Kin of coauthorship in five decades of health science literature

Mattia Prosperi[a,b,c,1,2], Iain Buchan[c], Iuri Fanti[d], Sandro Meloni[e,f,1], Pietro Palladino[g], and Vetle I. Torvik[h]

[a]Department of Epidemiology, College of Public Health and Health Professions, University of Florida, Gainesville, FL 32610; [b]Department of Epidemiology, College of Medicine, University of Florida, Gainesville, FL 32610; [c]Centre for Health Informatics, Institute of Population Health, University of Manchester, Manchester M13 9PL, United Kingdom; [d]Infectious Diseases Clinic, Catholic University of the Sacred Heart, Rome 00168, Italy; [e]Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza 50018, Spain; [f]Department of Theoretical Physics, University of Zaragoza, Zaragoza 50009, Spain; [g]Genomics, Genetics and Biology Innovation Pole, Perugia 06134, Italy; and [h]Graduate School of Library and Information Science, University of Illinois at Urbana–Champaign, Champaign, IL 61820

Edited by Yu Xie, Princeton University, Princeton, NJ, and approved June 7, 2016 (received for review September 4, 2015)

Family background—kinship—can propagate careers. The evidence for academic nepotism is littered with complex associations and disputed causal inferences. Surname clustering, albeit with very careful consideration of surnames' flows across regions and time periods, can be used to reflect family ties. We examined surname patterns in the health science literature, by country, across five decades. Over 21 million papers indexed in the MEDLINE/PubMed database were analyzed. We identified relevant country-specific kinship trends over time and found that authors who are part of a kin tend to occupy central positions in their collaborative networks. Just as kin build potent academic networks with their own resources, societies may do well to provide equivalent support for talented individuals with fewer resources, on the periphery of networks.

social capital | kinship | health science literature | PubMed | demographic model

Bellow's book *In Praise of Nepotism: A Natural History* (1) was a controversial review of family enterprises through history, from King David to George W. Bush.

Critics of Bellow's positive stance pointed out that kinship in politics and business can lead to corruption and stagnation (2). In academia, nepotism has been blamed for poor graduate career support, gender inequality, and emigration of the intelligentsia (3–6). In economics, building on close relatives' experiences and networks has been analyzed as an expression of "social capital," without the negative connotations of nepotism (7, 8). In health science, Chervenak and McCullough (9) observed that opposition to nepotism is largely intuitive and not all instances of nepotism are ethically unjustified. With respect to professional outcomes, Pinchot et al. (10) investigated whether the children of surgeons were more likely to pursue a surgical career than their peers: they found a positive association with the career intentions of medical students, but not with the actual career paths of young clinicians after residency.

The evidence for academic nepotism is littered with complex associations and disputed causal inferences. Allesina (11) reported an unnatural scarcity of distinct surnames among tenured faculties in Italy. Ferlazzo and Sdoia (12) repeated the same analysis among professors in the United Kingdom, sustaining a more objective expression of social capital. Durante et al. (13) used readership of nonsport newspapers as a proxy for educational social capital when looking at kinship (*SI Appendix, Supplementary Background*).

Albeit with very careful consideration of surnames' distributions and flows across regions and time periods, surname clustering can be used to reflect family ties or kinship, and interpreted in relation to measures of social capital (including corruption, income inequality, and scientific output).

In the present study, we examined coauthorship surname patterns in the health science literature over five decades, by country. We used the 2013 MEDLINE/PubMed database, indexing over 21 million papers (*SI Appendix*, Fig. S2). The database is available on request to the provider, subject to acceptance of a license agreement.

Results

We considered papers with two or more authors with a known affiliation (which is a free text field) and an encoded country of publication (n = 11,910,186 and 13,945,281 respectively, of a total of 21,507,644). The country of affiliation, which was inferred using MapAffil (*SI Appendix, Materials and Methods*), a robust classification algorithm (14, 15), was mildly but significantly correlated with the country of publication (48% match overall, $P < 0.0001$; *SI Appendix*, Table S1). Using the country of affiliation, nine countries—United States, Japan, United Kingdom, Germany, France, Italy, Canada, Spain, and The Netherlands—gave rise to 70% of articles, and 90% came from 25 countries.

We defined kinship of a PubMed paper as the occurrence of the same surname more than once in the list of authors. However, this definition has been shown to be flawed and prone to surnames' demographic distributions/flows (11–13). To account for these potential biases, three independent surname filtering procedures were defined and compared: (*i*) excluding surnames with high frequency in the database; (*ii*) excluding surnames based on high frequency in a single country or continental area, using an external resource (another database); and (*iii*) filtering surnames based on the probability that they would appear by chance twice or more in an article, by conditioning on the country of the first/senior author, the calendar year, and the number of authors. For method *i*, we counted all distinct surnames in the PubMed database and ordered them by frequency, leaving out those with a frequency above the

Significance

Family background—kinship—can propagate careers. Five decades of health science literature worldwide show that, among family-tied authorships, there are country-specific patterns of publication evolving over time; authors who are part of a kin tend to form distinctive collaborative networks. A certain level of kinship may have beneficial effects on the research outputs of a country, whereas greater or lesser amounts of kinship could have adverse effects. It is perhaps more important for nations to promote equal opportunities in academic careers than to attempt to contain nepotism. Just as kin build potent academic networks with their own resources, societies may do well to provide equivalent support for talented individuals with fewer resources, on the periphery of networks.

Author contributions: M.P. and I.B. designed research; M.P., I.F., S.M., and P.P. performed research; M.P., I.F., S.M., P.P., and V.I.T. analyzed data; V.I.T. provided additional expertise and a tool for classifying free-text PubMed affiliations; and M.P., I.B., S.M., and P.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]M.P. and S.M. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: m.prosperi@phhp.ufl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1517745113/-/DCSupplemental.

SOCIAL SCIENCES

50th percentile. For *ii*, we crawled Wikipedia webpages listing most common surnames per country or continental areas, identifying 2,735 distinct entries; then all PubMed surnames matching this list were excluded. For *iii*, we implemented a Monte Carlo sampling of the PubMed data generating a number of fictional articles in a 5:1 ratio with respect to the original, nearly a hundred million records. A fictional article was generated sampling year, country, and number of authors from a real PubMed article and then extracting a random list of authors from all PubMed papers with the same year and country. For each surname, we counted the frequency of being found at least twice in each of the fictional articles, and this was used as a proxy for the conditional probability (*SI Appendix, Materials and Methods*).

We calculated the unfiltered (raw) proportion where the same surname appeared more than once in the author list, i.e., the kinship, for the PubMed data, and we repeated the procedure on the Monte Carlo fictional data. Prevalence of kinship was stratified by vigintile year and country. Then we adjusted the kinship proportions by using the filters above described, i.e., (*i*) frequency, (*ii*) Wikipedia, and (*iii*) Monte Carlo. The filtering was operated at the article level, considering first the surnames appearing twice or more, if any, or the first authors' surname (*SI Appendix, Probability of Kinship*). After applying the three filtering procedures, the number of retained records was n = 5,926,296 using the frequency (threshold on median frequency), n = 9,483,133 using the Wikipedia flag, and n_1 = 7,836,755, n_2 = 10,828,449 using the Monte Carlo year-country-conditioned probability (excluding records with $P > 0$ or $P > 0.0001$, respectively). There was a fair level of concordance in kinship prevalence between the three filtering procedures (*SI Appendix*, Table S2 and Figs. S3 and S4). The stratified proportions as estimated by Monte Carlo filter at $P > 0.0001$ were highly correlated with the Wikipedia ones ($R = 0.99$) and almost equivalent to the prevalence obtained by subtracting the kinship proportion of the shuffled data from that of the real data (Pearson's correlation, R = 0.95). The correlation decreased significantly when comparing them against the frequency filter, the unfiltered (whole), and the shuffled data ($R = 0.92$, 0.88, and 0.41, respectively).

Fig. 1 (red plots) shows the prevalence of kinship through time by country of affiliation (the 16 countries publishing the most papers), using the Monte Carlo filter with $P > 0.0001$ probability and by further stratifying the data with the consideration of papers with only four or five authors (blue plots, see the next paragraphs). The worldwide trend, increasing by 2% from the early 1970s, was driven by United States. Looking at other country-specific trends, the United Kingdom and The Netherlands had consistently lower kinship than the global average, and the same was found to a lesser extent with Canada, Australia, and Switzerland (and Soviet Union before 1992); Germany followed the trend of the Federal Republic before unification, in line with the world's trend; Japan exhibited a slight increase in the last decade; Russia, starting from an overall low kinship prevalence before 1992 (as Soviet Union), showed a steeper increase after the dissolution of Soviet Union; and France and Spain showed a counter trend over time, more pronounced in the former. The countries with the steepest rises in kinship were Italy, India, and Poland; however, India started from a proportion doubling the worldwide average before the 1980s and then kept it more or less constant over the subsequent years. Italy, starting from values below the world's average, exceeded it in the 1980s, whereas Poland did so in the 1990s. Among the countries not displayed in Fig. 1, Argentina, Belgium, Czech Republic, Denmark, Finland, and Norway showed decreasing kinship trends; Brazil showed a substantial above-average prevalence, and Greece had an increasing trend. The estimates obtained using the Wikipedia filter did not diverge significantly from the Monte Carlo estimation. Also, using the country of publication, even if the absolute numbers were different, the country-specific trends were maintained (*SI Appendix*, Figs. S5 and S6).

We repeated the whole analysis by selecting papers with a constant number of authors. Worldwide and country-specific kinship trends were flattened for papers composed by only one or two authors. By selecting papers with only 4 or 5 authors, where the overall median number of authors (interquartile range) was 4 (3–6) and the average was 4.6, the worldwide trend was constant over time, close to 3%; therefore, the observed world's increase was likely due to the number or authors. Nonetheless, the country-specific kinship trends were in general consistent with the all-authors analysis, e.g., lower for The Netherlands, higher for India, increasing for Italy/Poland (less pronounced for Italy), and decreasing for France/Spain (Fig. 1, blue plots; *SI Appendix*, Fig. S6); the slight increase of Japan was not found.

By fitting a multivariable main-effect linear model we found that journal impact factor (based on the 2012 Thomson Reuter's Journal Citation Reports) was associated with higher prevalence of kinship and that the greater the number of authors in the paper, the greater the prevalence of kinship (*SI Appendix*, Table S3); we therefore sought to investigate more in detail these factors, stratifying the multivariable modeling per country. Fig. 2*A* shows how a more recent year of publication was positively associated to a higher number of authors in a paper. Also, from Fig. 2*A*, it can be seen that the number of authors between 1983 and 1995 does not go beyond 10, the maximum number allowed by PubMed in that period. Before 1983, no limit was imposed, whereas between 1996 and 2000, the limit was 25, and after 2000, the limit was lifted. Beginning in 2005, editing of authors' list in older papers was allowed, regardless previous restrictions. Country-wise, the number of authors was associated with higher odds of kinship (Fig. 2*B*), being less pronounced in Italy, France, and Germany (after 1990), and more relevant in Soviet Union, India, and East Germany. In general, the difference in the median number of authors between kinship and nonkinship papers, given that both increase over time, tends to be larger in more recent calendar years (*SI Appendix*, Fig. S7). The journal's impact factor exhibited a more diverse strength of association with kinship (Fig. 2*D*): higher impact factor was negatively associated with kinship in Italy, India, The Netherlands, Unites States, Spain, United Kingdom, and Germany, whereas a positive association was not clearly found for other countries. Finally, as outlined also in Fig. 1, a more recent calendar year (Fig. 2*C*) showed higher odds for kinship in Russia/Soviet Union, Poland, Italy, and East Germany and an opposite trend for France, Spain, Switzerland, Germany (after 1990), Canada, and The Netherlands. The per-year higher odds were confirmed when using the subset of papers with four to five authors (significant at the 0.0001 level for the aforementioned countries). We also repeated the multivariable analysis by taking into account specific time periods, i.e., 1983–1995, 1996–2000, and after 2000, according to PubMed's restrictions on authorship, and we found odds ratios consistent with the main analysis (*SI Appendix*, Table S4).

To explore the association between publication kinship and the "fairness" in academic research that might be inherited by wider social factors, we used the "corruption perceptions index," estimated annually since 1995 by Transparency International, and the Gini income inequality index. We found that low perceived corruption was strongly associated to low kinship (overall, Pearson's $R = 0.73$, $P < 0.0001$), confirmed using the country of publication, but the correlation with the Gini income inequality index (which was calculated as the World's Bank, before and after taxes) was lower and varied in dependence of the measure used ($R = 0.33$–0.22, $P < 0.0001$–0.017); of note, it was higher (up to $R = 0.53$) when using the country of publication. The overall estimates may be heavily biased by the nonuniform missing values in time periods (*SI Appendix*, Fig. S8).

To gain more insights about the effects of kinship in the structure of the health science community, we analyzed the collaborative networks formed by researchers from the top 25 publishing countries (90% of the articles). Two scientists were linked by coauthorship

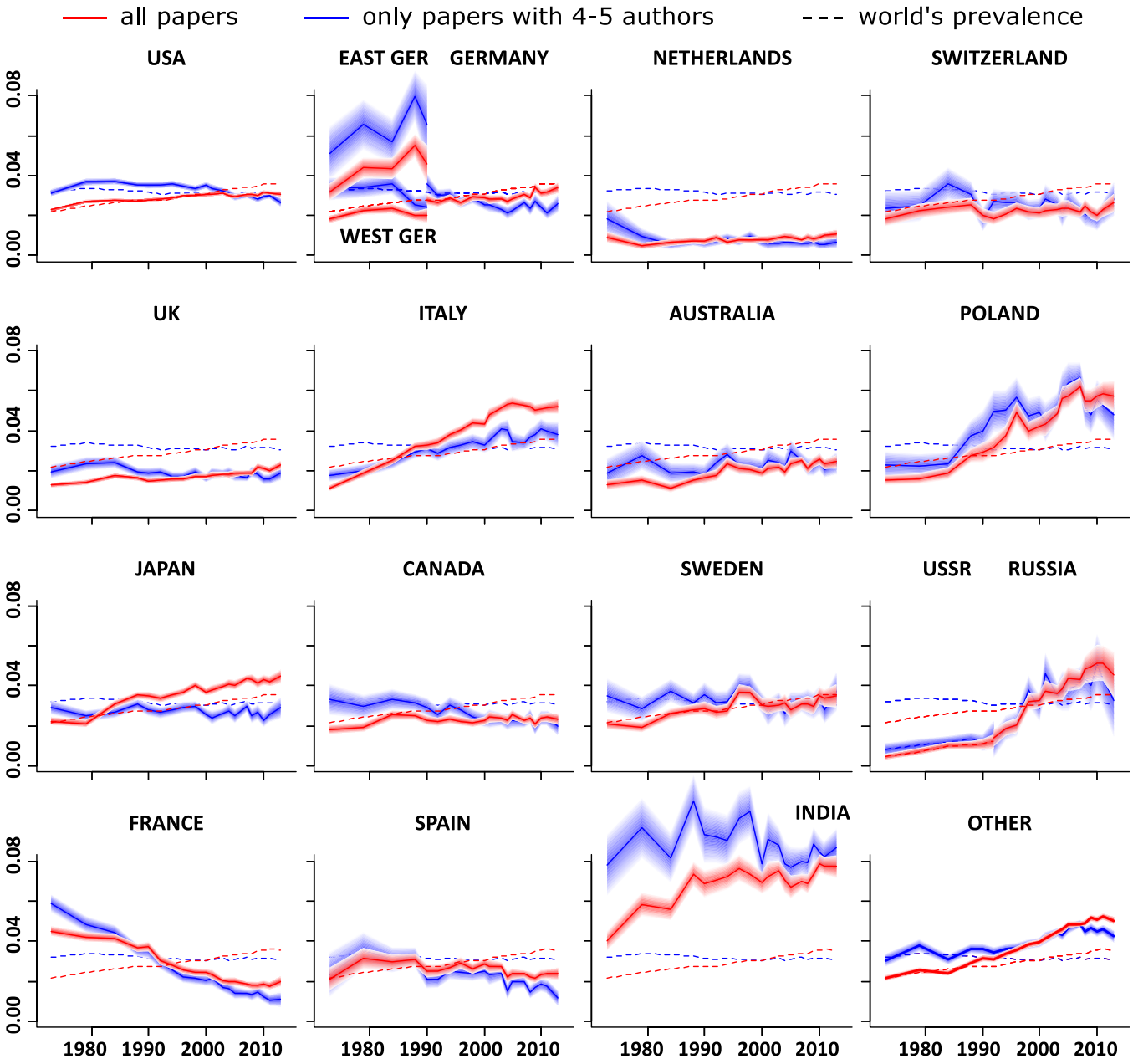

Fig. 1. Prevalence of kinship through calendar years by country estimated on MEDLINE/PubMed article records. Top countries in terms of publication output are shown in order from top to bottom, and then from left to right; countries are assigned on the basis of the first/senior author's affiliation. Kinship is defined as the presence of at least two identical surnames in a list of coauthors of the same paper. Data were prefiltered according to the potential demographic bias of surnames (expressed as probability of unrelated surname sharing conditioned on calendar year and country) using a Monte Carlo simulation. Shaded areas represent 95% CIs. Calendar year estimates are in red color for the full set of papers and blue when considering the subset of papers with only four to five authors. The dashed lines represent worldwide yearly averages (red for the full set of papers, blue for the subset of papers with only four to five authors).

in at least one paper. For each author, we flagged his surname if it was part of a kin or not, using the Wikipedia flag, as it is time-independent (*SI Appendix*, Tables S4 and S5 and Fig. S9). The topological analysis of the network highlighted significant differences between authors of a kin and those of nonkin. Although the distribution of connections for both types of authors is highly heterogeneous and in line with other coauthorship and collaboration networks (16), kin authors seems to have a higher number of collaborators and tend not to be placed randomly in the network (*SI Appendix*, Figs. S10–S13). The bootstrap analysis in Fig. 3*A* confirms that not only were kin authors characterized by a larger number of connections with respect to nonkin ones, but that they also occupied central places in the network, as they were characterized by a higher betweenness centrality, a measure of the fraction of shortest paths passing through a node (17, 18). With some speculation, we can say that authors of a kin occupy "important" positions. This suggestion is also supported by the k-core decomposition of the network (16), which demonstrated how kin authors occupied innermost parts of the network (higher core number) with an average core number of 15.6 respect to an average of 8.17 of rest of the authors (two-sample Kolmogorov–Smirnoff test, $D = 0.259$, $P < 0.0001$). Strong differences between kin and nonkin authors were found not only in the number of coauthors but also in the structure of collaborations itself. Specifically, coauthorship and social networks in

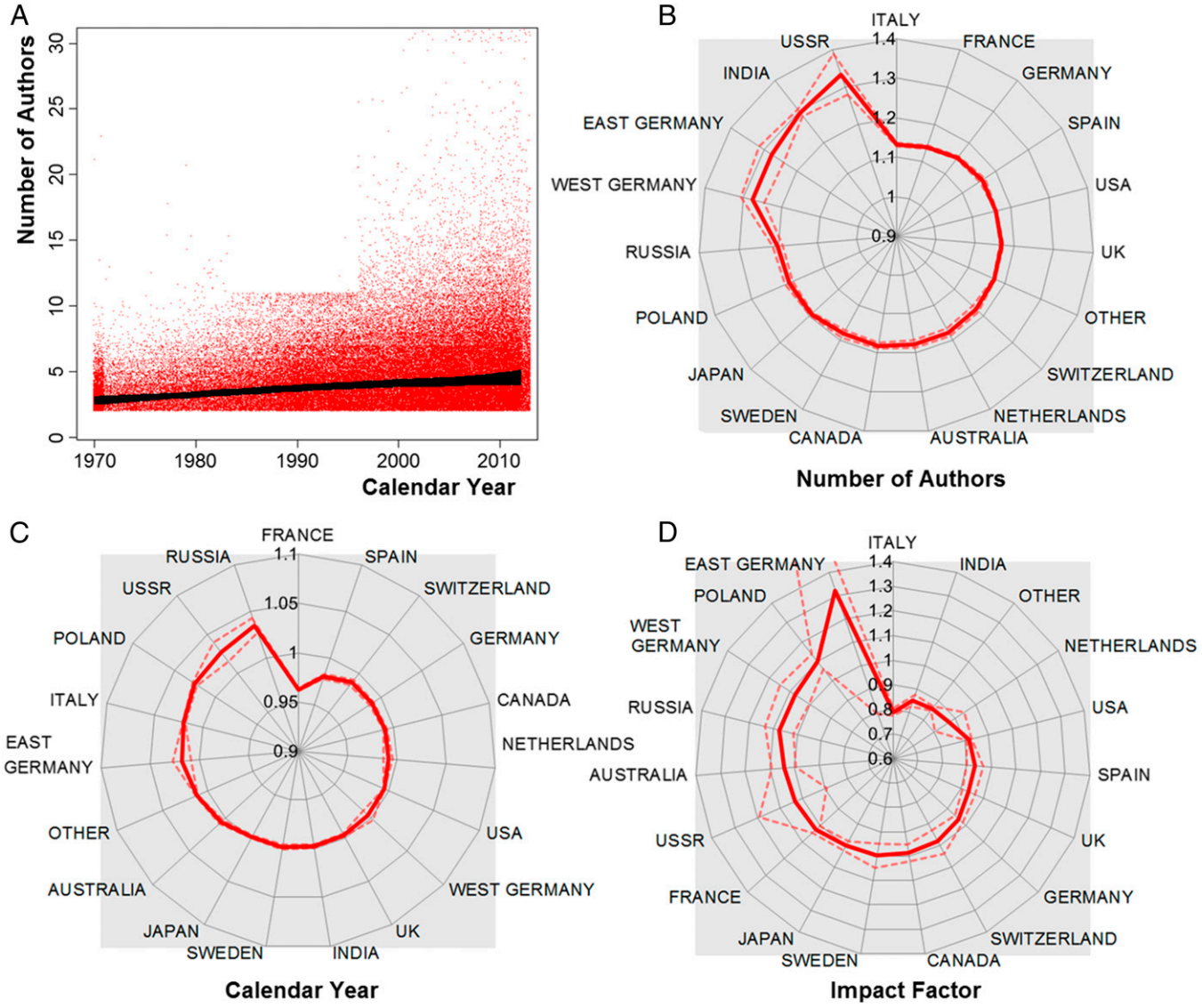

Fig. 2. Article features and their association with kinship. (*A*) Smoothing estimator of the distribution of the number of authors per calendar year. (*B–D*) Radar plots of the strength of association with kinship (odds ratio, with 95% CIs) for number of authors, calendar year, and impact factor, respectively.

general are characterized by a high clustering coefficient (triadic relationships between individuals) accounting for cohesive groups of authors, e.g., scientists in the same research group (19, 20). On the contrary, kin nodes showed a sensibly lower clustering coefficient if compared with the rest of the network, with an average of $\langle c \rangle = 0.38$ against $\langle c \rangle = 0.61$ of nonkin authors ($D = 0.289$, $P < 0.0001$). The latter result was not only due to the larger average degree of kin authors—more coauthors imply a lower probability for each couple of them to collaborate—but it seemed a peculiar feature of the network around kin nodes as demonstrated by stratifying $\langle c \rangle$ by degree class (Fig. 3*B*) for low, intermediate, and high number of connections. The local structure around kin nodes resembled a star graph with the kin node at its center and almost no connections between her neighbors, whereas the neighborhood of nonkin nodes showed a more clique-like structure with several connections between the neighbors of the central node (Fig. 3*C*). In addition to the degree-stratified analysis, we carried out two subanalyses for authors active in the periods 1983–1995 and 1996–present, in accordance to the different authorship rules defined by PubMed. For both periods we found significant differences in the betweenness and clustering coefficient between kin and nonkin authors (*SI Appendix*, Tables S8 and S9). Finally, we studied the relationship between authors taking into account the mixing patterns of the connections, using the Pearson's correlation coefficient R of the degrees at either ends of an edge (21). A positive coefficient denotes that high-degree nodes tend to create connections among each other; that is, important authors tend to collaborate more with other important authors, whereas if R is negative, high-degree nodes try to avoid connections with nodes of a similar degree. To highlight differences in the relationships between kin and nonkin authors, we measured the correlation coefficient separately for the three possible types of connections: a kin author with a nonkin one, kin-kin authors' connections, and nonkin links. Although the values of R for the entire network (without considering the type of linkage) and for nonkin connections were similar ($R = 0.175$ and $R = 0.125$) and in line with other types of coauthorship networks (20), kin–kin interactions showed a stronger correlation ($R = 0.351$), whereas connections between kin authors and nonkin ones were characterized by a weakly negative correlation ($R = -0.0239$). Although important nonkin nodes tend to form connections among each other, kin authors have a different behavior depending on the type of interaction. In kin–kin connections, important authors have a strong tendency to connect

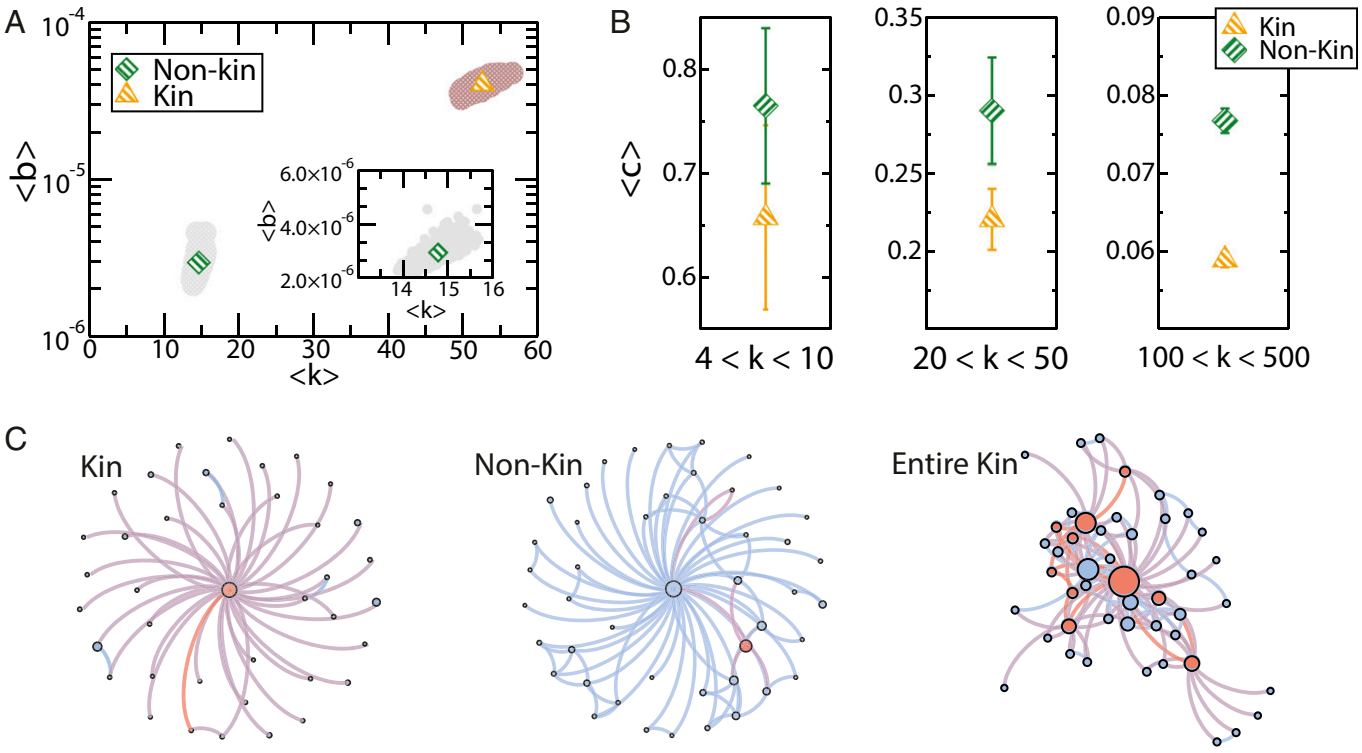

Fig. 3. Coauthorship network. (*A*) Scatterplot of degree (*k*) vs. betweenness centrality (*b*) of nodes (i.e., authors), showing average values for the nonkin (squares) and kin (triangles) nodes and their bootstrap estimates (gray/brown points, 10,000 replications). (*B*) Average clustering coefficient <c> of nonkin (diamonds) and kin (triangles) authors stratified by degree class: low degree nodes with *k* between 4 and 10 (two-sample Kolmogorov–Smirnoff's $D = 0.161$, $P < 0.0001$), intermediate nodes with *k* between 20 and 50 ($D = 0.190$, $P < 0.0001$), and hub nodes with *k* between 100 and 500 ($D = 0.210$, $P < 0.0001$). (*C*) Egocentric networks of two authors, one being part of a kin (orange) and the other one being not (blue) and for an entire kin composed by 10 authors. Nodes size is proportional to their degree. Type of connections are nonkin ↔ nonkin in blue, kin ↔ kin in orange, and mixed connection in violet.

together, whereas they tend to avoid other important authors who are not part of a kin (*SI Appendix*, Tables S6 and S11 and Figs. S14–S17).

Discussion

This study showed that kinship in the worldwide health science literature has increased modestly over the last 50 y, and this is probably an effect of the increased number of authors observed in scientific literature (22), but with substantial differences between nations. For example, Italy and Poland exhibited a dramatic increase in kinship starting from very low values and crossing the overall trend in the early 1980s and 1990s, which should be considered in the context of changing demographic and social factors. Exploring the concept of social capital, we observed low publication kinship among countries with low perceived corruption, and milder association of kinship with income inequality.

When investigating the coauthorship networks of authors from the top publishing countries, we found that authors who are part of a kin tend to have a larger number of collaborators and to occupy central positions in their networks. We could interpret this as increased information flows and allied activities such as grant applications, emanating from influential individuals who are more commonly kin coauthors. This interpretation is supported by the clustering analysis showing that the local structure of collaborations of a kin coauthor is usually very centralized, whereas authors who are not part of a kin tend to create democratic structures with groups that are more cohesive. In addition, the analysis of the mixing patterns in the network strongly supports the idea that important kin authors form robust collaborations among their peers and do not collaborate with those productive scientists who are not part of a kin.

Altogether the results from the network analysis suggest a distinctive portrait of the scientific community around kin networks: each kin group has several satellite members (including people sharing the same surname), and there is a centralized organization with a single scientist at the core that usually collaborates with other kins.

Our work has limitations. First, the methods for filtering surnames based on their potential demographic bias may need refinement and may not work in the same ways for all countries, given the fact that each country has its own peculiar surname distributions and flows. The differences between the Monte Carlo simulation and Wikipedia filter may be due to the fact that the Monte Carlo method also accounts for dynamic changes of surnames' distributions through time, whereas the Wikipedia filter is time-invariant. For instance, in the United States, the Chinese population has increased from 0.13% in 1960 to 1.23% in 2010. Second, our definition of kinship is blind to changes in a person's surname, for example, at marriage or from parents to children. National estimates may be biased by local practices in changing surname, and time trends could be misinterpreted if these practices varied over time. In Italy, since 1975, women were no longer legally required to adopt their husband's surname. Third, stratification by country was based on the first/senior author's affiliation, whose extraction may not be always accurate and by the country of publication. Although we observed very similar trends with the two classification methods, we implicitly assumed that the research activity in one article was led either in the country of the first author or where the paper was eventually published, and both are arbitrary assumptions. Shen and Barabasi (23) have introduced a method for identifying authors' credit in a publication that does not necessarily correspond to their position in the authors' list, e.g., first or last.

Their approach could be used in this work, but it would require imputing the country for all nonsenior authors whose affiliation was not recorded in the PubMed record and would be possible only if the author had at least one paper with known affiliation in a certain period. Fourth, the increasing kinship trends observed overall and in specific countries over time are correlated to the increased number of authors in more recent calendar years, which has been previously reported and discussed (22); also, PubMed operated changes in authorship policies (e.g., maximum number of authors allowed per paper, or use of collective names) for different time periods. In our work, the number of authors was a parameter of the Monte Carlo simulation, and we verified that the increasing trend was observed both in kinship and nonkinship papers. By selecting papers with a constant number of authors (four/five), we observed that the country-specific trends are maintained, yet the worldwide trend of kinship prevalence stabilized around 3% across calendar years. However, we could not exclude that country-specific inflation of the number of authors over time may be a driver of the phenomenon and that the maximum number of authors imposed in some time periods could affect the trends. Fifth, the corruption perceptions index is only one metric of one aspect of civic capital. The index is available only for the last 20 y, and we loosely normalized all values into a 10-point scale, without taking into account the modifications of the scoring that have been made over the years. For the income inequality, the analysis may be even less reliable as the data were sparser. There is a plethora of measures of income inequality, of perceived corruption, and of human development, which can be linked to social capital, but a careful evaluation of potentially spurious correlations is warranted. Sixth, in the network analysis we used the simple coauthorship as a proxy for collaboration between scientists, not considering individual contributions in each paper nor other dynamics, such changes of affiliations. We did not weight! the intensity of links, for instance, using the number of coauthored papers.

In conclusion, a certain level of kinship may have beneficial effects on the research outputs of a country, whereas greater or lesser amounts of kinship could have adverse effects. It is perhaps more important for nations to promote equal opportunities in academic careers than to attempt to contain nepotism. Just as kin build potent academic networks with their own resources, societies may do well to provide equivalent support for talented individuals with fewer resources, on the periphery of networks.

ACKNOWLEDGMENTS. We acknowledge the University of Manchester's Health eResearch Centre, funded by the Medical Research Council Grant MR/K006665/1. S.M. is supported by the Ministry of Economy and Competitiveness (MINECO), Spain, through the Juan de la Cierva Program. V.I.T. is supported by NIH Grant P01AG039347. M.P. is supported by European Union's Project EC H2020-PHC-2014 634650.

1. Bellow A (2003) *In Praise of Nepotism: A Natural History* (Doubleday, New York), 1st Ed.
2. Ciulla JB (2005) In praise of nepotism? *Bus Ethics Q* 15(1):153–160.
3. Morano Foadi S (2006) Key issues and causes of the Italian brain drain. *Innovation: Eur J Social Sci Res* 19(2):209–223.
4. Editorial (2008) Situations vacant. *Nature* 456(7219):142.
5. Martin HA (2008) Persistent nepotism in peer-review. *Scientometrics* 74(2):175–189.
6. Wenneras C, Wold A (1997) Nepotism and sexism in peer-review. *Nature* 387(6631): 341–343.
7. Gintis H, Osborne Groves M, Bowles S (2005) *Unequal Chances: Family Background and Economic Success* (Russel Sage Foundation, Princeton Univ Press, New York).
8. Bowles S, Gintis H (2002) Social capital and community governance. *Econ J* 112(483): F419–F436.
9. Chervenak FA, McCullough LB (2007) Is ethically justified nepotism in hiring and admissions in academic health centers an oxymoron? *Physician Exec* 33(5):42–45.
10. Pinchot S, Lewis BJ, Weber SM, Rikkers LF, Chen H (2008) Are surgical progeny more likely to pursue a surgical career? *J Surg Res* 147(2):253–259.
11. Allesina S (2011) Measuring nepotism through shared last names: The case of Italian Academia. *PLoS One* 6(8):e21160.
12. Ferlazzo F, Sdoia S (2012) Measuring nepotism through shared last names: Are we really moving from opinions to facts? *PLoS One* 7(8):e43574.
13. Durante R, Labartino G, Perotti R; National Bureau of Economic Research (2011) Academic dynasties: Decentralization and familism in the Italian academia NBER Working Paper Series Working Paper 17572 (National Bureau of Economic Research, Cambridge, MA). Available at www.nber.org/papers/w17572.
14. Torvik VI (2015) MapAffil: A bibliographic tool for mapping author affiliation strings to cities and their geocodes worldwide. *Dlib Mag* 21(11-12):34–44.
15. Torvik VI, Smalheiser NR (2009) Author name disambiguation in MEDLINE. *ACM Trans Knowl Discov Data* 3(3):11.
16. Dorogovtsev SN, Goltsev AV, Mendes JF (2006) k-Core organization of complex networks. *Phys Rev Lett* 96(4):040601.
17. Newman M (2010) *Networks: An Introduction* (Oxford Univ Press, Oxford, UK).
18. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: Structure and dynamics. *Phys Rep* 424(4–5):175–308.
19. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442.
20. Newman ME, Park J (2003) Why social networks are different from other types of networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 68(3 Pt 2):036122.
21. Newman ME (2002) Assortative mixing in networks. *Phys Rev Lett* 89(20):208701.
22. Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039.
23. Shen HW, Barabási AL (2014) Collective credit allocation in science. *Proc Natl Acad Sci USA* 111(34):12325–12330.