Informal Connections Outweigh Co-authorship Ties in Academic Impact

Lluís Danús[a], William Dinneen[b], Carolina Torreblanca[b], Guy Grossman[b], and Sandra González-Bailón[a,1]

This manuscript was compiled on August 11, 2025

Research has documented the importance of teamwork in the form of co-authorship for research productivity and innovation, but we know much less about how informal collaborations relate to academic success. Informal ties allow intangible exchanges like mentoring, guidance, and feedback to flow among scholars: these interactions weave a support structure that improves ideas and encourages project growth. However, these informal exchanges are more difficult to measure because they do not leave as clear a trail as co-authorship ties. Here, we uncover this layer of informal communication around scholarly outputs by parsing the information contained in the acknowledgment sections of published articles. Our data include $N \sim 130,000$ articles authored by $N \sim 86,000$ scholars from the period 2003-2023. We analyze scholars' embeddedness in this informal structure of collaboration and reveal that (1) informal ties create a larger and denser network of support than co-authorship ties; (2) disconnection from informal networks is associated with gaps in productivity and impact; and (3) informal ties are a more relevant predictor of academic success than formal collaborations, even after matching for gender, seniority, methodology, and geographical location. Using coarsened exact matching and random forest regressions we show that informal structures of support are significantly associated with academic impact, creating gaps in who benefits from those connections.

Science of Science | Invisible College | Social Networks | Computational Social Science

Significance Statement

The term 'invisible college' has been used for decades to describe the informal networks of communication that help scientists advance knowledge. Measuring these informal structures and who is more likely to benefit (or be excluded) from the exchanges they enable is an empirical challenge, given that those exchanges are usually intangible. Here we analyze the 'thank you' notes published in journal articles to approximate these informal ties. We show that scholars disconnected from this layer of the invisible college do worse in terms of publication impact. We also show that informal ties offer a type of support not captured by co-authorship ties, a more rigid type of collaboration. Documenting how informal structures of support operate can help leverage that collective resource in the pursuit of shared intellectual goals.

Science is an inherently collaborative endeavor. Researchers exchange ideas, provide feedback, and collectively advance knowledge through the activation of formal and informal networks. Yet the exact mechanisms through which collaboration shapes scholarly success remain only partially understood. We are bound by what we can measure, which is why past research has focused predominantly on co-authorship structures. Indeed, co-authorship offers one of the main collaboration mechanisms to integrate knowledge and expertise: co-authorship networks influence professional trajectories, shape academic impact, and facilitate breakthrough discoveries (1–5). These networks have provided the main measures to approximate the 'invisible college', a term coined to describe the channels of knowledge exchange that transcend institutional affiliations (6, 7). However, joint publications are not the only mechanism to engage in intellectual collaboration. Researchers often rely on other, more informal types of information sharing that do not require institutionalized structures of communication, like publications through academic journals. Collaboration networks also arise through more intangible exchanges that feed relevant information (e.g., guidance and feedback) into the exercise of scientific research (8). These intangible exchanges create another layer of communication through which the invisible college operates. Here, we aim to grant some visibility to this intangible structure by retrieving the 'thank you' notes from the acknowledgment sections of published articles. We aim to test if embeddedness in these informal structures of communication helps explain variation in academic success.

Researchers have always relied on personal connections and shared intellectual interests to develop their work and advance the state of the art in their fields. Scientific ideas evolve and gain traction through those connections, which are activated through informal and often private communication. Darwin and Einstein, for instance, wrote thousands of letters during their lifetime, on occasion writing as many as 12 letters a day (9). That these two prominent scientists devoted substantial portions of their time to engage in correspondence reveals the importance they attributed to this type of intellectual exchange. And so did their correspondents, who sometimes received encouragement, other times a critical response to their research efforts. Today, the range of communication tools has greatly expanded but many scholars still use them with the same goal: to seek or offer feedback to

Author affiliations: [a]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, 19104; [b]Political Science Department, University of Pennsylvania, Philadelphia, PA, 19104

L.D. and S.G-B. conceptualized the research, analyzed the data, and produced the figures. W.D., C.T., and G.G. provided materials and methods. All authors contributed to the writing.

The authors declare no competing interest.

[1]To whom correspondence should be addressed. E-mail: sgonzalezbailon@asc.upenn.edu

projects and ideas outside the institutionalized mechanisms of peer-review. In other words, scholars, including the most prominent ones (10), activate professional connections to improve their research as it brews and evolves towards publication.

These exchanges rely, for the most part, on private (and, these days, increasingly encrypted) communication. Absent a historical archive of surviving letters, and given the difficulty of accessing private digital records, how are we to measure connections in this important layer of the invisible college? We offer an answer to this question that relies on the convention to acknowledge peers and colleagues once an article is published. Past research suggests that the acknowledgment sections encode information that can help explain academic performance (10–14). The convention to acknowledge peers is more closely followed in some fields than others. In Political Science, acknowledging colleagues who provided feedback is a common feature in published articles, perhaps because of the long time it takes to complete the submission-review-publication cycle. This field has a double blind review process, and acknowledgment sections are not permitted at the review phase, so strategic 'thank you' notes are less likely than in disciplines with a single-blind review process. Informal ties capture structures of support, but whether they help explain academic performance or how they compare to more formal collaborations are open questions. Acknowledgment networks operate in parallel to those formed by co-authoring relationships — but which one matters more to explain variation in the impact of scholars' work?

Formal and Informal Collaborations

Collaborations that result in co-authorship usually require a minimum level of investment in a joint project. Who qualifies as a co-author is often a subjective criterion and, occasionally, a contested decision (15). But there is general agreement that co-authors make substantive contributions to the work, either in its conceptualization or its execution. This type of formal collaboration aims to bring together skills and expertise that strengthen the quality of the output — or so is the hope. The goal of formal collaborative work is, ultimately, to be published in a well-regarded journal and to have an impact in future work.

However, formal collaborations are not the only mechanism to integrate ideas and expertise. It is also common to circulate and discuss papers ahead of (or in parallel to) peer-review, mostly through scholars' professional networks. This type of research communication often results in feedback that ends up being incorporated in the work to increase its value and its chances of successful publication. Informal ties also help disseminate ideas prior to the research being published, which may contribute to more favorable reviews (if perceptions of the value of the ideas broaden with their dissemination). Journals are still the main mode of communication for academic research, but given the time it takes for the publication cycle to complete, proactive scholars activate communication networks on their own: they send their working papers to colleagues and present them in seminars and conferences, activities that create opportunities for informal exchange. Informal ties reflect how embedded scholars are in these academic networks of support that grow and evolve in parallel to more formal types of collaboration.

The 'thank you' notes appended in the acknowledgment section of published articles offer a window to these informal ties (11, 14). The acknowledgment section allows authors to recognize colleagues who provided support or feedback. These contributions are not substantial enough to warrant authorship but they are deemed important enough to merit appreciation. These acknowledged scholars are what past work called 'helpful scientists', a category that shifts attention from productivity measures (the usual benchmark of academic success) to helpfulness to others (16). The mechanisms explaining the creation of these helpful, informal ties are likely varied, as varied as the motivations attracting researchers to academic pursuits. But, whatever their nature, those mechanisms result in structures that place scholars in either networked or disconnected positions. Our main motivating question is: do these connections matter? Or, to put it differently, does membership to this layer of the invisible college increase the publication impact of scholars' work?

To address this question, we analyzed 129,750 articles published in 174 Political Science journals from 2003 to 2023 (see Materials and Methods). Using information contained in these articles, we built two longitudinal networks: one mapping the structure of co-authorships and the second mapping the structure of acknowledgments (Fig. 1A). For each author we collected additional information on their gender, seniority, geographical region, and methodological orientation (see Materials and Methods). We also compiled three measures of their academic impact: number of articles published (within the journals we track and in total); h-index (calculated with in-sample articles); and Euclid score (also calculated with in-sample articles; in the SI we offer additional analyses using the h-index and Euclid scores derived from all publications).

Results

Informal networks are larger, denser, and less hierarchical than formal networks. The networks formed by formal and informal ties exhibit substantial differences, as summarized in table 1. The acknowledgment network is larger and less fragmented, as assessed by the number of components (in the SI we show that the observed level of fragmentation is substantially higher than expected by chance). The acknowledgment network has twice as many connections, on average, per scholar (since the acknowledgment graph is directed, we add up all arcs to report these statistics). There is a similar fraction of disconnected scholars (isolates) in both structures: 28% and 30%, respectively. About 19% of all scholars in our data are disconnected from both networks, i.e., they have no formal or informal connections with colleagues.

The plots in Fig. 1B-C expand on these descriptive statistics. The acknowledgment network is less hierarchical, in terms of degree distribution, which means there are less extreme differences in the connectivity of scholars. These scholars are predominantly male in both networks (about 62% of all scholars), a disparity that is also present among the disconnected subset (insets). Both networks have a similar growth rate, but the acknowledgment network shows a tendency to retain more disconnected scholars over time. Fig. 1D shows that most of the articles published by scholars who are disconnected in the co-authorship structure are qualitative, consistent with the different epistemological and

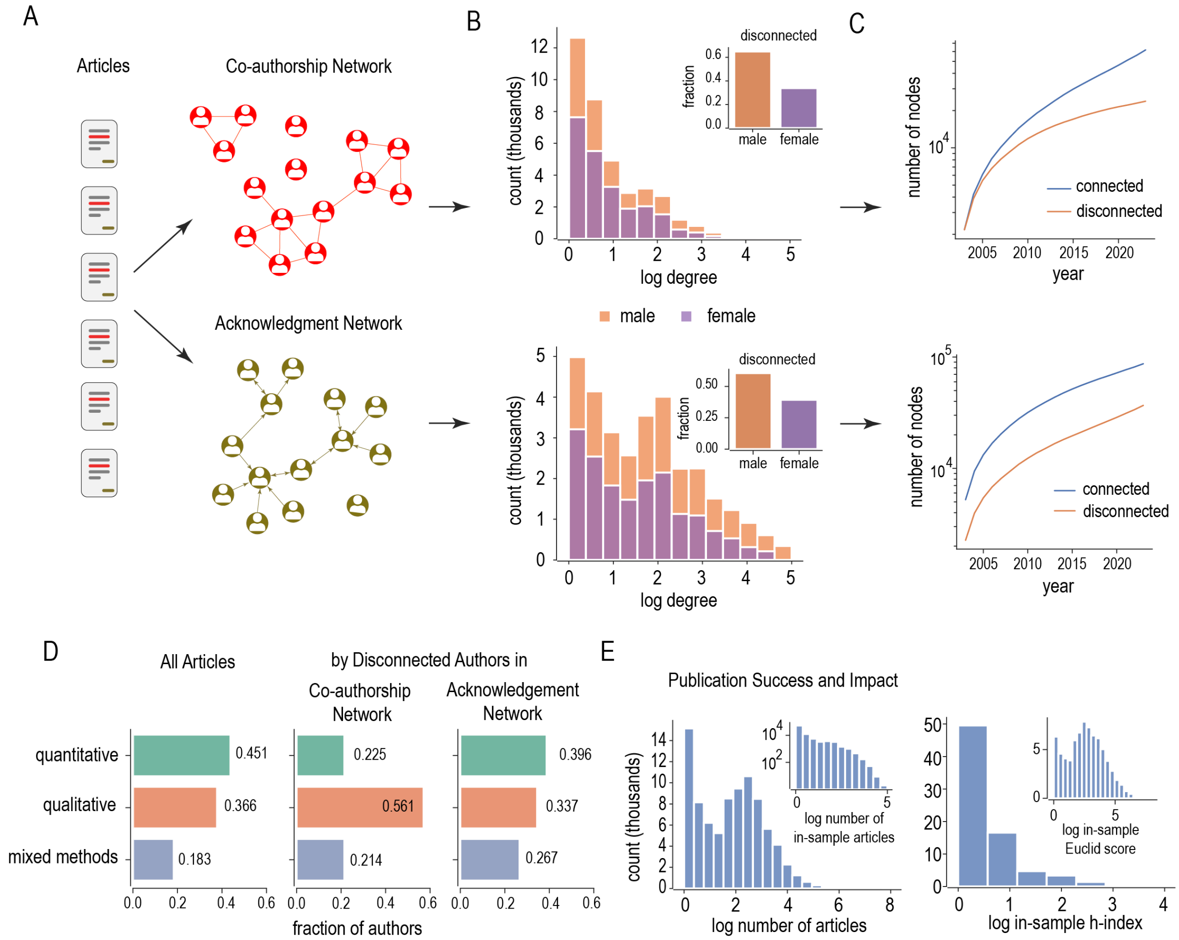

Fig. 1. Description of the Data. (A) We extract co-authorship and acknowledgment ties from published articles and build the corresponding networks. (B) The acknowledgment network is denser and less skewed. About 28% and 30% of all scholars are disconnected in each network, respectively. (C) Both networks grow at similar rates, but the group of disconnected scholars grows more in the acknowledgment network (D) There is a slight predominance of quantitative articles in our sample. Most of the disconnected scholars in the co-authorship network publish qualitative articles, but disconnected scholars in the acknowledgment network are more evenly distributed across methods. (E) Distributions of our three measures of academic impact: number of articles published (total and in-sample); h-index; and Euclid score (inset), all measured in 2023.

technical requirements in that methodological subdomain; disconnected scholars in the acknowledgment network have a more even distribution of methodological approaches.

In panel 1E we show the distribution of our two measures of academic impact: number of publications (total and in-sample), and citation impact, as measured by the in-sample h-index and Euclid scores. The former leverages the number of citations relative to the total volume of scientific output, while the latter emphasizes highly cited work (see Materials and Methods for formal definitions). We want to determine if being disconnected from support structures is associated with lower academic impact, as assessed by these conventional metrics.

Disconnection is associated with gaps in productivity. In Fig. 2 we examine variation in publication success for each decile of the within-field productivity distribution. Scholars with multi-disciplinary backgrounds also publish in non-political science journals, thus having many of their articles excluded from our data. We take this into account by looking at the ratio of articles published in the journals we analyze over all articles authored. We then bin the scholars using the deciles of this distribution, such that those in the upper deciles are publishing most of their papers within the field. Contingent on being published in one of the political science journals we analyze, and contingent on being in the same decile of within-field productivity, disconnected authors have consistently a lower impact, both as measured by the h-index and the Euclid scores. This difference is especially noticeable in the acknowledgment network: in the first decile, for instance, the mean h-index is more than twice as high in the connected group; it is seven times higher for the Euclid score. These gaps are noticeable for most of the within-field productivity

Table 1. Characteristics of the Two Networks

Statistics	Co-Authorship	Acknowledgment
Number of nodes	85,653	124,023
Degree range	[0, 139]	[0, 499]
Average degree	2.85	5.81
Median degree	1.0	1.0
Number of components ($size \geq 2$)	8,871	1,630
Number of isolates	23,807	36,841

distribution. (In the SI we include additional analyses that consistently show the presence of these gaps across a range of comparisons: using a longer, 5-year lag between the network measurement and the two metrics of impact; using the h-index and Euclid scores calculated with all publications, not just those within our sample; and using a subset of the data that only includes scholars connected in the co-authorship network).

Informal ties are the most relevant predictor of publication impact. In Fig. 3A-B we show the estimates of two generalized linear models explaining variation in the h-index and Euclid scores using matched data with balanced groups (see Materials and Methods). Being connected to the acknowledgment network (as opposed to being an isolate) is the most important predictor of impact (we use, as in Fig. 2, a two-year lag). The relative relevance of the other variables varies depending on the measure of impact used. Seniority, for instance, is more important than gender to explain variation in the Euclid score, a measure that is more sensitive to career length. Likewise, the methodological approach of scholars has also different effects depending on the measure of impact used: quantitative research fares better with Euclid scores. Consistent across models, being based in North America or Europe (as opposed to elsewhere in the world) is positively associated with publication impact, but the magnitude of that association is not as large as that of having co-authors and, especially, having informal ties: connections to helpful colleagues are the most relevant correlate regardless of the metric of impact used.

Panels 3C-D display the rank order of variable importance according to random forest models fitted to the full data (not just matched observations). Having connections in the acknowledgment network is, again, the most important correlate of publication impact. Both the percentage increase in root mean square error (RMSE, our primary measure of feature importance) and the increase in node purity (our secondary measure) suggest that being connected in the acknowledgment network is the most important covariate to explain variance in publication impact. The second most important correlate is being connected to co-authors, consistent with the linear models and also with prior research (2, 4). In the SI we provide the full regression tables and alternative model specifications (with consistent results). As is common in all multivariate analysis, we can compare the importance of network connectedness only to the covariates included in the model specification. While seniority, methodological expertise, and institutional affiliation are the most emphasized correlates in the literature (e.g., (11, 17)), there could be other unobserved factors that are also predictive of academic impact. However, to the degree that these factors shape career trajectories, they are likely implicitly shaping authors' co-authorship networks and are thus tacitly accounted for.

Discussion

Our analyses confirm that embeddedness in informal networks of academic exchange is associated with higher publication impact. Informal collaborations allow scholars to draw advice and feedback from a wider range of colleagues than those formally attached through co-authorship ties. These informal relationships create the circuits for information exchange that are core to the operation of the 'invisible college'. Our analyses show that these connections are key predictors of publication impact – but also that there is an uneven distribution of opportunities in how research communities self-organize. Informal ties are a source of support but also a gateway to opportunities that not every scholar can (or decides to) cross.

Our evidence is consistent with the idea that informal structures of support enable differential access to knowledge and advice. Of course, embeddedness in these structures is not a binary category: those connected can hold very different positions within the network, some more advantageous than others. Here we focused on the differences between disconnected and connected scholars (especially since more than a quarter of researchers in our data have no connections in these structures). In future research we plan on unpacking the different positions that those embedded in the networks hold, and whether variation in those positions also helps explain variation in academic impact. Here we demonstrate that network connectivity status is strongly associated with performance differences.

Our acknowledgment data has a broader coverage than past research in terms of number of journals and the length of the observation window. There is only a handful of studies using the acknowledgment sections to explain publication outcomes, and a key limitation of this past work (e.g., (10, 13)) is that they only measure the number of acknowledged 'commenters', without constructing the overall network or comparing it to more formal forms of collaboration. The few studies that consider acknowledgment networks either cover a single subfield (e.g., Finance Economics, based on six purposively selected journals (11)) or focus on topic-specific publications over short time periods (e.g., wind energy publications over a 2-year window, (12). In addition, this past research centers on trying to explain paper-level measures of impact, like the citation counts accumulated by specific articles. Our focus on the overall impact of scholars allow us to analyze (and theorize about) their network embeddedness as a structural position that enables (or not) opportunities for information exchange. Our approach, in other words, allows us to more directly consider the role of the invisible college in shaping scholarly success.

Like all empirical research, our approach also has limitations. The acknowledgment network we analyze is a proxy to an underlying, unmeasured structure that is likely only partially represented in our measurements. Most scholars have the sort of informal exchange we aim to capture with the 'thank you' notes, but often those exchanges do not make it to the acknowledgment section. It is almost certain that our network of informal collaboration is a partial and noisy representation of the unobserved structure. In addition, being disconnected from the acknowledgment network is likely correlated with unobserved individual characteristics that could (also) affect publication impact. Our regression estimates should be viewed as informative and robust associations but not necessarily as direct causal pathways. Absent a randomized experiment, using observational estimates is the closest we can get to estimating the importance of embeddedness in informal structures of support. But determining causality from these observational data presents significant challenges. Like all empirical research, our approach also has limitations.

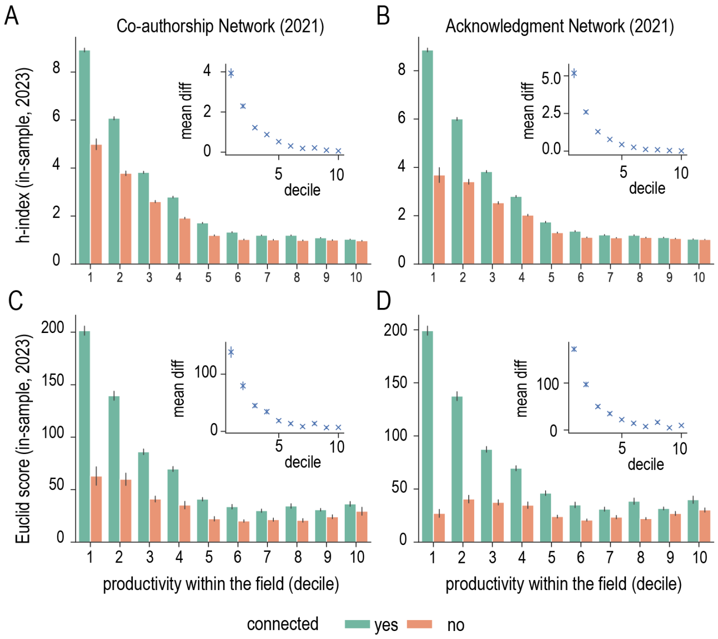

Fig. 2. Differences in Publication Impact for Each Productivity Decile. Disconnected authors in both networks have, on average, lower impact across all deciles of the within-field productivity distribution. The mean difference gap is higher for disconnected authors in the acknowledgment network (insets). Whiskers indicate 95% confidence intervals.

The acknowledgment network we analyze is a proxy to an underlying, unmeasured structure that is likely only partially represented in our measurements. Most scholars have the sort of informal exchange we aim to capture with the ‘thank you’ notes, but often those exchanges do not make it to the acknowledgment section. It is almost certain that our network of informal collaboration is a partial and noisy representation of the unobserved structure. In addition, being disconnected from the acknowledgment network is likely correlated with unobserved individual characteristics that could (also) affect publication impact. Our regression estimates should be viewed as informative and robust associations but not necessarily as direct causal pathways. Absent a randomized experiment, using observational estimates is the closest we can get to estimating the importance of embeddedness in informal structures of support. But determining causality from these observational data presents significant challenges.

Future research should also consider if the gaps we identify here also arise in other research fields. Different disciplines follow different conventions in acknowledging peers in published work. Whether our findings generalize to other fields is an empirical question that requires being able to measure informal ties of exchange – and this is more difficult if there is no general practice to acknowledge colleagues. In the field we consider (Political Science), informal ties leave a visible footprint because acknowledgment sections are commonly used. It is possible that the impact of helpful colleagues varies across fields of inquiry. Future work should consider other measurement strategies to more accurately capture these informal structures of academic support across disciplines.

Overall, our evidence suggests it is important to document how informal structures of support operate so that more scholars can leverage this type of collective resource. At the very least, understanding these structures and how they shape knowledge generation can help give credit where credit is due, beyond the usual metrics of academic performance focused on publication counts and the accumulation of citations.

Materials and Methods

Data. We obtained the full list of Political Science Journals from the Clarivate's Social Science Index. We excluded non-English journals, journals with an impact factor lower than 1, and journals that do not use peer review. In the SI we list the full list of journals included in our sample ($N = 174$). We then used the citation database Elsevier Scopus to (1) obtain the index of all articles published in each journal (which we downloaded as pdfs) and (2) collect additional metadata for the authors of these articles, including yearly number of publications, annual citation count, and country of institutional affiliation (see the SI for a map of the global distribution of scholars in our sample). For each author we also obtained basic metadata on articles published in journals other than those in our list (i.e., out-of-sample publications). In total, the articles are authored by 85,653 unique scholars. In order to have enough data to calculate the in-sample measures of publication impact, we filtered out authors with fewer than 10 publications over the period we analyze (see SI for more details). A significant fraction of the articles (57.56%) are single-authored and more than half (55.61%) have an acknowledgment section.

Measures of Publication Impact. We calculated the h-index and Euclid scores using only in-sample publications (see SI for a description of the same metrics using all articles). The h-index (or hirsch-index (18)) is one of the most commonly used metrics of impact in the science of science (19). It is defined as:

$$h = \max \{ i \in \{1, 2, \dots, N\} \mid c_i \geq i \} \qquad [1]$$

where c_i represents the number of citations of paper i. It is thus defined as the highest number h, which means that a scholar has h papers each cited at least h times. The h-index measures both productivity (number of papers) and impact (citations).

The Euclid score (20) is defined as:

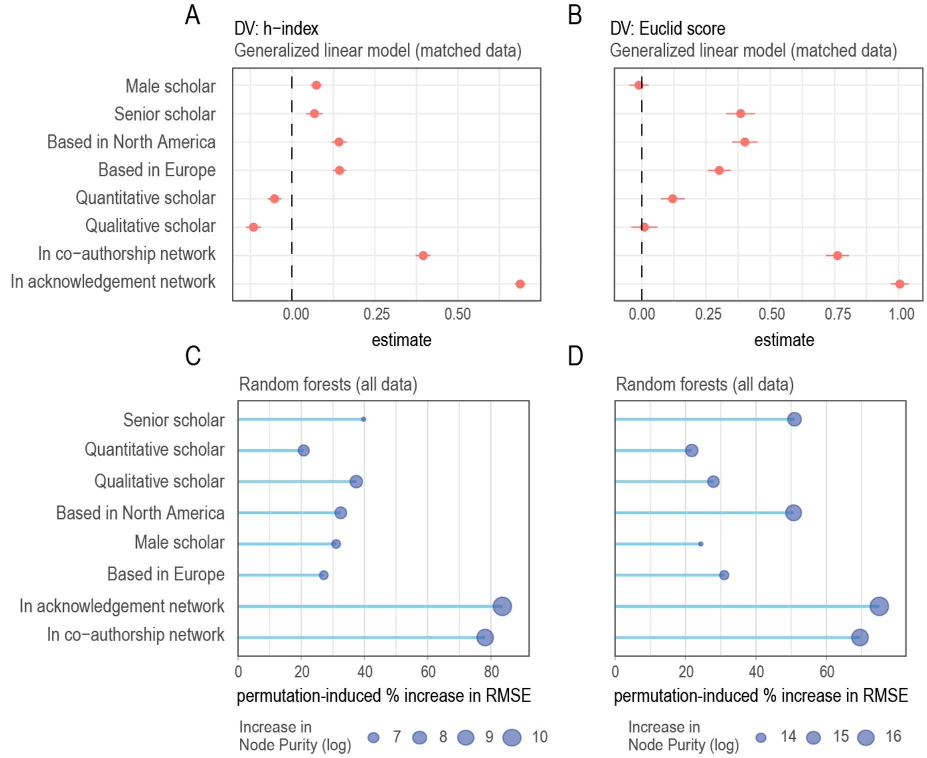

Fig. 3. Correlates of Publication Success. (A-B) Estimates of a generalized linear model using a matched dataset. (C-D) Variable importance according to random forests using all data.

$$e = \sqrt{\sum_{k=0}^{N} C_k^2} \qquad [2]$$

where c_k represents the number of citations of paper k. The Euclid score provides an alternative measure of impact that emphasizes the distribution of citations rather than just their total count. A higher score indicates that a scholar has a few highly cited works, whereas lower scores reflect broader contributions with a more moderate number of citations across.

Gender Classification. We predicted the gender of authors using *genderize.io*, a name-to-gender classification tool that predicts binary gender based on the frequency of first names (and country when available) in a labeled dataset of over one billion public social media profiles. We were able to assign gender for about 97% of authors based on their name and country of institutional affiliation when available.

Authors' Seniority. We define seniority using the time of first publication: senior authors are those who published their first article more than 10 years prior to 2023. Junior authors are those who published their first article after 2013.

Methodological Orientation. We classify the main methodological approach of the articles using a combination of supervised machine learning and AI-assisted fine-grained classification. First, we cleaned and prepared the raw text of articles by removing preambles and bibliography, white spaces, lines with fewer than 5 words or less than 50% text characters, and stop-words. We then used the manual coding of $N = 1,694$ articles, taken from (17), to train a two-step classifier for three categories: quantitative, formal theory, qualitative/normative. In the first stage, we vectorize the text and train a TF-IDF + Logistic Regression model to identify key features (unigrams and bigrams) associated to each of the three methods categories. In stage two, we used the top 70 most significant features and retrained the model. We evaluated model performance with a 10 fold cross validation (we achieve a mean macro F1 score of 82%, +/- 0.06 across, see the SI for precision and recall metrics as well as the list of top 40 features used during training). This yielded the classification of the papers in our sample into two categories: 'quantitative/formal theory' and 'qualitative/normative'. Once the papers were classified, we identified the predominant category at the scholar level: if an author has $\frac{2}{3}$ of their articles labeled as 'quantitative/formal theory' or 'qualitative/normative', those are the assigned labels (respectively); if they do not reach that cutoff, the category we assign them is 'mixed methods' (which means that scholars have a portfolio of publications that varies in the methodological approach).

Coarsened Exact Matching Regressions. Using a randomized experiment to manipulate the network embededdness of scholars and determine its impact on publication success is not a feasible design. Instead, we try to alleviate concerns related to confounders by reweighting observations to ensure that connected and disconnected scholars are as similar as possible along all observed covariates. If connected and disconnected scholars are different only across observable characteristics, this strategy is sufficient to retrieve the difference attributable to connectedness alone by comparing the two groups of interest: disconnected and connected scholars. We reduce imbalances in covariates identified as relevant in past research: gender and seniority (21), geographical location (22), and methodological orientation (23). To reduce imbalances we use a statistical method known as coarsened exact matching (24, 25). This method groups observations into strata based on coarsened values of covariates to make the 'treated' (connected scholars) and 'control' (disconnected scholars) units comparable. This matching approach results in a dataset that excludes unmatched authors and balances the two groups so that the covariates follow similar empirical distributions. In the analyses presented here we use the k-to-k restriction to ensure that within each matched stratum there are the same number k of 'treated' and 'control' observations. In the SI we also produce results without this restriction (and thus matched strata of different sizes). We applied generalized linear regression models to the matched dataset to estimate the effects of being connected in both networks on publication impact. See the SI for full regression tables and other model specifications.

Random Forest Regression. We use a machine learning algorithm known as random forests (26, 27) to assess the importance of informal connections using the full dataset (not just matched observations). Random forests do not assume linear relationships and are better at capturing complex interactions without these being explicitly specified in the model. Overfitting is also less of an issue than in other regression models because random forests operate with ensemble averaging. Random forests are also less sensitive to outliers. We use the percentage increase in root mean squared error (RMSE) as the main measure of variable importance. This measure quantifies how much worse the model performs when a specific variable is randomly shuffled (thus breaking its relationship with the dependent variable, in our case, publication impact). This metric is considered more reliable than other measures like increase in node purity, but we also report this secondary measure to offer more context. Node purity captures the total reduction in mean squared error when a variable is used for splitting across all trees in the forest. In the SI we provide additional robustness tests using a classifier version of the random forest model, and two additional variations using XGBoost and Support Vector classifiers.

Replication Materials. We obtained our data from Elsevier Scopus, which has use policies that limit data sharing. We have deposited the code that allows redownloading the data (including the author and article metadata) in a GitHub repository, which includes a description about how to access the database. The repository also includes the code necessary to replicate our analyses.

ACKNOWLEDGMENTS. We wish to extend a special thank you to Irene Tang and Shuning Ge for their help setting the data pipeline. We also thank Noa Rubinstein, Micaela Montero Johnson, Ria Rege, Clarice Wang, Erika Kishino, and Arihant Tripathi for excellent research assistance. G.G. thanks Penn's School of Arts and Sciences and S.G-B. thanks the Annenberg School for Communication for providing research funds necessary to conduct this study.

1. V Larivière, Y Gingras, CR Sugimoto, A Tsou, Team size matters: Collaboration and scientific impact since 1900. *J. Assoc. for Inf. Sci. Technol.* **66**, 1323–1332 (2015).
2. V Sekara, et al., The chaperone effect in scientific publishing. *Proc. Natl. Acad. Sci.* **115**, 12603–12607 (2018).
3. L Wu, D Wang, JA Evans, Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
4. E Sarigöl, R Pfitzner, I Scholtes, A Garas, F Schweitzer, Predicting scientific success based on coauthorship networks. *EPJ Data Sci.* **3**, 9 (2014).
5. O Vásárhelyi, I Zakhlebin, S Milojević, Eorvát, Gender inequities in the online dissemination of scholars' work. *Proc. Natl. Acad. Sci.* **118**, e2102945118 (2021).
6. DJ De Solla Price, D Beaver, Collaboration in an invisible college. *Am. Psychol.* **21**, 1011–1018 (1966).
7. D Crane, Social structure in a group of scientists: a test of the "invisible college" hypothesis. *Soc. Networks* pp. 161–178 (1977).
8. N Desrochers, A Paul-Hus, J Pecoskie, Five decades of gratitude: A meta-synthesis of acknowledgments research. *J. Assoc. for Inf. Sci. Technol.* **68**, 2821–2833 (2017).
9. JG Oliveira, AL Barabási, Darwin and einstein correspondence patterns. *Nature* **437**, 1251–1251 (2005).
10. W Lou, J He, L Zhang, Z Zhu, Y Zhu, Support behind the scenes: the relationship between acknowledgement, coauthor, and citation in nobel articles. *Scientometrics* **128**, 5767–5790 (2023).
11. ME Rose, CP Georg, What 5,000 acknowledgements tell us about informal collaboration in financial economics. *Res. Policy* **50**, 104236 (2021).
12. S Tian, X Xu, P Li, Acknowledgement network and citation count: the moderating role of collaboration network. *Scientometrics* **126**, 7837–7857 (2021).
13. Q Xie, X Zhang, Exploring the correlation between acknowledgees' contributions and their academic performance. *Scientometrics* **128**, 6003–6027 (2023).
14. M Doehne, C Herfeld, How academic opinion leaders shape scientific ideas: an acknowledgment analysis. *Scientometrics* **128**, 2507–2533 (2023).
15. M Bikard, F Murray, JS Gans, Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Manag. Sci.* **61**, 1473–1495 (2015).
16. A Oettl, Reconceptualizing stars: Scientist helpfulness and peer performance. *Manag. Sci.* **58**, 1122–1140 (2012).
17. DL Teele, K Thelen, Gender in the journals: Publication patterns in political science. *Polit. Sci. & Polit.* **50**, 433–447 (2017).
18. JE Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci.* **102**, 16569–16572 (2005).
19. D Wang, AL Barabási, *The Science of Science.* (Cambridge University Press, New York, NY), (2021).
20. M Perry, PJ Reny, How to count citations if you must. *Am. Econ. Rev.* **106**, 2722–41 (2016).
21. J Huang, AJ Gates, R Sinatra, AL Barabási, Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl. Acad. Sci.* **117**, 4609–4616 (2020).
22. L Danús, C Muntaner, A Krauss, M Sales-Pardo, R Guimerà, Differences in collaboration structures and impact among prominent researchers in europe and north america. *EPJ Data Sci.* **12**, 12 (2023).
23. V Larivière, Y Gingras, On the relationship between interdisciplinarity and scientific impact. *J. Am. Soc. for Inf. Sci. Technol.* **61**, 126–131 (2010).
24. SM Iacus, G King, G Porro, Causal inference without balance checking: Coarsened exact matching. *Polit. Analysis* **20**, 1–24 (2012).
25. S Iacus, G King, G Porro, cem: Software for coarsened exact matching. *J. Stat. Softw.* **30**, 1 – 27 (2009).
26. L Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
27. A Liaw, M Wiener, Classification and regression by randomforest. *R News* **2**, 18–22 (2002).