Thus, similar analysis could be extended in certain specific disciplines to provide deeper insights on the evolution and performance of scientific collaboration along with the development of ICTs.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (#71871065, 91846302), Shanghai Philosophy and Social Science Plan (#2017BGL019) and Fundamental Research Funds for the Central Universities (#30919013203).

References

Abramo, G., D'Angelo, C.A., & Di Costa, F. (2009). Research collaboration and productivity: Is there correlation? Higher Education, 57(2), 155–171.

Agrawal, A., & Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of institution innovation. American Economic Review, 98(4), 1578–1590.

Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., & Song, M. (2017). Standing on the shoulders of giants. Journal of Informetrics, 11 (1), 307–323.

Beaver, D.D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. Scientometrics, 52(3), 365–377.

Birnholtz, J.P. (2007). When do researchers collaborate? Toward a model of collaboration propensity. Journal of the American Society for Information Science and Technology, 58(14), 2226–2239.

Bohlmann, J.D., Calantone, R.J., & Zhao, M. (2010). The effects of market network heterogeneity on innovation diffusion: An agent-based modeling approach. Journal of Product Innovation Management, 27(5), 741–760.

Boudreau, K.J., Brady, T., Ganguli, I., Gaule, P., Guinan, E., Hollenberg, A., & Lakhani, K.R. (2017). A field experiment on search costs and the formation of scientific collaborations. Review of Economics and Statistics, 99(4), 565–576.

Bozeman, B., & Boardman, C. (2014). Research collaboration and team science: A state-of-the-art review and agenda. New York: Springer.

Bu, Y., Ding, Y., Xu, J., Liang, X., Gao, G., & Zhao, Y. (2018). Understanding success through the diversity of collaborators and the milestone of career. Journal of the Association for Information Science and Technology, 69(1), 87–97.

Cairncross, F. (1997). The death of distance: How the communications revolution will change our lives. Cambridge, MA: Harvard Business School Publishing.

Chang, H.W., & Huang, M.H. (2013). Prominent institutions in international collaboration network in astronomy and astrophysics. Scientometrics, 97 (2), 443–460.

Chinchilla-Rodríguez, Z., Bu, Y., Robinson-García, N., Costas, R., & Sugimoto, C.R. (2018). Travel bans and scientific mobility: Utility of asymmetry and affinity indexes to inform science policy. Scientometrics, 116(1), 569–590.

Ding, W.W., Levin, S.G., Stephan, P.E., & Winkler, A.E. (2010). The impact of information technology on academic scientists' productivity and collaboration patterns. Management Science, 56(9), 1439–1461.

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. Journal of Informetrics, 5 (1), 187–203.

Donath, J.S. (1995). Visual who: Animating the affinities and activities of an electronic community. In *Proceedings of the third ACM international conference on multimedia* (pp. 99–107). New York: ACM.

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. Journal of Informetrics, 4(4), 540–553.

Freeman, R.B., & Huang, W. (2014). Collaboration: Strength in diversity. Nature News, 513(7518), 305.

Freeman, R.B., & Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. Journal of Labor Economics, 33(S1), S289–S318.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. Journal of Informetrics, 3(3), 222–232.

Galison, P., Hevly, B., & Weinberg, A.M. (1992). Big science: The growth of large-scale research. Physics Today, 45, 89–90.

Gazni, A., & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: A case study of Harvard Institution's publications. Scientometrics, 87(2), 251–265.

Gazni, A., Sugimoto, C.R., & Didegah, F. (2012). Mapping world scientific collaboration: Authors, institutions, and countries. Journal of the Association for Information Science and Technology, 63(2), 323–335.

Hara, N., Solomon, P., Kim, S.L., & Sonnenwald, D.H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. Journal of the American Society for Information Science and Technology, 54(10), 952–965.

Hoekman, J., Frenken, K., & Tijssen, R.J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. Research Policy, 39(5), 662–673.

Jones, B.F., Wuchty, S., & Uzzi, B. (2008). Multi-institution research teams: Shifting impact, geography, and stratification in science. Science, 322(5905), 1259–1262.

Jonkers, K., & Tijssen, R. (2008). Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity. Scientometrics, 77(2), 309–333.

Kling, R., & McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. Journal of the American Society for Information Science, 51(14), 1306–1320.

Knorr, K.D., Mittermeir, R., Aichholzer, G., & Waller, G. (1976). In F. F. Andrews (Ed.), *Scientific productivity: The effectiveness of research groups in science*. Cambridge Institution Press: Cambridge, MA.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C.R. (2016). Contributorship and division of labor in knowledge production. Social Studies of Science, 46(3), 417–435.

Larivière, V., Gingras, Y., Sugimoto, C.R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. Journal of the Association for Information Science and Technology, 66(7), 1323–1332.

Laudel, G. (2002). What do we measure by co-authorships? Research Evaluation, 11(1), 3–15.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. Social Studies of Science, 35(5), 673–702.

Leiner, B.M., Cerf, V.G., Clark, D.D., Kahn, R.E., Kleinrock, L., Lynch, D.C., … Wolff, S.S. (1997). The past and future history of the Internet. Communications of the ACM, 40(2), 102–108.

Li, J., & Li, Y. (2015). Patterns and evolution of coauthorship in China's humanities and social sciences. Scientometrics, 102(3), 1997–2010.

Li, Y., Li, H., Liu, N., & Liu, X. (2018). Important institutions of interinstitutional scientific collaboration networks in materials science. Scientometrics, 117(1), 85–103.

Liang, X. (2015). The changing impact of geographic distance: A preliminary analysis on the co-author networks in scientometrics (1983–2013). In *System sciences (HICSS), 2015 48th hawaii international conference on (pp. 722–731)*. IEEE.

Lundberg, J., Tomson, G., Lundkvist, I., & Brommels, M. (2006). Collaboration uncovered: Exploring the adequacy of measuring institution-industry collaboration through co-authorship and funding. Scientometrics, 69(3), 575–589.

Marmolejo-Leyva, R., Perez-Angon, M.A., & Russell, J.M. (2015). Mobility and international collaboration: Case of the Mexican scientific diaspora. PLoS One, 10(6), e0126720.

Milojevic, S. (2010). Modes of collaboration in modern science: Beyond power laws and preferential attachment. Journal of the American Society for Information Science and Technology, 61, 1410–1423.

Pao, M.L. (1992). Global and local collaborators: A study of scientific collaboration. Information Processing & Management, 28(1), 99–109.

Pennington, D.D., Simpson, G.L., McConnell, M.S., Fair, J.M., & Baker, R. J. (2013). Transdisciplinary research, transformative learning, and transformative science. Bioscience, 63(7), 564–573.

Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. Papers in Regional Science, 86(3), 423–443.

Rosenblat, T.S., & Mobius, M.M. (2004). Getting closer or drifting apart? The Quarterly Journal of Economics, 119(3), 971–1009.

Schaltegger, S., Beckmann, M., & Hansen, E.G. (2013). Transdisciplinarity in corporate sustainability: Mapping the field. Business Strategy and the Environment, 22(4), 219–229.

Schilling, M.A., & Phelps, C.C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. Management Science, 53(7), 1113–1126.

Sidone, O.J.G., Haddad, E.A., & Mena-Chalco, J.P. (2017). Scholarly publication and collaboration in Brazil: The role of geography. Journal of the Association for Information Science and Technology, 68(1), 243–258.

Sie, R., Drachsler, H., Bitter-Rijpkema, M., & Sloep, P. (2012). To whom and why should I connect? Co-author recommendation based on powerful and similar peers. International Journal of Technology Enhanced Learning, 4(1–2), 121–137.

Srinivasan, R. (2017). *Whose global village?: Rethinking how technology shapes our world*. New York: NYU Press.

Stevens, M.J., & Campion, M.A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. Journal of Management, 20(2), 503–530.

Thijs, B., & Glänzel, W. (2010). A structural analysis of collaboration between European research institutes. Research Evaluation, 19(1), 55–65.

Van Alstyne, M., & Brynjolfsson, E. (1996). Could the Internet balkanize science? Science, 274(5292), 1479–1480.

Van Alstyne, M., & Brynjolfsson, E. (2005). Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. Management Science, 51(6), 851–868.

Wang, J. (2013). Citation time window choice for research impact evaluation. Scientometrics, 94(3), 851–872.

Wuchty, S., Jones, B.F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. Science, 316(5827), 1036–1039.

Yao, X., Li, X., Zhang, C., & Ling, H. (2017). Fueling virtual teams with creativity through composition of private and public workspaces. In 2017 38th International Conference on Information Systems (ICIS).

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. Journal of the Association for Information Science and Technology, 69(1), 72–86.

DOI: 10.1002/asi