https://doi.org/10.51408/issi2025_020

The Increasing Fragmentation of Global Science Limits the Diffusion of Ideas

Alexander J. Gates[1], Indraneel Mane[2], Jianjian Gao[3]

[1]*agates@virginia.ed at*
School of Data Science, University of Virginia, Charlottesville, Virginia (USA)

[2]*maneindraneel@gmail.com at*
Network Science Institute, Northeastern University, Boston, Massachusett (USA)

[3]*psp2nq@virginia.edu at*
School of Data Science, University of Virginia, Charlottesville, Virginia (USA)

Abstract

The global scientific landscape emerges from a complex interplay of collaboration and competition, where nations vie for dominance while simultaneously fostering the diffusion of knowledge on a global scale. This raises crucial questions: What underlying patterns govern international scientific recognition and influence? How does this structure impact knowledge dissemination? Traditional models view the global scientific ecosystem through a core-periphery lens, with Western nations dominating knowledge production. Here, we investigate the dynamics of international scientific recognition through the lens of citation preferences, introducing a novel signed measure to characterize national citation preferences and enabling a network analysis of international scientific recognition. We find that scientific recognition is related to cultural and political factors in addition to economic strength and scientific quality. Our analysis challenges the conventional core-periphery narrative, uncovering instead several communities of international knowledge production that are rapidly fragmenting the scientific recognition ecosystem. Moreover, we provide a comprehensive statistical model that shows this network significantly constrains the diffusion of ideas across international borders. The resulting network framework for global scientific recognition sheds light on the barriers and opportunities for collaboration, innovation, and the equitable recognition of scientific advancements, with significant consequences for policymakers seeking to foster inclusive and impactful international scientific endeavours.

Introduction

The global scientific research ecosystem is shaped by the emergent interplay between international collaboration, competition, and recognition, which collectively drive the diffusion of ideas and the cross-border flow of knowledge (Hagstrom, 1974; Chinchilla-Rodr´ıguez et al., 2019; Marginson, 2022a). Strong national research infrastructures empower nations to vie for competitive advantages in technology, economics, security, and health. Concurrently, scientific knowledge flows on a global scale, with scientific ideas disseminating from their nation of origin and influencing research around the world. This diffusion and adoption of scientific information transcends national boundaries, forming a global network of scientific recognition and influence. However, the strength of influence is not uniform across all communities, leading to status stratification where nations are differentially recognized for their scientific contributions (Moravcsik, 1985; Schott, 1998; Galvez et al.´, 2000; Tickner, 2013; Collyer, 2014; Gomez et al., 2022). This raises two central questions to be explored: What structural patterns underlie international scientific recognition and influence? and What are the consequences of that structure for knowledge dissemination?

The prevailing theories for the structure and consequences of global scientific recognition closely mirror economic models, with a clear hierarchy and power dynamics between the "core" of scientific knowledge production and its "periphery" such that certain regions or countries dominate the production and dissemination of scientific research while others occupy a peripheral or marginalized position (Prebisch, 1962; Shils, 1975; May, 1997; King, 2004; Zelnio, 2012). This core-periphery structure is hypothesized to have important consequences for international science by hindering diverse perspectives and knowledge diffusion. The core-periphery model tends to oversimplify the complex relationships between nations, reducing influence dynamics to a binary classification of 'core' or 'periphery', while overlooking the nuances and inter-dependencies that shape global science (Schott, 1988a). By relying on this model, policy and funding decisions risk becoming skewed in favor of established centers, reinforcing existing national disparities. Core countries dominate research agendas and attract greater resources, while peripheral regions struggle to keep pace, further entrenching their marginal position in the global scientific network (Sumathipala et al., 2004; Kozlowski et al., 2022; Abramo et al., 2020; Heimeriks and Boschma, 2014).

Quantitative support for the core-periphery structure of global scientific recognition is evident across various dimensions of academic activity, including international collaboration, researcher mobility, and citation patterns. For example, international collaboration networks show that core countries have higher degrees of centrality and connectivity than periphery countries, indicating their dominant role in global science (Leydesdorff and Wagner, 2008; Zelnio, 2012; Gui et al., 2019; Choi, 2012; Wagner et al., 2015), and the global embeddedness of a nation, quantified by proportion of internationally co-authored publications, is a significant predictor of traditional scientific impact (Wagner and Jonkers, 2017). Additional analysis utilizing hierarchical clustering and dominant flow methodologies on international collaboration networks suggest that the global scientific community consists of four tiers: core, strong semi-periphery, semiperiphery, and periphery (Gui et al., 2019). Under this model, the United States consistently occupies the core, maintaining collaborations with nearly every major scientific nation, while emerging powers like China and South Korea have only recently ascended to the core. Mobility patterns also reveal that core countries attract more foreign scientists and researchers than periphery countries, suggesting their greater availability of resources and opportunities (Freeman, 2010; Scott, 2015; Adams, 1998; Urbinati et al., 2021; Bauder et al., 2018). Scott (2015) refers to this phenomenon as "hegemonic internationalisation" where internationalization becomes an extension of global inequality and the struggle for dominance, driven by competition, rankings, and the concentration of academic power in certain geopolitical centers. Analysis of raw citation networks further demonstrate that core countries generate more citations than periphery countries, implying their higher impact and influence on scientific research (Schott, 1988b, 1998; Choi, 2012; Gomez et al., 2022). Notably, Gomez et al. (2022) draws on the existing classification of countries into core and periphery to reveal a growing disparity between the number of citations a country receives and the textual similarity of the publications they produce.

Yet, it is often argued that the core-periphery model is entrenched in a Western-centric perspective that prioritizes resources and personnel, and thus overlooks the diverse cultural influences and research priorities shaping global scientific recognition and influence (Schott, 1988b; Seth, 2009; Marginson, 2022b). As early as 1988, Schott (1988b) suggested that the coreperiphery structure is primarily attributed to the volume of a nations' scientific output which obfuscates the importance of other key factors related to ties between countries, such as geopolitical relationships, linguistic similarities, colleagueship, scientific cooperation, and educational connections. Indeed, publication output remains heavily concentrated in the United States and a few European nations, implying that most quantitative indicators of scientific recognition—such as those based on raw publication, collaboration, and citation counts—tend to be notoriously Western-centric (May, 1997; King, 2004; Gomez et al., 2022). These metrics often overlook contributions from regions with smaller output, failing to recognize the diverse intellectual contributions and local innovations that may not fit neatly within dominant Western frameworks (Anderson, 2018). These limitations highlight the need for more nuanced approaches that account for regional and contextual variations in scientific production and influence.

Recent observations challenge the Western-centric narrative, indicating that emerging scientific nations are reshaping the global landscape of scientific recognition. Countries like China, Singapore, and South Korea are increasingly disrupting the traditional dominance of Western nations, signaling a shift in the concentration of global scientific influence (Lariviere et al., 2018; Basu et al., 2018; Leydesdorff et al., 2013; Gui et al., 2019; Choi, 2012). However, there is a growing tension between two perspectives: one that focuses on individual nations' transitions from the periphery to the core, and another that critiques the vertical stratification and lower visibility of researchers from regions like Latin America, the Middle East, and East Asia. The latter perspective is best articulated by Marginson (Marginson, 2022b) who discusses "the collapse of the centre-periphery model" which he attributes to internal collaboration and regional alliances rather than through traditional engagement with Euro-American scientific hubs (Marginson and Xu, 2023). AdamsAdams (2012) further characterizes such regional collaboration as a form of mutual recognition among partners within the region, fostering the development of emerging research economies.

Despite these qualitative insights, the tension remains unresolved due to a lack of robust quantitative evidence comparing the rise of individual countries within the existing core-periphery hierarchy with the creation of distinct regional scientific communities. Quantitative analyses are crucial for determining whether these regional networks are merely reinforcing the global hierarchy or truly reshaping it.

Without data-driven comparisons, it remains unclear whether the traditional core-periphery model still applies or if a more nuanced framework is needed to capture the evolving dynamics of global scientific influence.

Here, to map the structure underlying global scientific recognition and evaluate its implications for scientific influence, we analyzed the evolution of citation networks constructed from scientific publications and geolocated by their authors' affiliations. Although citations are only one–among many–means to acknowledge scientific recognition, the accessibility and quantity of such data provide a useful perspective of how scientific influence accumulates. Specifically, our data is built from 57,558,268 papers contained in the OpenAlex publication database from 1990 to 2022, which can be attributed to the countries from which their authors are affiliated, in total capturing the output of 223 countries and independent states. We must acknowledge that the OpenAlex database has known limitations, including incomplete affiliation coverage (Zhang et al., 2024) and a primary focus on English-language journals, which may introduce a selection bias towards Western countries (Gong et al., 2019). Despite these constraints, our results effectively identify significant patterns in scientific recognition. We then extracted 242 million citation relationships and calculated the number of country-specific citations to each paper within 5 years of publication (see SI, section S2).

To quantitatively capture scientific recognition, we adopt a popular measure of rank overrepresentation or under-representation (Methods, and SI, section S3), which empowers us to measure when one nationality over- or under-cites the papers from another nationality, accompanied by a level of statistical significance. To our knowledge, this marks the first application of such a method to determine whether one nation exhibits a preference or aversion towards another's scientific publications. The recognition relationship between countries in scientific output is influenced by a complex interplay of factors, including nationality bias, disparities in research quality, and international collaboration. Our study, through this measure, aims not to disentangle these individual factors but to elucidate the overall landscape resulting from their combined effects. We compare these citation patterns to a baseline constructed from the citation distribution of the source country to all other countries in the same year. This baseline is specifically tailored to each source country and year, representing the actual distribution of citations accumulated over a 5-year window from the source country to all global publications within that year. The resulting measure of citation preference between a source and target country can be interpreted as the probability that a randomly selected publication from the target country has more citations from the source country than a randomly selected publication from anywhere else in the world, and assumes a value between 0 (strong preference against) and 1 (strong preference for), where a value of 0.5 captures no preference. Since our method aggregates over a 5-year citation window, the most recent year for our analysis is 2017.

There are many possible mechanisms that may contribute to strong citation preferences; our data lets us further control for two potential contributions. First, scientists are known to self-cite (Aksnes, 2003) at rates which vary based on culture, discipline, and demographics (King et al., 2017; Azoulay and Lynn, 2020). Second, sharing an affiliation can increase the propensity to engage with a scientist's work (Wuestman et al., 2019). To control for the possible influence of these two factors, we removed all citations between publications that share at least one author or at least one affiliation (SI, Section S2). This framework further accommodates controlling for specific factors which may influence national citation preferences, including scientific disciplines and journals, by modifying the citation baseline (see Methods).

Data and Methods

Bibliometric Data

The dataset was drawn from the OpenAlex (Priem et al., 2022) bibliometric database in July 2022. OpenAlex is built upon the Microsoft Academic Graph (MAG), which Microsoft shuttered in December 2021, CrossRef, and ORCID. We used all indexed "journal-article" and "proceedings-article" records listed as published after 1990 and excluded any publication that did not list an institutional address.

Publications are associated with countries using the institutional addresses listed by the authors. We assign a full unit credit of a publication to every country of affiliation on the paper's author byline ("full counting"). For example, a paper listing ten authors– three with affiliations in Hungary, five with affiliations in the United States, and two in Canada— would count one paper to all three countries. See Supplementary Information for more details.

National Co-variate Data

We use data on national GDP, GDP per capita, and Population from the World Bank (Fantom and Serajuddin, 2016) to approximate the economic wealth and size of each country. The dataset covers 264 countries from 1960 to 2023. The official spoken language is provided for 195 countries and is encoded as a binary variable denoting common language for country pairs (Melitz and Toubal, 2014). We also source the bilateral distances (in kilometres) for most country pairs across the world from the *GeoDist* dataset provided by the Centre for Prospective Studies and International Information (CEPII) (Mayer and Zignago, 2011). This dataset also provides the continent each country belongs to, which we convert into a binary indicator denoting whether two countries belong to the same continent. In addition, Science and Technology Agreements (STA) are regarded as an important tool to achieve strategic Science Diplomacy (SD) objectives (Langenhove, 2017). We select records of STAs between countries (Nicolas Ruffin and¨ Schreiterer, 2017) to obtain the cumulative number of STAs between two countries over time.

National citation preference

We fix a year y and a source country (citing country) s and identify all publications with at least one affiliation in the source country over the next 5 years (y to y+5). We then find all publications worldwide published in year y that also received citations from the source country's 5-year publications. This process generates country-specific citation frequencies ($c_{s,5}$) over the fiveyear observation window, enabling us to establish a hierarchical ranking of $n_{s,y}$ publications that have garnered at least one citation from the source country ($c_{s,5}$ >= 1). This forms the baseline sample, comprising a citation distribution $p(c_5|s,y)$ specific to the source country s and year y, with a sample size of $n_{s,y}$. Next, we narrow our analytical focus to a designated target country t, identifying a subset of $n_{s,t,y}$ publications within our sample. These publications, represented by the distribution $p(c_5|s,t,y)$, must satisfy two criteria: they have received citations from the source country and maintain at least one institutional affiliation within the target country.

The national citation preference, $P_{s,t,y}$, from the source country s to the target country t in year y is found using the Area Under the receiver-operator Curve (AUC) as a measure of the extent to which the target country's publications are randomly distributed throughout the source country's ranking. Specifically, the national preference is found as:

$$P_{s,t,y} = \frac{1}{n_{s,t,y}\, n_{s,y}} \sum_{i=1}^{n_{s,t,y}} \sum_{j=1}^{n_{s,y}} \mathbb{I}\left(c_5^{(i)} > c_5^{(j)}\right) \quad (1)$$

where $c_5^{(i)}$ is the i-th sample from $p(c_5|s,t,y)$, $c_5^{(j)}$ is the the j-th sample from $p(c_5|s,y)$, and $\mathbb{I}$ is the indicator function, which is 1 if $c_5^{(i)} > c_5^{(j)}$ and 0 otherwise. The AUC is a measure of the probability (between 0 and 1) that a randomly chosen publication from the cited country is ranked higher than a randomly chosen publication from any other country; a value of 1 reflects the cited country's publications are over-expressed towards the top of the ranking, 0 occurs when the cited country's publications are under-expressed towards the bottom of the ranking, and 0.5 denotes a random distribution throughout the ranking.

We can further quantify the statistical significance of the over/under-representation of a specific country in the citation counts due to the equivalence of the AUC and Mann-Whitney U statistic (a.k.a. the Wilcox rank sum statistic). Specifically, we follow DeLong et al. to compare the observed AUC to 0.5 (DeLong et al., 1988) using the algorithm's fast implementation (Sun and Xu, 2014).

International citation preference network

The international citation preference network is a temporal network, independently constructed for each year. To avoid multiple hypothesis testing, we used the Holm step-down method (Holm, 1979) using Bonferroni adjustments as implemented in Statsmodels with $\alpha = 0.01$. The cumulative network aggregates over of all time slices and adopts the sign of the most recent slice in which the edge appeared.

The community structure within the positive international citation preference network is found using the Degree Corrected Stochastic Block Model (DCSBM) as implemented in graphtool (Peixoto, 2017). Network centrality for the positive international citation preference network is found using the PageRank algorithm with a return probability of $\alpha = 0.85$.

Stratified bootstrap baseline

To account for potential explanatory factors such as disciplinary focus and journal quality, we refine the assumptions underlying the random baseline in our national citation preference measure. We achieve this by implementing a stratified bootstrap approach, where we sample from the conditional citation distribution while ensuring that the sampled set exactly matches the observed publication counts for each journal in the observed citation distribution. Specifically, given the sample of $n_{s,t,y}$ publications affiliated with the target country t in year y and cited by the source country s, we track the frequency with which each journal appears, denoted $j_{s,t,y}$. We then sample with replacement from the source country's baseline distribution $p(c_5|s,y)$ such that the journal counts remain consistent with the observed values. This adjustment controls for the influence of journal-specific factors and disciplinary differences. We then perform 100 samples of this bootstrap procedure and use the mean and standard deviation of the AUCs to identify statistically significant links.

Scientific ideas

To identify scientific ideas, we follow the methodology introduced in Cheng et al. 2023 (Cheng et al., 2023). Specifically, we analyze the titles and abstracts for all of the publications in our OpenAlex corpus to identify the publications that mention at least one of 46,535 scientific ideas derived by Cheng et al. using the data-driven phrase segmentation algorithm, AutoPhrase (Shang et al., 2018). We then post-process these ideas, removing cases that were first mentioned before 2000 and focusing only on those ideas that were mentioned by only one country in their first year of usage, resulting in 7,327 unique ideas mentioned in 202,932 publications. Finally, we derive a dyadic variable, for all pairs of countries in our network that mentioned at least one idea, denoting the fraction of ideas whose first usage was in the Origin country and then were later used in the Destination country.

Logistic regression analysis

We use a logistic regression model to investigate the potential relationship between the propensity for scientific ideas to spread between countries and their connectivity in the international citation preference network. The model is written as follows:

$$\log \frac{y_c}{1-y_c} = \beta_0 + \beta_1 X_{1,c} + \beta_2 X_{2,c} + \cdots + \beta_k X_{k,c} \quad (2)$$

Where c denotes countries and y_c is the dependent variable. For the first group of models, we use the fraction of ideas that originate in the origin country and are later mentioned by the destination country (see Methods and SI, section S2). The included control variables are the GDP and Population for both the Origin and Destination countries. The investigated independent variables are the total number of ideas mentioned by the Origin and Destination countries, the Topical Distance between the countries' publications, the Physical Distance between the countries, a binary indicator of common official language, the one-hot encoding of a directed positive edge from the Destination to the Origin in the international citation preference network, the onehot encoding of a directed negative edge from the Destination to the Origin in the international citation network, and the PageRank centrality of the Origin and Destination countries in the positive international citation preference network. We apply log-transformation with base 10 to GDP, Population, and Physical Distance. All features besides the binary features (Same Official Lang, Positive Edge, Negative Edge) are standarized by subtracting the mean and dividing by the standard deviation.

Fixed-effect multinomial logistic regression

We use the multinomial logit model to predict the trinary citation preference between countries (e.g. positive, negative, or no preference). The multinomial logit model assumes that the log odds of each category $s \in \{-1,1\}$ relative to the reference category of no citation preference ($s = 0$) is a linear combination of the independent variables. Specifically, the model is defined as follows:

$$\log\left(\frac{P(Y_{ijt}=s)}{P(Y_{ijt}=0)}\right) = \beta_{s0} + \beta_{s1}X_{it} + \beta_{s2}X_{jt} + \beta_{s3}X_{ijt} + \alpha_t \quad (3)$$

where $P(Y_{ijt} = s)$ is the probability of the edge sign between source country i and target country j at time t taking value $s \in \{-1,1\}$; X_{it} and X_{it} capture potential country-specific characteristics in the country i and j at time t, respectively, while X_{ijt} represents potential pair-specific barriers or catalysts between country i and j at time t; α_t are the time-specific effects (intercepts) that capture the heterogeneity across time periods. β_{s0} is the intercept for category s; β_{s1}, β_{s2} and β_{s3} are the coefficients associated with the independent variables X_i, X_j and X_{ijt} for category s. We investigate different variants of the above model to study different combinations of countryspecific and country-pair-specific variables. The included control variables are the GDP per capita, population, and the fraction of top journal publications for both the Source and Target countries. The investigated pair-specific independent variables are physical distance, field distance, the same continent, the same official language, the cumulative number of bilateral science and technology agreements and scientific collaboration strength. We apply log-transformation with a base 10 to GDP per capita, population, physical distance, the cumulative number of bilateral science and technology agreements and scientific collaboration strength. All non-binary features are standardized by subtracting the mean and dividing by the standard deviation.

Results

International network of scientific recognition

We first build the network of international scientific recognition (Fig. 1A). The international scientific recognition network is a temporal signed and directed network in which each country is a node, and a source country is linked to a target country by a positive (negative) edge if the source country over-cites (under-cites) the target country's publications. To begin, we consider the cumulative network in which we aggregate over time, taking any edge that appears at least one throughout the 27 years. We find that 147 countries had at least one statistically significant relationship to be included in the network. Of the 17,030 possible international relationships, only 541 are positive interactions and 1538 are negative interactions. The positive recognition network is shown in Fig. 1B. Scientific publications from Switzerland are over-cited the most, with 36 incoming edges, followed by Great Britain, Germany, and the Netherlands (Fig. 2B). On the other hand, publications from China are the most under-cited, with 86 incoming undercitation edges, followed by Japan, Iran and India (Fig. 2C).

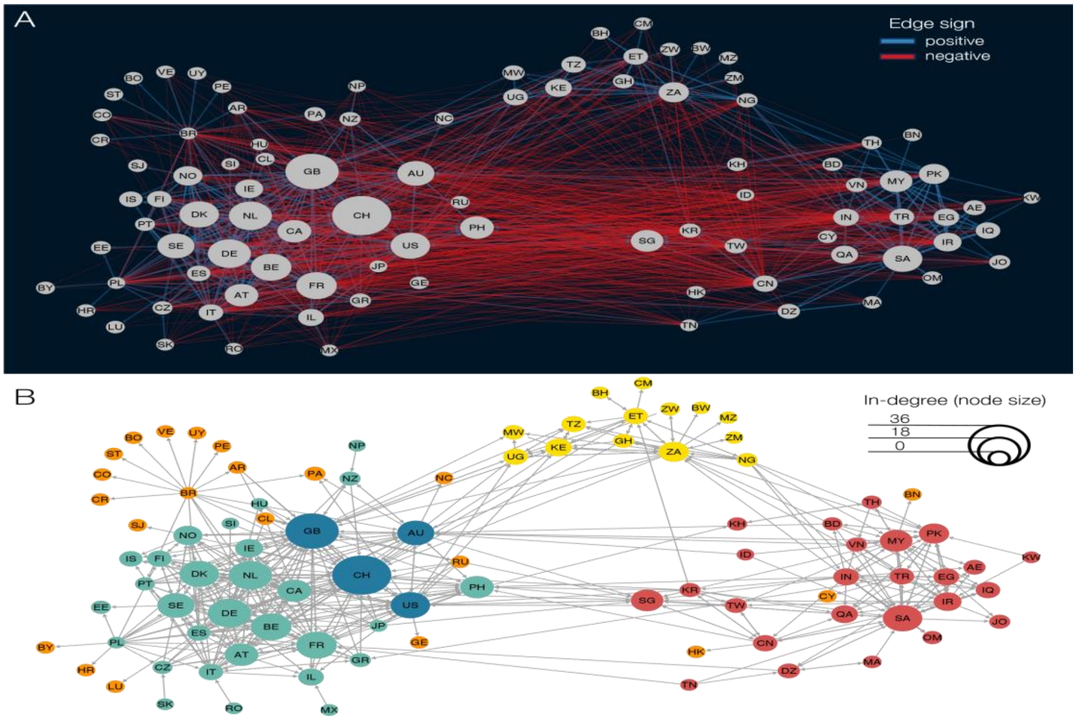

Figure 1. International network of scientific citation preferences. A) The full international network of scientific citation preferences. Edge color reflects positive (blue) or negative (red) citation preferences. B) The network filtered to positive relationships. The node size captures the country in-degree, while node colour reflects membership in one of five communities inferred using the degree-corrected stochastic block model. Node position is the same in both panels and was derived using only the positive relationships.

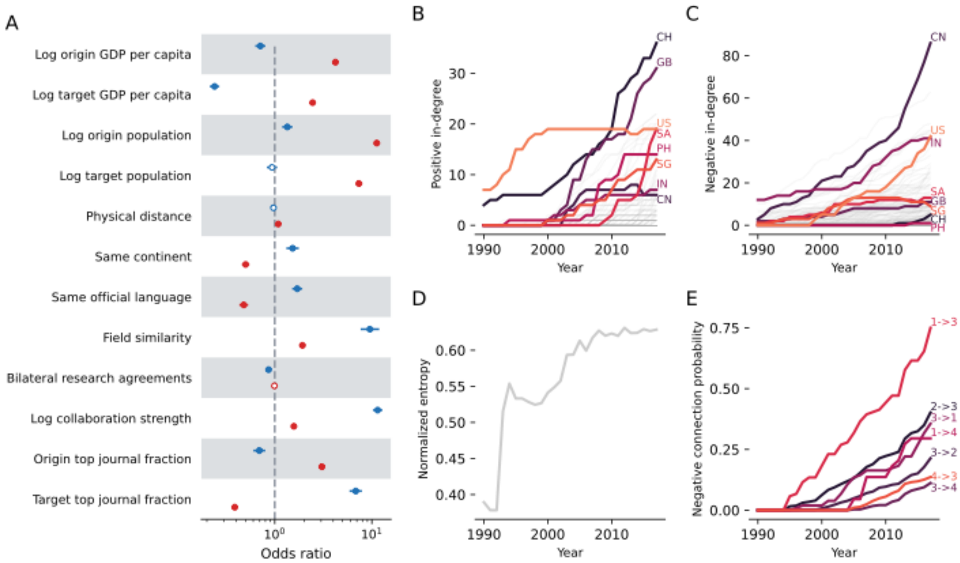

Figure 2. Properties of the international network of scientific citation preferences. (A) The odds ratios for a multinomial logit model with temporal fixed-effects to predict the positive (blue) or negative (red) citation preference compared to the baseline of no preference. Solid points are statistically significant at $p < 0.05$ with the 95% confidence intervals shown. The full regression table can be found in the SI, Table S3. The (B) positive and (C) negative in-degrees highlight 6 prominent countries, including the most positively viewed country in 2017, Switzerland (CH), and the most negatively viewed country, China (CN). (D) The normalized entropy for the distribution of PageRank centrality over the nodes has been increasing over the last 30 years. (E) The probability for a negative citation preference between a country in a source community and a country in a target community.

To identify key country-specific and country-pair-specific factors related to national citation preferences, we build a multinomial logit model with temporal fixed-effects to predict the citation preference between all pairs of countries from 1990 through 2017. We find that, while most independent variables play a statistically significant role in the prediction, many of them do not differentiate in terms of the contribution direction between positive and negative citation preferences (Fig. 2A and SI, Table S1). In particular, collaboration strength, while indicative of a link between countries, does not help differentiate the sign of that preference, and topical similarity only contributes to the prediction of positive preferences. However, three cultural indicators: common language ($\beta_{positive} = 0.53$, 95% $CI = [0.41, 0.65]$; $\beta_{negative} = -0.74$, 95% $CI = [-0.84, -0.63]$), same continent ($\beta_{positive} = 0.42$, 95% $CI = [0.27, 0.57]$; $\beta_{negative} = -0.69$, 95% $CI = [-0.78, -0.6]$), and participation in Science and Technology Agreements

(bilateral research agreement, $\beta_{positive} = -0.17$, 95% $CI = [-0.19, -0.15]$; $\beta_{negative} = -0.01$, 95% $CI = [-0.02, 0.0]$), relate to both the presence and sign of the national citation preference (Fig. 2A). Finally, we use the fraction of publications in top journals to capture one aspect of research quality (see Methods and SI, Section S1) and find that higher-quality publications in the target country are associated with a higher probability of a positive citation preference from the origin country, while lower-quality publications in the target country are associated with a negative citation preference ($\beta_{positive} = 1.51$, 95% $CI = [1.39, 1.63]$; $\beta_{negative} = -0.75$, 95%$CI = [-0.8, -0.69]$). Overall, this model suggests a mutual-influence relationship between scientific quality, national culture, science diplomacy and international scientific recognition.

Mapping the network of international preferences over time reveals the changing landscape of scientific diplomacy. Specifically, the network of international citation preferences has evolved away from a core-periphery structure dominated by a few hubs to a more distributed structure, a change which we measure by the increasing normalized entropy for the distribution of PageRank centrality scores (Fig. 2D). For example, before 2000, the network was dominated by the United States, with relatively little positive scientific recognition of countries in Asia or Africa (Fig. 2B). However, by 2010, Switzerland and Great Britain surpassed the United States in global recognition, and there were notable rises in recognition to Saudi Arabia, the Philippines, and Singapore (Fig. 2B). Throughout this period, China and Japan remained the most under-cited, dominating the negative citation preference network (Fig. 2C).

Growing international scientific fragmentation

The preference of some nations for the scientific work of others, combined with the proliferation of negative biases against groups of countries, is a characteristic hallmark of international scientific fragmentation (Aref et al., 2020). This pattern in citation patterns can stem from various factors, such as disciplinary biases, prevailing research trends, language barriers, geographical disparities, or ideological preferences. As a result, scientific fragmentation can distort the perception of research's importance and impact, reinforce existing knowledge gaps, and impede the equitable dissemination and recognition of diverse scientific contributions.

To measure the dynamics of international scientific fragmentation, we first detect the presence of international communities using the degree-corrected stochastic block model, finding strong evidence for a partition of the positive network in 5 distinct communities. Three blocks strongly resemble a three-layer core-periphery structure (Gallagher et al., 2021). Specifically, we find a dense core consisting of Western countries that tend to positively prefer each other's work (1, dark blue) and a weaker core of many European countries (2, light blue), while countries in the periphery (5, orange) are agnostic to each other, but prefer countries from both the weak and strong cores (Fig. 1).

At the same time, this analysis confirms that the core-periphery structure is an oversimplification of the diverse communities in global science. The international scientific recognition network reveals two additional communities outside of the Western scientific world: one community captures an international community predominately composed of Asian countries (3, red), including both East Asia and the Middle East, while another reflects the African nations (4, yellow).

The fragmentation of global science is evidenced by the distribution pattern of positive and negative citation preferences across scientific communities. Overall, only 34% of positive citation preferences occur between nations from different communities, whereas negative citation preferences predominantly cross community boundaries, with over 86% occurring between nations from different communities.
The structure of the international citation preference network and its communities provides a more nuanced view of the differing roles nations play in shaping global scientific recognition and knowledge dissemination. For example, while both Singapore and China have gained recognition for their scientific contributions (Zhou and Leydesdorff, 2006), our analysis shows that Singapore occupies a unique bridging role between different communities, whereas China, despite its prominence, remains within the Asian community without holding a central core position. Notably, our work highlights Saudi Arabia, Turkey, and Iran as occupying more central roles within the Asian scientific community. Similarly, South Africa (ZA) stands out as a central node within the African scientific community, while the network reveals the distinct roles of Uganda and Nigeria as key bridges—Uganda connecting to Western communities and Nigeria to the Asian community.
To assess the dynamics of international scientific fragmentation, we look at the probability of forming negative or positive links. Overall, we observe a growing tendency for nations to negatively judge the work of other nations as evidenced by the increase in negative connection probabilities (SI, Figure S1). However, the community structure of the international scientific recognition network reveals that these preferences are not evenly distributed and are not primarily directed at specific nations. Instead, the fragmentation of global science appears to be influenced by the detected geopolitical communities. As shown in Fig. 2E, the probability of inter-community negative preference links has grown significantly since 1990. The probability of negative inter-community links is largest between the Western and Asian communities, specifically communities 1−> 3 and 3− > 1 as well as 2−> 3 and 3−> 2, but has also significantly grown between the African and Western communities 1− > 4, 4− > 1 and the African and Asian communities 3− > 4, 4− > 3. Significantly, there are nearly symmetric negative inter-community link probabilities, indicating the true fragmentation of the global scientific landscape into distinct communities cannot be explained by a core-periphery model.

International recognition network and the diffusion of ideas

To explore the potential connection between the position of nations in the international scientific recognition network and the propensity for them to spread ideas, we investigate the flow of knowledge between countries. We operationalize scientific knowledge through the appearance of keywords in the title and abstract of scientific publications (Milojevic et al.´, 2011; Milojevic´, 2015; Cheng et al., 2023). Specifically, we identify the mention of over 40,000 n-grams defined as scientific ideas by a previous study (Cheng et al., 2023) and limit to 7,327 unique ideas originating in only one country after 2000 (see Methods and SI, Section S2). We then model the fraction of ideas originating in one country that are eventually mentioned in another target country at least once during the subsequent 22 years (2000-2022) using logistic regression. This approach allows us to gauge the spread of information through the global scientific ecosystem, reflecting the broader exchange of ideas without needing to follow each idea's continuous trajectory over time. Consequently, we use the cumulative international recognition network where we aggregate into a static snapshot using all links that appear in at least one time slice.

Since there are many factors which may influence the flow of knowledge between countries, in Model 1, we predict the fraction of ideas which spread between 9,635 country pairs based on the number of ideas which originate and terminate in the origin and target countries respectively, the countries' populations and GDP per capita. Unsurprisingly, the model coefficients in Table 1 show that the number of ideas originating in a country, the ability of a target country to take up ideas, and the country's population are all statistically significant. We also find that the topical distance between the countries' scientific publications and whether the origin and destination share a common language are also statistically significant in their relation to the spread of ideas.

Table 1. International diffusion of scientific ideas. Model coefficients for a series of logistic regression models to predict the fraction of scientific ideas that originate in one country that are eventually mentioned in the destination country. Confidence intervals in parentheses. Standard errors and p-values are reported.

	Dependent variable: Fraction of scientific ideas.			
	Model			
	(1)	(2)	(3)	(4)
Intercept	−1.08∗∗∗	−1.1∗∗∗	−1.08∗∗∗	−1.08∗∗∗
	(−1.13,−1.02)	(−1.16,−1.04)	(−1.15, −1.02)	(−1.15, −1.01)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0
Log Population origin	−0.38∗∗∗	−0.45∗∗∗	−0.44∗∗∗	−0.46∗∗∗
	(−0.44,−0.33)	(−0.51,−0.38)	(−0.51,−0.37)	(−0.53,−0.39)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0
Log Population destination	0.17∗∗	0.18∗∗	0.17∗∗	0.18∗∗
	(0.06,0.28)	(0.06,0.3)	(0.06,0.29)	(0.06,0.3)
	S.E. 0.06; p-v 0.0019	S.E. 0.06; p-v 0.0036	S.E. 0.06; p-v 0.0042	S.E. 0.06; p-v 0.0034
Log GDP per capita origin	−0.14∗∗∗	−0.21∗∗∗	−0.21∗∗∗	−0.25∗∗∗
	(−0.2,−0.09)	(−0.27,−0.15)	(−0.28,−0.15)	(−0.32,−0.18)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.04; p-v 0.0
Log GDP per capita destination	0.07	0.07	0.07	0.07
	(−0.05,0.19)	(−0.06,0.2)	(−0.06,0.2)	(−0.06,0.2)
	S.E. 0.06; p-v 0.2372	S.E. 0.06; p-v 0.2687	S.E. 0.07; p-v 0.2719	S.E. 0.07; p-v 0.2703
Number of ideas origin	0.16∗∗∗	0.19∗∗∗	0.19∗∗∗	0.18∗∗∗
	(0.1,0.21)	(0.13,0.25)	(0.12,0.25)	(0.11,0.24)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0
Number of ideas destination	0.81∗∗∗	0.82∗∗∗	0.83∗∗∗	0.83∗∗∗
	(0.69,0.94)	(0.68,0.96)	(0.69,0.97)	(0.68,0.97)
	S.E. 0.06; p-v 0.0	S.E. 0.07; p-v 0.0	S.E. 0.07; p-v 0.0	S.E. 0.08; p-v 0.0

Topic distance		−0.1** (−0.17,−0.03) S.E. 0.03; p-v 0.0046	−0.09** (−0.16,−0.03) S.E. 0.03; p-v 0.007	−0.09** (−0.16,−0.02) S.E. 0.03; p-v 0.0084
Log Physical distance		−0.01 (−0.07,0.05) S.E. 0.03; p-v 0.6742	0.01 (−0.05,0.07) S.E. 0.03; p-v 0.6855	0.01 (−0.05,0.07) S.E. 0.03; p-v 0.811
Common language		0.39*** (0.21,0.57) S.E. 0.09; p-v 0.0	0.34*** (0.15,0.53) S.E. 0.1; p-v 0.0004	0.32*** (0.13,0.51) S.E. 0.1; p-v 0.0008
Positive citation preference			0.4** (0.16,0.64) S.E. 0.12; p-v 0.001	0.32* (0.07,0.57) S.E. 0.13; p-v 0.0111
Negative citation preference			−0.23* (−0.41,−0.05) S.E. 0.09; p-v 0.0114	−0.22* (−0.4,−0.04) S.E. 0.09; p-v 0.0153
Network centrality origin				0.08* (0.01,0.16) S.E. 0.04; p-v 0.0218
Network centrality destination				0.01 (−0.05,0.08) S.E. 0.03; p-v 0.6583
Note:	*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$			
Observations	9635	8292	8292	8292
Pseudo-R^2	0.1652	0.1876	0.1924	0.1932
Log Likelihood	-4242.81	-3606.07	-3584.55	-3581.06
F statistic	243.97*** (d.f.=6.0)	148.93*** (d.f.=9.0)	123.34*** (d.f.=11.0)	104.5*** (d.f.=13.0)

The network of scientific recognition enhances our ability to predict the flow of ideas between countries, as shown in Models 3 and 4 (Table 1). The odds ratios suggest that a positive recognition edge between the target and originating countries leads to a 1.5 times increase in the fraction of ideas which spread between those countries compared to the baseline of no edge, while a negative recognition edge between the target and originating countries leads to 0.8x decrease in the fraction of ideas which spread between those countries. Beyond the immediate neighborhood, the global network topology is hypothesized to play a significant role in the spread of information over social networks (Kempe et al., 2005; Pei et al., 2018). We also find that the network centrality of the originating country is related to the diffusion of ideas (*p-value* < 0.0218; 95% IC = [0.01,0.16]) (see Table 1 for details).

Exploring the impact of journals on citation preferences

We now extend our analysis by introducing additional controls to further explore factors influencing citation preferences. Our framework seamlessly integrates a non-parametric approach that accounts for the field or journal in which each article is published, allowing us to control for variability in citation practices across disciplines and venues. By incorporating these controls and juxtaposing the new network against our original, this enhanced model provides a more refined understanding of how disciplinary and journal-specific effects interact with national-level citation behaviors, offering deeper insights into the structure of global scientific recognition.

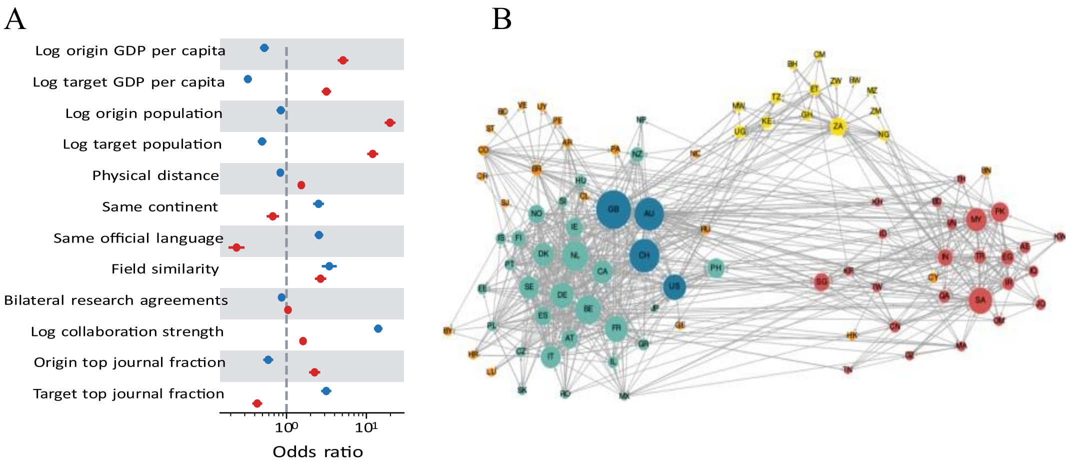

Figure 3. The international network of scientific citation preferences controlling for publication journal. (A) The odds ratios for a multinomial logit model to predict the positive (blue) or negative (red) citation preference compared to the baseline of no preference. Solid points are statistically significant at *p* < 0.05 with the 95% confidence intervals shown. The full regression table can be found in the SI, Table S4. (B) The journal bootstrap network filtered to positive relationships using the same layout as in Fig 1.

Instead of relying on the full citation distribution for all publications cited by the source country, we construct a new baseline citation distribution using a stratified bootstrap approach that accounts for journal frequency (see Methods for details). This technique samples from the source country's conditional citation distribution while ensuring the sampled set reflects the observed publication counts for each journal. By controlling for journal-level citation patterns– commonly used as proxies for scientific discipline and "quality"–this method provides a more detailed benchmark, isolating national citation preferences from journal-related con-founders. Shown in Fig. 3B, the resulting cumulative international network of citation preferences based on the journal bootstrap (N2) exhibits both notable similarities and differences when compared to the original network (N1). Specifically, N2 reveals more positive national preferences, with a total of 645 compared to 541 in N1, while it shows significantly fewer negative preferences, dropping from 1,538 in N1 to just 334 in N2. At the same time, there is considerable overlap between the networks: 448 positive preferences are present in both networks, accounting for 84% of the smaller N2, and 326 negative preferences are shared, representing 98% of the smaller N1. The variation in positive edges is largely concentrated in a small number of countries: 47% of the new edges are directed toward just 11 countries, while 30% originate from only 7 countries. Moreover, the edge distribution in N2 largely mirrors the community structure observed in N1 such that 60% of positive edges connect nations within the same community in N2, slightly down from 66% in N1, and 85% of negative edges link nations from different communities in N2, compared to 86% in N1. Using a similar multinomial logistic regression model with temperal fixed-effects to predict the presence and sign of national preferences, we find the same independent variables play remarkably similar patterns of importance for predicting the odds of a positive or negative edge, and differentiating between those signs (Fig. 3A).

Taken together, these observations suggest that about 80% of the negative citation preferences we initially identified can be attributed to disciplinary differences in scientific focus and journal "quality". At the same time, the increase in positive preferences primarily within the original communities, indicates the importance of those communities, suggesting they are highly influential in shaping collaborative networks and recognition. Ultimately, these findings emphasize the value of applying robust methodological frameworks to uncover the complexities of international citation preferences, providing deeper insights into the factors that influence scientific recognition on a global scale.

Discussion

The international scientific landscape, a complex and dynamic web of knowledge, people and practices, is molded by national interests grounded in historical events, cultural values, political agendas, economics, and technological innovations. These same forces shape interactions between nations through incentives for international collaboration, researcher mobility, and knowledge flows. By analyzing more than fifty-seven million scientific publications across 223 countries spanning the period 1990-2022, we provide a large-scale temporal and structural analysis of the collective structure of global scientific recognition. We find that the international citation preference network constructed from these publications is shaped by cultural elements, including language and political agreements, and augments insights from the study of scientific collaboration and scientific topics. Additionally, we quantify the network's departure from a core-periphery structure and identify five communities corresponding to major global regions, revealing a growing trend towards increased fragmentation. We then demonstrate that the international citation preference network imposes limitations on the dissemination of scientific ideas, reflecting a more efficient spread of concepts within a community compared to their transmission between distinct communities. Finally, we find that around 80% of negative citation preferences can be explained by disciplinary differences and journal "quality", while those same factors increase the prevalence of positive preferences within the original communities.

Our results reveal the collective structure of international citation preferences, complementing the viewpoints offered by collaboration, migration, and citation volume (Glanzel¨, 2001; Leydesdorff and Wagner, 2008; Wagner and Jonkers, 2017). While we were able to quantify the magnitude and significance of these preferences, and mapped how these preferences changed when controlling for scientific journals, but our data is unable to suggest all of causal mechanisms driving them. Additional work is needed to differentiate whether the observed patterns are rooted in social factors like cultural differences or accessibility. Nevertheless, the resulting model of global recognition reveals interesting features of the international state of scientific discourse.

Our quantitative results reveal the emergence of multiple distinct international communities, challenging the traditional core-periphery model of global science. These findings show that rather than a simple transition of countries from the periphery to the core, regional scientific communities are increasingly disconnected from each other. Notably, we identified negative citation links between these communities–evidence of declining mutual recognition–which would not be captured by a standard citation or collaboration network model. This suggests that these communities are growing apart, reinforcing their preference for internal knowledge sharing over external engagement. The implications of this are profound for the sociology of science and global science inequalities: instead of a unified global hierarchy, we may be witnessing the fragmentation of scientific influence, where certain regions strengthen internal ties at the cost of broader visibility and integration into the global scientific landscape. This deepens existing disparities, as regions that were once peripheral may develop more insular networks, further complicating efforts to address global inequalities in scientific recognition and collaboration.

The results of our study on the international scientific landscape carry several policy implications. Firstly, acknowledging the influence of national interests, historical events, cultural values, political agendas, economics, and technological innovations on global scientific recognition through citation suggests that the assessment of scientific impact to publications, authors, and organizations should also be sensitive to these multifaceted factors. It further suggests research into the implications of national vs international citation recognition on individual careers and potential inequalities in recognition that may arise (Huang et al., 2020). Secondly, the departure from a traditional core-periphery structure via the identification of five major global communities, underscores a growing trend towards increased fragmentation in scientific influence. Policymakers will need to adapt strategies to address this shift, ensuring inclusivity and collaboration across diverse scientific communities. Finally, the negative influences on national preferences of bilateral agreements for science and technology mirror results found for other types of treaties (Hoffman et al., 2022). This finding underscores that such agreements, which are intended to foster collaboration and knowledge exchange between nations, may encounter challenges that impede their effectiveness. Thus, they hamper an important tool that policymakers have to establish and nurture international scientific relationships, potentially hindering the full realization of the intended benefits of bilateral agreements in advancing global scientific cooperation.

Code and Data Availability

The primary dataset, OpenAlex, is freely available online at https://openalex.org/. All code used to conduct the analysis and generate the figures, as well as the processed data and network structure, is included as part of the pySciSci Python package (Gates and Barabasi´, 2023):
https://github.com/SciSciCollective/pyscisci/globalscience.

References

Abramo, G., C. A. D'Angelo, and F. Di Costa (2020, February). The role of geographical proximity in knowledge diffusion, measured by citations to scientific literature. *Journal of Informetrics 14*(1), 101010.

Adams, J. (1998). Benchmarking international research. *Nature 396*(6712), 615–618.

Adams, J. (2012). The rise of research networks. *Nature 490*(7420), 335–336.

Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics 56*(2), 235–246.

Althouse, B. M., J. D. West, C. T. Bergstrom, and T. Bergstrom (2009). Differences in impact factor across fields and over time. *Journal of the American Society for Information Science and Technology 60*(1), 27–34.

Anderson, W. (2018). Remembering the spread of western science. *Historical Records of Australian Science 29*(2), 73–81.

Aref, S., L. Dinh, R. Rezapour, and J. Diesner (2020). Multilevel structural evaluation of signed directed social networks based on balance theory. *Scientific Reports 10*(1), 15228.

Azoulay, P. and F. B. Lynn (2020). Self-citation, cumulative advantage, and gender inequality in science. *Sociological Science 7*, 152–186.

Basu, A., P. Foland, G. Holdridge, and R. D. Shelton (2018, October). China's rising leadership in science and technology: quantitative and qualitative indicators. *Scientometrics 117*(1), 249–269.

Bauder, H., O. Lujan, and C.-A. Hannan (2018). Internationally mobile academics: Hierarchies, hegemony, and the geo-scientific imagination. *Geoforum 89*, 52–59.

Cheng, M., D. S. Smith, X. Ren, H. Cao, S. Smith, and D. A. McFarland (2023, April). How new ideas diffuse in science. *American Sociological Review*, 000312242311669.

Chinchilla-Rodr´ıguez, Z., C. R. Sugimoto, and V. Lariviere (2019, June). Follow the leader:` On the relationship between leadership and scholarly impact in international collaborations. *PLoS One 14*(6), e0218309.

Choi, S. (2012, January). Core-periphery, new clusters, or rising stars?: international scientific collaboration among 'advanced' countries in the era of globalization. *Scientometrics 90*(1), 25–41.

Collyer, F. (2014, September). Sociology, sociologists and core-periphery reflections. *Journal of Sociology 50*(3), 252–268.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988, Sep). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics 44*(3), 837.

Fantom, N. and U. Serajuddin (2016). *The World Bank's classification of countries by income*. The World Bank.

Freeman, R. B. (2010, July). Globalization of scientific and engineering talent: international mobility of students, workers, and ideas and the world economy. *Economics of Innovation and New Technology 19*(5), 393–406.

Gallagher, R. J., J.-G. Young, and B. F. Welles (2021, March). A clarified typology of coreperiphery structure in networks. *Science Advances 7*(12), eabc9800.
Garfield, E. (2006). The history and meaning of the journal impact factor. *Jama 295*(1), 90–93.
Gates, A. J. and A.-L. Barabasi (2023). Reproducible science of science at scale:´ *pySciSci*. *Quantitative Science Studies*, 1–17.
Glanzel, W. (2001). National characteristics in international scientific co-authorship relations.¨ *Scientometrics 51*(1), 69–115.
Gomez, C. J., A. C. Herman, and P. Parigi (2022). Leading countries in global science increasingly receive more citations than other countries doing similar research. *Nature Human Behaviour 6*(7), 919–929.
Gong, K., J. Xie, Y. Cheng, V. Lariviere, and C. R. Sugimoto (2019). The citation advantage` of foreign language references for Chinese social science papers. *Scientometrics 120*(3), 1439–1460.
Gui, Q., C. Liu, and D. Du (2019). Globalization of science and international scientific collaboration: A network perspective. *Geoforum 105*, 1–12.
Galvez, A., M. Maqueda, M. Martinez-Bueno, and E. Valdivia (2000). Scientific publication´ trends and the developing world: What can the volume and authorship of scientific articles tell us about scientific progress in various regions? *American Scientist 88*(4), 526–533.
Hagstrom, W. O. (1974). Competition in science. *American Sociological Review 39*(1), 1–18.
Heimeriks, G. and R. Boschma (2014, March). The path- and place-dependent nature of scientific knowledge production in biotech 1986-2008. *Journal of Economic Geography 14*(2), 339–364.
Hoffman, S. J., P. Baral, S. Rogers Van Katwyk, L. Sritharan, M. Hughsam, H. Randhawa,
G. Lin, S. Campbell, B. Campus, M. Dantas, N. Foroughian, G. Groux, E. Gunn, G. Guyatt, R. Habibi, M. Karabit, A. Karir, K. Kruja, J. N. Lavis, O. Lee, B. Li, R. Nagi, K. Naicker, J.A. Røttingen, N. Sahar, A. Srivastava, A. Tejpar, M. Tran, Y.-q. Zhang, Q. Zhou, and M. J. P. Poirier (2022, August). International treaties have mostly failed to produce their intended effects. *Proceedings of the National Academy of Sciences 119*(32), e2122854119.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.
Huang, J., A. J. Gates, R. Sinatra, and A.-L. Barabasi (2020). Historical comparison of gender´ inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences 117*(9), 4609–4616.
Kempe, D., J. Kleinberg, and E. Tardos (2005). Influential nodes in a diffusion model for so-´ cial networks. In *Automata, Languages and Programming: 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005. Proceedings 32*, pp. 1127–1138. Springer.
King, D. A. (2004). The scientific impact of nations. *Nature 430*(6997), 311–316.
King, M. M., C. T. Bergstrom, S. J. Correll, J. Jacquet, and J. D. West (2017, Dec). Men set their own cites high: Gender and self-citation across fields and over time. *Socius 3*, 2378023117738903.
Kozlowski, D., V. Lariviere, C. R. Sugimoto, and T. Monroe-White (2022, January). Inter-` sectional inequalities in science. *Proceedings of the National Academy of Sciences 119*(2), e2113067119.

Langenhove, L. V. (2017). Tools for an eu science diplomacy. *Luxembourg: European Commission*.

Lariviere, V., K. Gong, and C. R. Sugimoto (2018). Citations strength begins at home. *Nature 564*(7735), S70–S70.

Leydesdorff, L., C. Wagner, H. W. Park, and J. Adams (2013). International collaboration in science: The global map and the network. *arXiv preprint arXiv:1301.0801*.

Leydesdorff, L. and C. S. Wagner (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics 2*(4), 317–325.

Marginson, S. (2022a, August). What drives global science? The four competing narratives. *Studies in Higher Education 47*(8), 1566–1584.

Marginson, S. (2022b). What drives global science? The four competing narratives. *Studies in Higher Education 47*(8), 1566–1584.

Marginson, S. and X. Xu (2023). Hegemony and inequality in global science: Problems of the center-periphery model. *Comparative Education Review 67*(1), 31–52.

May, R. M. (1997). The scientific wealth of nations. *Science 275*(5301), 793–796.

Mayer, T. and S. Zignago (2011). Notes on CEPII's distances measures: The geodist database. *CEPII Documentation*.

McKiernan, E. C., L. A. Schimanski, C. Munoz Nieves, L. Matthias, M. T. Niles, and J. P.˜ Alperin (2019). Use of the journal impact factor in academic review, promotion, and tenure evaluations. *Elife 8*, e47338.

Melitz, J. and F. Toubal (2014). Native language, spoken language, translation and trade. *Journal of International Economics 93*(2), 351–363.

Milojevic, S. (2015, October). Quantifying the cognitive extent of science.´ *Journal of Informetrics 9*(4), 962–973.

Milojevic, S., C. R. Sugimoto, E. Yan, and Y. Ding (2011, October). The cognitive structure´ of Library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology 62*(10), 1933–1953.

Moravcsik, M. J. (1985, March). Applied scientometrics: An assessment methodology for developing countries. *Scientometrics 7*(3-6), 165–176.

Nicolas Ruffin, N. R. and U. Schreiterer (2017). Case study. science and technology agreements¨ in the toolbox of science diplomacy: Effective instruments or insignificant add-ons? *ELCSID Working Paper, No.6*.

Pei, S., F. Morone, and H. A. Makse (2018). Theories for influencer identification in complex networks. *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks*, 125–148.

Peixoto, T. P. (2017). Nonparametric bayesian inference of the microcanonical stochastic block model. *Physical Review E 95*(1), 012317.

Prebisch, R. (1962). The economic development of Latin America and its principal problems. *Economic Bulletin for Latin America*.

Priem, J., H. Piwowar, and R. Orr (2022). OpenAlex: A fully open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Saha, S., S. Saint, and D. A. Christakis (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association 91*(1), 42.

Schott, T. (1988a). International influence in science: Beyond center and periphery. *Social science research 17*(3), 219–238.

Schott, T. (1988b, September). International influence in science: Beyond center and periphery. *Social Science Research 17*(3), 219–238.

Schott, T. (1998, August). Ties between center and periphery in the scientific world-system: Accumulation of rewards, dominance and self-reliance in the center. *Journal of World-Systems Research*, 112–144.

Scott, P. (2015). Dynamics of academic mobility: Hegemonic internationalisation or fluid globalisation. *European Review 23*(S1), S55–S69.

Seth, S. (2009, December). Putting knowledge in its place: Science, colonialism, and the postcolonial. *Postcolonial Studies 12*(4), 373–388.

Shang, J., J. Liu, M. Jiang, X. Ren, C. R. Voss, and J. Han (2018). Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering 30*(10), 1825–1837.

Shils, E. (1975). *Center and Periphery: Essays in Macrosociology*. University of Chicago Press.

Stringer, M. J., M. Sales-Pardo, and L. A. N. Amaral (2008). Effectiveness of journal ranking schemes as a tool for locating information. *PLoS One 3*(2), e1683.

Sumathipala, A., S. Siribaddana, and V. Patel (2004, December). Under-representation of developing countries in the research literature: Ethical issues arising from a survey of five leading medical journals. *BMC Medical Ethics 5*(1), 5.

Sun, X. and W. Xu (2014, Nov). Fast implementation of delong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Processing Letters 21*(11), 1389–1393.

Tickner, A. B. (2013, sep). Core, periphery and (neo)imperialist international relations. *European Journal of International Relations 19*(3), 627–646.

Urbinati, A., E. Galimberti, and G. Ruffo (2021). Measuring scientific brain drain with hubs and authorities: A dual perspective. *Online Social Networks and Media 26*, 100176.

Wagner, C. S. and K. Jonkers (2017, October). Open countries have strong science. *Nature 550*(7674), 32–33.

Wagner, C. S., H. W. Park, and L. Leydesdorff (2015, July). The continuing growth of global cooperation networks in research: A conundrum for national governments. *PLoS One 10*(7), e0131816.

Wuestman, M. L., J. Hoekman, and K. Frenken (2019). The geography of scientific citations. *Research Policy 48*(7), 1771–1780.

Zelnio, R. (2012, May). Identifying the global core-periphery structure of science. *Scientometrics 91*(2), 601–615.

Zhang, L., Z. Cao, Y. Shang, G. Sivertsen, and Y. Huang (2024). Missing institutions in openalex: possible reasons, implications, and solutions. *Scientometrics*, 1–23.

Zhou, P. and L. Leydesdorff (2006). The emergence of china as a leading nation in science. *Research policy 35*(1), 83–104.

Supplemental Information

S1 Top Journals

Assessing the quality of a publication by its venue, often the journal in which it is published is a common practice in academic research (Saha et al., 2003; McKiernan et al., 2019). This approach is based on the premise that the reputation and rigour of the peer-review process of a journal are indicative of the quality of the articles it publishes. Top journals are traditionally identified by their *impact factor*, the average number of citations to publications in that journal over a 2-year window, which is susceptible to temporal and disciplinary variations (Garfield, 2006; Althouse et al., 2009). To control for these, we focus on all publications in each journal and field (OpenAlex Level 1) and find the number of citations to the publication over 5 years (c_5). We then leverage that the journal-specific citation distributions are log-normal (Stringer et al., 2008), and rank each journal in each field and each year by the mean log number of citations over 5 years. Finally, we take the top 50 journals in each field and each year, giving the set of top journals.

Using the yearly top journal set, we identify the fraction of publications from each country in the top journals, and normalize by the overall fraction of the global publications in these top journals (this quantity was decreasing over the time period considered).

S2 Identifying Scientific Ideas

To identify scientific ideas, we follow the methodology introduced in Cheng et al. (2023) (Cheng et al., 2023). We begin by pre-processing OpenAlex texts in several ways. First, we generate our input corpus by combining the abstract and title of each OpenAlex article. Then we remove the last sentence of an abstract if it contains copyright information. Next, we lowercase the text, remove digits, and replace punctuation except commas and periods with spaces. Finally, we use Porter lemmatization on the corpus for all words longer than five characters to collapse different variations of the same word (e.g., singular versus plural forms).

We then identify all publications that mention at least one of the ideas from the master list of 46,535 scientific ideas derived by Cheng et al. using a data-driven phrase segmentation algorithm, AutoPhrase (Shang et al., 2018). This results in a corpus of 1,191,364 publications from 221 countries. Next, we post-process these ideas, removing cases that were first mentioned before 2000 and focusing only on those ideas that were mentioned by only one country in their first year of usage, resulting in 7,327 unique ideas mentioned in 202,932 publications. Finally, we derive a dyadic variable denoting the fraction of ideas whose first usage was in the Origin country and then were later used in the Destination Country sometime between 2000 and 2022.

S3 International citation preference network

Table S1. Fixed-effect multinomial logit regression for 1990-2017. Model coefficients labelled by *p*-value. Standard errors in parentheses.

	Dependent variable: Citation preference					
	Model					
	(1)	(2)	(3)	(4)	(5)	(6)
Citation Preference : Positive						
Intercept	−15.92***	−16.64***	−17.22***	−17.4***	−13.55***	−15.57***
	(−16.61,−15.23)	(−17.34,−15.93)	(−17.99,−16.44)	(−18.18,−16.62)	(−14.34,−12.76)	(−16.4,−14.73)
	S.E. 0.35; p-v 0.0	S.E. 0.36; p-v 0.0	S.E. 0.4; p-v 0.0	S.E. 0.4; p-v 0.0	S.E. 0.4; p-v 0.0	S.E. 0.43; p-v 0.0
Log origin GDP per capita	2.27***	2.05***	1.68***	1.71***	−0.56***	−0.34***
	(2.2,2.33)	(1.98,2.12)	(1.61,1.75)	(1.64,1.78)	(−0.67,−0.45)	(−0.46,−0.22)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.06; p-v 0.0	S.E. 0.06; p-v 0.0
Log target GDP per capita	2.07***	1.89***	1.46***	1.5***	−0.79***	−1.42***
	(2.01,2.13)	(1.82,1.95)	(1.39,1.53)	(1.43,1.57)	(−0.9,−0.69)	(−1.53,−1.31)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.05; p-v 0.0	S.E. 0.06; p-v 0.0
Log origin population	2.17***	2.36***	2.19***	2.26***	0.14*	0.27***
	(2.11,2.23)	(2.29,2.42)	(2.12,2.26)	(2.19,2.33)	(0.03,0.24)	(0.16,0.38)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.05; p-v 0.0105	S.E. 0.06; p-v 0.0
Log target population	1.97***	2.14***	1.91***	1.97***	−0.22***	−0.06
	(1.91,2.02)	(2.08,2.2)	(1.84,1.97)	(1.9,2.04)	(−0.32,−0.11)	(−0.17,0.05)
	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.05; p-v 0.0001	S.E. 0.06; p-v 0.3037
Physical distance		−0.79***	−0.56***	−0.55***	−0.03	−0.03
		(−0.84,−0.75)	(−0.6,−0.51)	(−0.6,−0.51)	(−0.08,0.03)	(−0.09,0.02)
		S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.03; p-v 0.3344	S.E. 0.03; p-v 0.2762
Same continent		0.07	0.11	0.14*	0.32***	0.42***
		(−0.04,0.18)	(−0.01,0.23)	(0.02,0.26)	(0.18,0.46)	(0.27,0.57)
		S.E. 0.06; p-v 0.2189	S.E. 0.06; p-v 0.0769	S.E. 0.06; p-v 0.0264	S.E. 0.07; p-v 0.0	S.E. 0.08; p-v 0.0
Same official language			1.25***	1.2***	0.61***	0.53***
			(1.14,1.35)	(1.1,1.31)	(0.49,0.72)	(0.41,0.65)
			S.E. 0.05; p-v 0.0	S.E. 0.05; p-v 0.0	S.E. 0.06; p-v 0.0	S.E. 0.06; p-v 0.0
Field similarity			3.19***	3.21***	1.63***	2.14***
			(2.99,3.39)	(3.01,3.41)	(1.43,1.82)	(1.93,2.35)
			S.E. 0.1; p-v 0.0	S.E. 0.1; p-v 0.0	S.E. 0.1; p-v 0.0	S.E. 0.11; p-v 0.0
Bilateral research agreements				−0.1***	−0.18***	−0.17***
				(−0.12,−0.08)	(−0.2,−0.16)	(−0.19,−0.15)
				S.E. 0.01; p-v 0.0	S.E. 0.01; p-v 0.0	S.E. 0.01; p-v 0.0
Log collaboration strength					2.73***	2.57***
					(2.62,2.85)	(2.45,2.69)
					S.E. 0.06; p-v 0.0	S.E. 0.06; p-v 0.0
Origin top journal fraction						−0.29***
						(−0.4,−0.18)
						S.E. 0.06; p-v 0.0
Target top journal fraction						1.51***
						(1.39,1.63)
						S.E. 0.06; p-v 0.0
Citation Preference : Negative						
Intercept	−14.79***	−14.49***	−11.41***	−11.38***	−10.5***	−10.63***
	(−15.09,−14.5)	(−14.79,−14.2)	(−11.72,−11.11)	(−11.69,−11.07)	(−10.81,−10.18)	(−10.96,−10.3)
	S.E. 0.15; p-v 0.0	S.E. 0.15; p-v 0.0	S.E. 0.16; p-v 0.0	S.E. 0.16; p-v 0.0	S.E. 0.16; p-v 0.0	S.E. 0.17; p-v 0.0
Log origin GDP per capita	2.64***	2.56***	2.48***	2.48***	2.1***	1.43***
	(2.6,2.69)	(2.52,2.61)	(2.43,2.53)	(2.43,2.53)	(2.04,2.16)	(1.37,1.5)
	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0
Log target GDP per capita	1.14***	1.13***	0.95***	0.94***	0.57***	0.89***
	(1.11,1.17)	(1.1,1.16)	(0.92,0.99)	(0.91,0.98)	(0.52,0.63)	(0.84,0.95)
	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0

	(1)	(2)	(3)	(4)	(5)	(6)
Log origin population	2.66***	2.6***	2.53***	2.51***	2.13***	2.17***
	(2.62,2.7)	(2.56,2.64)	(2.48,2.57)	(2.47,2.55)	(2.08,2.19)	(2.11,2.23)
	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0
Log target population	2.44***	2.41***	2.25***	2.23***	1.86***	1.79***
	(2.4,2.48)	(2.37,2.45)	(2.21,2.29)	(2.19,2.28)	(1.8,1.91)	(1.73,1.85)
	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0
Physical distance		−0.19***	−0.07***	−0.07***	0.03	0.08***
		(−0.22,−0.15)	(−0.11,−0.04)	(−0.11,−0.04)	(−0.01,0.06)	(0.05,0.12)
		S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.1488	S.E. 0.02; p-v 0.0
Same continent		−0.72***	−0.71***	−0.7***	−0.7***	−0.69***
		(−0.8,−0.64)	(−0.79,−0.63)	(−0.79,−0.62)	(−0.78,−0.61)	(−0.78,−0.6)
		S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0	S.E. 0.04; p-v 0.0
Same official language			−0.63***	−0.61***	−0.75***	−0.74***
			(−0.73,−0.53)	(−0.71,−0.51)	(−0.85,−0.65)	(−0.84,−0.63)
			S.E. 0.05; p-v 0.0	S.E. 0.05; p-v 0.0	S.E. 0.05; p-v 0.0	S.E. 0.05; p-v 0.0
Field similarity			0.73***	0.73***	0.53***	0.62***
			(0.67,0.79)	(0.67,0.79)	(0.47,0.6)	(0.56,0.69)
			S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0	S.E. 0.03; p-v 0.0
Bilateral research agreements				0.01*	−0.01	−0.01
				(0.0,0.02)	(−0.02,0.0)	(−0.02,0.0)
				S.E. 0.01; p-v 0.0488	S.E. 0.01; p-v 0.0708	S.E. 0.01; p-v 0.1369
Log collaboration strength					0.4***	0.48***
					(0.37,0.44)	(0.44,0.52)
					S.E. 0.02; p-v 0.0	S.E. 0.02; p-v 0.0
Origin top journal fraction						0.88***
						(0.82,0.94)
						S.E. 0.03; p-v 0.0
Target top journal fraction						−0.75***
						(−0.8,−0.69)
						S.E. 0.03; p-v 0.0
Observations	930798	844882	744760	744760	744760	744760
Pseudo R^2	0.5613	0.5793	0.5954	0.5963	0.6198	0.6386
Log Likelihood	-32842.41	-30453.84	-28640.06	-28580.11	-26918.1	-25582.81
LLR χ^2	84024.71*** (d.f.=62.0)	83875.63*** (d.f.=66.0)	84307.51*** (d.f.=70.0)	84427.42*** (d.f.=72.0)	87751.43*** (d.f.=74.0)	90422.0*** (d.f.=78.0)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: *p < 0.05; **p < 0.01; ***p < 0.001

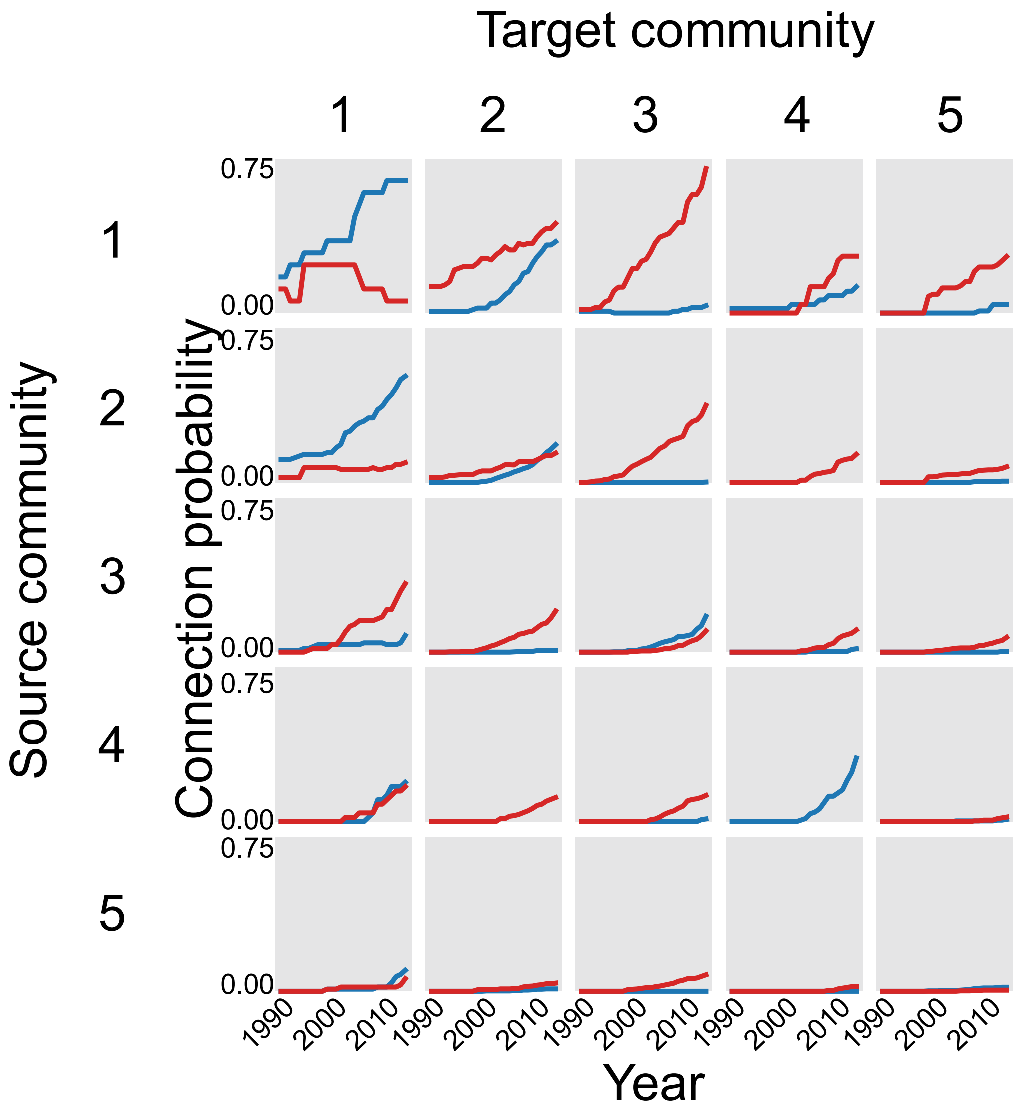

Figure S1. International network fragmentation. The probability of a positive (blue) or negative (red) directed edge from a country in the source community (rows) to a country in the target community (columns) from 1990 until 2017 (x-axis).