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 October 4, 2025

Past work has documented the importance of formal collaboration, particularly co-authorship, in increasing research productivity and innovation. However, we know much less about how informal collaboration relates to academic success. Informal ties facilitate the exchange of intangible resources like mentoring, guidance, and feedback. These interactions form a support structure that improves ideas and facilitates the successful development of research projects. However, these informal exchanges are difficult to measure because they do not leave as clear a trail as co-authorship ties. 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 $N \sim 130,000$ political science articles authored by $N \sim 86,000$ scholars from 2003 to 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 lower productivity and impact; and (3) informal ties are a more relevant predictor of academic success than formal collaborations (i.e., co-authorship), even after matching for gender, seniority, methodological orientation, geographical location, and institutional prestige. Using coarsened exact matching and random forest regressions, we demonstrate that informal support structures are significantly associated with academic impact, creating gaps in who benefits from these connections.

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

Significance Statement

The term 'invisible college' has been used for decades to describe the networks of communication that help scientists advance knowledge. These networks often rely on informal interactions, and they create a form of social capital that gives access to resources like information or support. Measuring informal structures of communication to determine who is more likely to benefit (or be excluded) from the exchanges these structures 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 academic social capital 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 by activating formal and informal networks. However, the 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 and citation patterns. Indeed, co-authorship offers one of the primary mechanisms for integrating knowledge and expertise. Co-authorship networks influence professional trajectories, shape academic productivity and impact, and facilitate novel discoveries (1–5). The analysis of citation patterns, in turn, has uncovered clusters of interconnected scholars that shape the flow of ideas in scientific communities (6–8). Co-authorship and citation networks have so far provided the primary measures to approximate what past work refers to as the 'invisible college', a term that describes the channels of knowledge exchange that transcend institutional affiliations (9, 10). Invisible colleges embody a form of social capital, a broader concept that describes the resources individuals derive from social networks (11–13). If social capital focuses broadly on networks of relationships, the invisible college focuses on a narrower set of ties: those connecting scientists.

Joint publications and citations help us trace formal and institutionalized networks through which scholars circulate and test new ideas. However, co-authoring and citing other research are not the only channels of scholarly communication. Scientists also rely on more informal modes of information sharing, beyond institutionalized structures. Collaboration networks often emerge through intangible exchanges that also feed relevant guidance and recommendations into the practice of research (14). These exchanges create an additional layer of communication through which social capital operates, benefiting those with access to the resources it confers. To illuminate this intangible structure, we retrieve the 'thank you' notes from the acknowledgment sections of published articles. We aim to test if connections in these informal networks of communication help explain variation in academic success.

Researchers have long relied on personal connections and shared intellectual interests to advance their work. Scientific ideas evolve and gain traction through these connections, often activated in informal and private exchanges. Darwin

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

and Einstein, for instance, wrote thousands of letters — on occasion as many as a dozen in a single day— which underscores the value they and their correspondents placed on such intellectual exchange (15). Today, the range of communication tools has expanded, but scholars continue to use them with the same purpose: seeking and offering feedback outside the institutionalized mechanisms of peer review. Even the most prominent researchers rely on professional connections to shape their work as it develops toward publication (16).

These exchanges rely, for the most part, on private (and, 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 help and feedback received from peers and colleagues once an article is published. Past research suggests that the acknowledgment sections encode information that can help explain publication success (16–20). The convention of acknowledging 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 due to the lengthy time it takes to complete the submission, review, and publication cycle. This field has a double-blind review process, and acknowledgment sections are not permitted during 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 is an open question. 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 (21). However, there is general agreement that co-authors make substantive contributions to the work, either in its conceptualization or its execution (e.g., data analysis, writing). This type of formal collaboration brings together skills and expertise that strengthen the quality of the output — or so is the hope. The goal of formal collaborative work is to enhance productivity, increase the placement of research outputs in prestigious outlets, and have a lasting impact on future work.

Formal collaborations, however, 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, primarily through scholars' professional networks. This type of communication often results in feedback that is incorporated into the research flow, enhancing its value and increasing 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 remain the primary mode of communication for academic research; yet, given the time it takes for the publication cycle to complete, scholars proactively activate communication networks on their own. They send their working papers to colleagues and present their projects in seminars and conferences. These activities create opportunities for informal exchange. Informal ties reflect the extent to which scholars are embedded in these academic networks of support, which grow and evolve in parallel with more formal forms of collaboration

Why Scholars Use Acknowledgments. The 'thank you' notes that are part of the acknowledgment section of published articles offer a window to academic informal ties (17, 20). In these notes, the authors recognize colleagues who provided support or feedback. These contributions are not substantial enough to warrant authorship, but they are deemed important enough to merit public appreciation. A qualitative analysis of $N \sim 3,700$ acknowledgment sentences found that thanking colleagues who contributed to the research is the most common feature in this section, followed by funding disclosures and disclaimers (22). Past work also highlights the lack of standardization in the use of acknowledgments, except for the common recognition of peer communication and intellectual debt (23). Tellingly, some researchers have proposed using acknowledgments to calculate measures of influence that complement citations in performance assessments (24–26). The acknowledgment sections recognize the contributions of 'helpful scientists', which broadens the conventional focus on productivity and citation counts (the usual benchmark of academic success) to also include helpfulness to others (another pathway of influence and indirect impact).

Research that parses the content of acknowledgment sections predominantly focuses on the number of 'commenters' and, generally, not the network of scholars these informal exchanges create (e.g., (27)). When the analyses include measures of academic success, the focus is on article-level metrics like citation counts (e.g., (16, 19, 28)), not scholar-level metrics of impact. Moreover, the few studies that consider acknowledgment networks have a narrow scope: they cover single subfields (e.g., Finance Economics, based on six purposively selected journals (17)) or focus on publications on narrow topics over short periods (e.g., wind energy publications over a 2-year window, (18)). If informal ties enable opportunities for valuable exchange, this narrow empirical focus is unlikely to capture the full context or downstream impact of the social capital those ties create.

The mechanisms explaining the creation of helpful, informal ties are varied — as varied as the motivations attracting researchers to academic pursuits (29). 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 informal connections matter? Does membership in this layer of social capital increase the publication impact of scholars' work?

Academic Networks in Political Science. As is common across academia (30), Political Science is a status-based hierarchical discipline with institutional stratification within the field, captured, for example, through the analysis of hiring and the placement of PhDs (31, 32). While the study of job mobility patterns focuses on networks formed by institutions sending and recruiting scholars (not networks connecting scholars themselves), the patterns suggest the presence of an

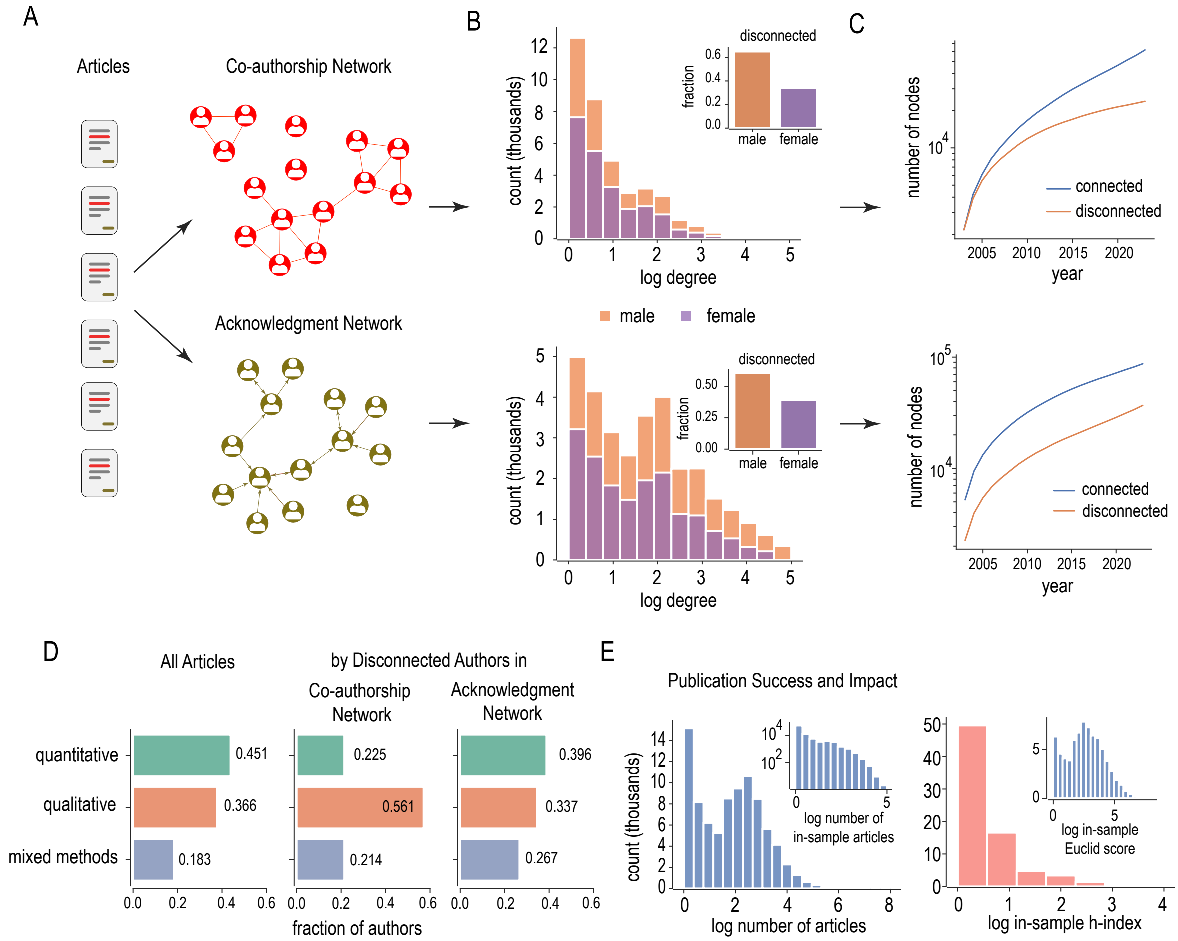

Fig. 1. Description of the Data. (A) We extract co-authorship and acknowledgment ties from published articles and build the corresponding networks. (B) Overlay bar charts with gender composition by binned degree; inset plots suggest about 28% and 30% of all scholars are disconnected in each network, 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.

underlying structure of interpersonal ties that shapes hiring decisions and, therefore, the flow of ideas and influence. These ties, anchored as they are in brick-and-mortar institutions, are different than the ties that leave a footprint in the acknowledgment section. Formal and informal ties, as captured by co-authorship and acknowledgment connections, may or may not have an overlap with the type of interpersonal exchange that happens routinely in the workplace. However, the prestige of these institutions likely affects which ties are formed in the first place.

Political science is also male-dominant. According to the largest professional association in the field, about 60% of members identify as men and 40% as women (33). This 3:2 ratio suggests that, merely due to compositional constraints, women will appear less frequently in co-authorship teams. Whether they are also less likely to be included as co-authors and listed only in the acknowledgments (as prior research suggests is the case in some STEM disciplines, (34)) is an open empirical question. The analysis of journal placement and citation patterns in Political Science has uncovered gender gaps (35–37), which are partly explained by differences in submission practices, and not necessarily bias (38).

Gender and institutional inequalities co-exist with other gaps, including those between North and Global South institutions (39). Whether formal and informal ties manage to span those divides and bring benefits to those connected is the question we consider here. Does this type of academic social capital help increase academic impact, regardless of gender or institutional affiliation?

To answer this question, we analyzed 129,750 articles published in 174 Political Science journals from 2003 to 2023 (see Materials and Methods). We built two longitudinal networks using information contained in the articles we analyze: one network maps the structure of co-authorships and the second network maps the structure of acknowledgments (Fig. 1A). For each author we collected additional information on their gender, seniority, the prestige of the institution to which they are affiliated, geographical region, and methodological orientation (see Materials and Methods). We also compiled three measures of 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 Figs. S6-S7 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, see Figs. S13-S14). The acknowledgment network has, on average, twice as many connections 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. Approximately 19% of all scholars in our data are disconnected from both networks, meaning they have no formal or informal connections with peers. In the SI (Fig. S15) we show that the correlation of scholars' degree centrality in the two networks is moderate, which suggests they occupy different positions in the two structures. In Table S4 we also show that the gender differences are minimal in the case of the co-authorship network: contingent on the 3:2 ratio, women have, on average, similar centrality to men. However, men are more central in the acknowledgment structure.

The plots in Fig. 1B-C expand on these descriptives. The acknowledgment network is less hierarchical, in terms of degree distribution, which means there are fewer extreme differences in the connectivity of scholars. Consistent with the membership figures provided by APSA, the largest professional association in the field, scholars are predominantly male in both networks (about 62% of all scholars). The insets in 1B show that this disparity is also present among the disconnected subset of scholars. Both networks have a similar growth rate (1C), but the acknowledgment network retains more disconnected scholars over time. Fig. 1D shows that most of the articles published by disconnected scholars in the co-authorship structure are qualitative, consistent with the different epistemological and technical requirements in that methodological subdomain; in contrast, 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 success: 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 next examine whether being disconnected from the informal network of scientific support 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 across deciles of the within-field productivity distribution. Because scholars with multidisciplinary profiles often publish outside Political Science, many of their articles fall outside our dataset. To account for this, we compute the ratio of articles published in the journals under study to a scholar's total output and assign scholars to deciles of this distribution, with higher deciles corresponding to those concentrating most of their publications 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 consistently have a lower impact, both as measured by the h-index and the Euclid scores (with connectivity assessed two years prior to the impact measure). 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 distribution. (In the SI, Figs. S6, S7, S26, S27 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).

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

Informal ties are the most relevant predictor of publication impact. In Fig. 3A-B, we present estimates from two generalized linear models of variation in the h-index and Euclid scores using matched data with balanced groups (see Materials and Methods). The strongest predictor of impact is being connected to the acknowledgment network (with a two-year lag, as in Fig. 2). Other predictors vary by impact measure: seniority outweighs gender in explaining Euclid scores, which are sensitive to career length; methodological approach also matters, with quantitative research more consequential for Euclid scores. Institutional prestige is consistently significant —scholars at top-100 universities, or based in North America or Europe, show higher impact—highlighting the role of academic stratification (39). Still, the most consequential covariates overall are co-authorship and, especially, informal ties: connections to helpful peers are the most robust correlate across metrics.

Panels 3C-D display the rank order of variable importance according to random forest models fitted to the complete 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 the 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

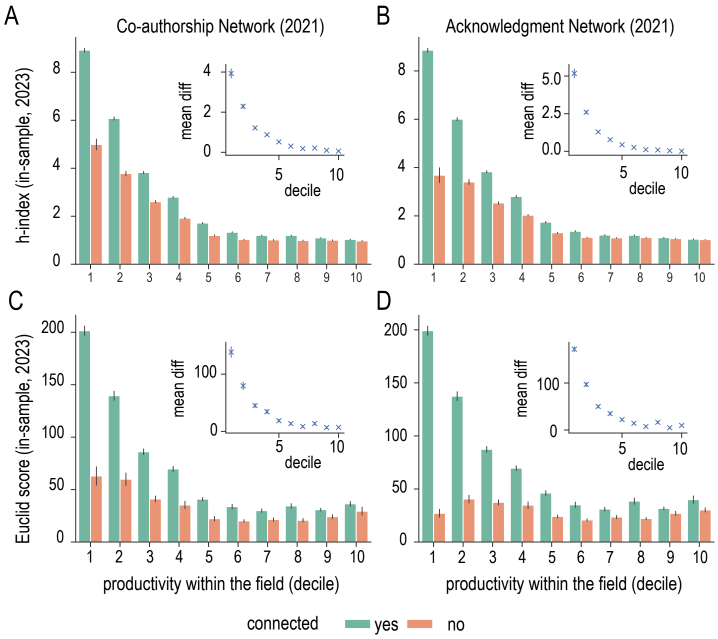

Fig. 2. Differences in Publication Impact for Each Productivity Decile. Average academic impact for connected (green) and isolated (orange) scholars by within field productivity decile (fraction of papers published in Political Science). Disconnected authors in both networks have, on average, lower impact 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.

in explaining variance in publication impact. The second most important correlate is being connected to co-authors, consistent with the linear models and with prior research (2, 3). In the SI, Fig. S25 and Tables S7-S10, we provide the full regression outputs and demonstrate the robustness of our findings to alternative model specifications.

In Fig. 4, we evaluate the importance of additional network features, including degree centrality (i.e., number of connections) and several measures of network clustering that assess similarity across connected scholars. For each focal scholar, we calculate the fraction of alters that share the same gender, seniority, country of institutional affiliation, and institutional prestige (in SI, Figs. S18-S23, we offer descriptive plots for all these measures of network homophily). Centrality in the acknowledgment network is again the most critical predictor of academic success, especially for the h-index (which is less sensitive than the Euclid score to one-hit publication track records). The SHAP plots in the right column reveal that higher values in the degree centrality of both networks, but especially the acknowledgment network, have the largest effect on the accuracy of the model predictions. In SI Figs. S28-S35, we report additional analyses that confirm the robustness of these findings to different modeling approaches and sensitivity tests.

In multivariate analyses, we can compare the importance of network connectedness only to the covariates included in the model specification. While gender, seniority, institutional prestige, methodological orientation, and country of institutional affiliation are the most emphasized correlates in the literature (e.g., (17, 35)), there could be other unobserved factors that are also predictive of academic success. However, to the extent 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 offer a form of social capital that allows scholars to draw advice and feedback from a broader range of colleagues than those formally linked through co-authorship. These informal relationships create the circuits of information exchange that sustain the invisible college. Our analyses show that these connections are key predictors of publication impact—but also that opportunities are unevenly distributed in how research communities self-organize. Informal ties function as sources of support and gateways to opportunity, yet their unequal accessibility reinforces stratification within academia.

Our evidence supports the view that informal structures of support shape access to knowledge, advice, and resources. Embeddedness in these structures divides scholars into two groups: those with informal connections, who benefit from added social capital, and those without. Of course, as we also show, those connected can still hold different positions within the structure, as assessed by their degree centrality or the composition of their personal networks, some being more homogeneous than others. The contrast between connected and disconnected scholars is stark: over a quarter of researchers in our data have no informal ties and perform significantly worse, while those with such ties —and especially those in more central positions— achieve higher academic impact. In future research, we plan on unpacking the different positions held by this connected subset of scholars, beyond the size and composition of their networks. But the evidence we present here offers robust support to the theoretical expectation that informal structures of support are key to explaining variation in academic success.

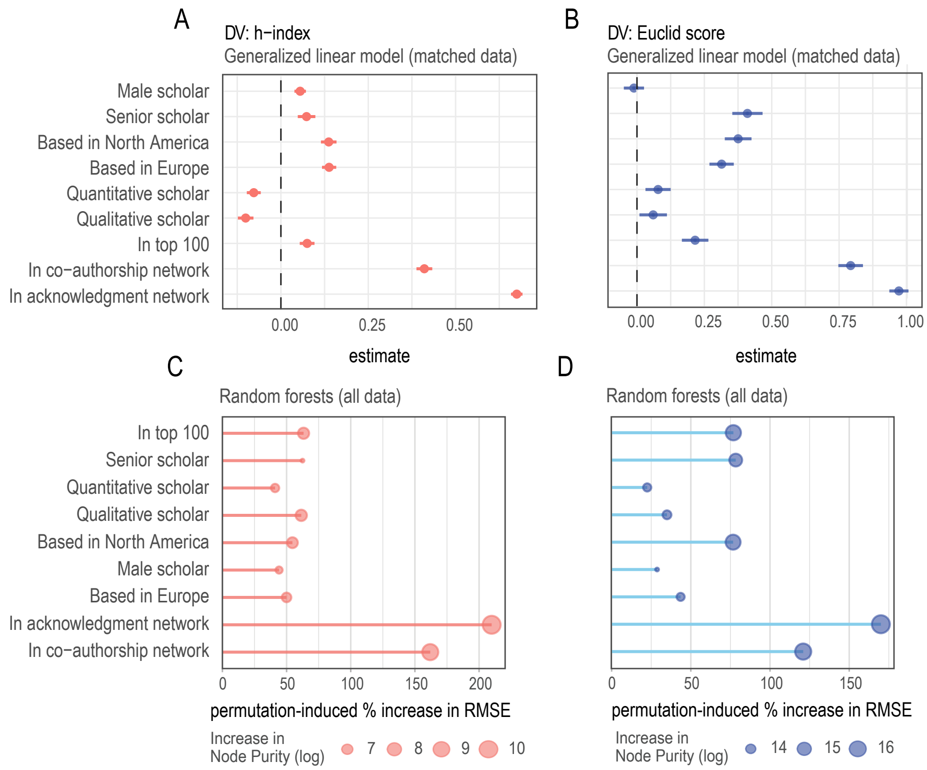

Fig. 3. Connectivity and Publication Impact. (A-B) Estimates of a generalized linear model using a matched dataset. (C-D) Variable importance according to random forests using all data. Regardless of the model and measure of academic impact, connectivity to the informal collaboration structure stands out as the most important predictor.

Our acknowledgment data offer broader coverage than past research, both in the number of journals and the length of the observation window. Only a handful of studies use acknowledgment sections to explain publication outcomes, and most face notable limitations. With few exceptions, prior work (e.g., (16, 19)) does not construct the overall network or compare it to more formal types of collaboration. Many studies also rely on the Web of Science, which indexes only acknowledgments containing funding information (e.g., (28)), likely introducing sampling bias. By contrast, we extract all available acknowledgment sections from every published Political Science article over two decades. When prior research does consider acknowledgment networks, it typically focuses on narrow subfields (e.g., (17)) or topic-specific publications over short periods (e.g., (18)). Moreover, it centers on paper-level measures of impact, such as citation counts for individual articles. Our focus on the overall impact of scholars allows us to analyze their embeddedness as a structural position that enables——or constrains——access to information and resources, thereby more directly capturing the role of social capital in shaping scholarly success.

Like all empirical research, our approach also has limitations. The acknowledgment network we analyze is a proxy to an underlying, latent 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 sure 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 exogenous variation, our estimates are 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 whether the gaps we identify here 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, and the double-masking peer review process mutes any strategic use of this section to manipulate the pool of reviewers. It is possible that the impact of helpful colleagues varies across fields of inquiry. Future work should consider other measurement strategies to capture these informal structures of academic support across disciplines more accurately.

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,

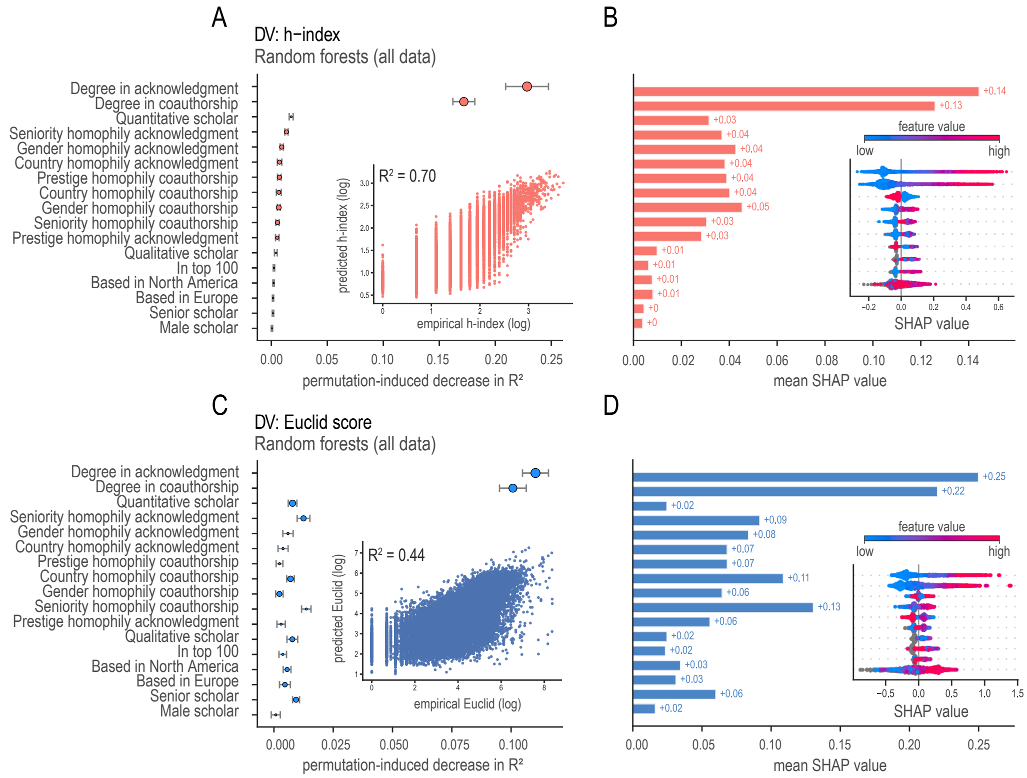

Fig. 4. Degree centrality and Publication Impact. (A-C) Variable importance to predict h-index and Euclid scores (insets show the correlation between observed and predicted values). (B-C) Average SHAP values in both models, which quantify the contribution of each variable to the prediction. Insets show the full distributions: color encodes high to low values in each variable (the further they are from 0, the higher the importance of each value); to improve visualization, the last row of the SHAP values aggregates the last eight variables. These plots reveal that higher values in the degree centrality of both networks but, especially, the acknowledgment network, have the greatest effect in the accuracy of the model predictions.

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 Supplementary Table S1 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 Fig. S4 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 section 15 for sensitivity analyses that use a 5-publication filter and no filter, which show our main conclusions remain unchanged). A significant fraction of the articles (57.56%) are single-authored and more than half (55.61%) have an acknowledgment section (see SI Figs. S2-S3 for additional information on the scholars and papers that do not use an acknowledgment section).

Measures of Publication Impact. We calculated the h-index and Euclid scores using only in-sample publications (see SI Fig. S5 for a description of the same metrics using all articles). The h-index (or hirsch-index (40)) is one of the most commonly used metrics of impact in the science of science (41). It is defined as:

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

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 (42) is defined as:

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

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. Previous research shows that, despite a bias towards western names, the classification tool has an overall accuracy of 96.6% (43). However, it is important to note that name-based gender inference methods have lower accuracy for certain demographic groups (44). Gender is also a complex social construct: the use of these classification methods may erase minority groups or non-binary populations (45).

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 2021. Junior authors are those who published their first article after 2011.

Institutional Prestige. We define institutional prestige using the Academic Ranking of World Universities, or Shanghai ranking. The Shanghai ranking returned the highest percentage overlap with the affiliations in our data compared to two alternative rankings (the QS World University and Times Higher Education rankings, see SI Fig. S8). In total, we retrieved ranking information for $N = 761$ institutions covering 47,619 scholars' affiliations, 55.59% of our data. Given the large number of unmatched cases, we operationalized the measure of institutional prestige as a binary variable that tracks whether a scholar is in a top 100 institution.

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 (35), 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 SI Tables S2-S3 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 (46), institutional prestige (31, 32), geographical location (47), and methodological orientation (48). To reduce imbalances we use a statistical method known as coarsened exact matching (49, 50). 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 SI Fig. S25 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 Tables S7-10 for full regression outputs and other model specifications.

Random Forest Regression. We use a machine learning algorithm known as random forests (51, 52) 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; they are also less sensitive to outliers. We use the percentage increase in root mean squared error, RMSE (Fig. 3), and the permutation-induced decrease in R^2 (Fig. 4) as the main measures of variable importance. Both measures quantify 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). These metrics are considered more reliable than other measures like increase in node purity, but we also report this secondary measure in Fig. 3 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. SHAP values, reported in Fig. 4, help identify which values in each variable contribute the most to the model's predictive accuracy. In the SI, sections 14-15 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. E Sarigöl, R Pfitzner, I Scholtes, A Garas, F Schweitzer, Predicting scientific success based on coauthorship networks. *EPJ Data Sci.* **3**, 9 (2014).
4. M Newman, The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.* **98**, 404–409 (2001).
5. O Vásárhelyi, I Zakhlebin, S Milojević, EÁ Horvát, Gender inequities in the online dissemination of scholars' work. *Proc. Natl. Acad. Sci.* **118**, e2102945118 (2021).
6. D Price, Networks of scientific papers. *Science* **149**, 510–515 (1965).
7. L Wu, D Wang, JA Evans, Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
8. Q Ke, AJ Gates, AL Barabási, A network-based normalized impact measure reveals successful periods of scientific discovery across disciplines. *Proc. Natl. Acad. Sci.* **120**, e2309378120 (2023).
9. DJ De Solla Price, D Beaver, Collaboration in an invisible college. *Am. Psychol.* **21**, 1011–1018 (1966).
10. D Crane, Social structure in a group of scientists: a test of the "invisible college" hypothesis. *Soc. Networks* pp. 161–178 (1977).
11. P Bourdieu, *The Forms of Capital*, ed. JG Richardson. (Greenwood Press, Westport, CT), (1983).
12. JS Coleman, *Social Capital.* (Harvard University Press, Cambridge, Ma), pp. 300–321 (1990).
13. N Lin, K Cook, RS Burt, *Social Capital: Theory and Research.* (Aldine de Gruyter, New York, NY), (2001).
14. 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).
15. JG Oliveira, AL Barabási, Darwin and einstein correspondence patterns. *Nature* **437**, 1251–1251 (2005).
16. 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).
17. ME Rose, CP Georg, What 5,000 acknowledgements tell us about informal collaboration in financial economics. *Res. Policy* **50**, 104236 (2021).
18. S Tian, X Xu, P Li, Acknowledgement network and citation count: the moderating role of collaboration network. *Scientometrics* **126**, 7837–7857 (2021).
19. Q Xie, X Zhang, Exploring the correlation between acknowledgees' contributions and their academic performance. *Scientometrics* **128**, 6003–6027 (2023).
20. M Doehne, C Herfeld, How academic opinion leaders shape scientific ideas: an acknowledgment analysis. *Scientometrics* **128**, 2507–2533 (2023).

21. 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).
22. A Paul-Hus, N Desrochers, Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015. *PLOS ONE* **14**, e0226727 (2019).
23. JA Teixeira da Silva, P Tsigaris, QH Vuong, Acknowledgments in scientific papers. *Publ. Res. Q.* **39**, 280–299 (2023).
24. B Cronin, SG Weaver, The praxis of acknowledgement: from bibliometrics to influmetrics. *Revista Espanola De Documentacion Cientifica* **18**, 172–177 (1995).
25. CL Giles, IG Councill, Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proc. Natl. Acad. Sci.* **101**, 17599–17604 (2004).
26. A Oettl, Reconceptualizing stars: Scientist helpfulness and peer performance. *Manag. Sci.* **58**, 1122–1140 (2012).
27. A Baccini, E Petrovich, Normative versus strategic accounts of acknowledgment data: The case of the top-five journals of economics. *Scientometrics* **127**, 603–635 (2022).
28. A Paul-Hus, P Mongeon, M Sainte-Marie, V Larivière, The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements. *J. Informetrics* **11**, 80–87 (2017).
29. E Petrovich, Acknowledgments-based networks for mapping the social structure of research fields. a case study on recent analytic philosophy. *Synthese* **200**, 204 (2022).
30. A Clauset, S Arbesman, DB Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1** (2015).
31. JH Fowler, B Grofman, N Masuoka, Social networks in political science: Hiring and placement of ph.d.s, 1960-2002. *Ps-Political Sci. & Polit.* **40**, 729–739 (2007).
32. HJ Kim, B Grofman, Job mobility, tenure, and promotions in political science phd-granting departments, 2002–2017: Cohort, gender, and citation-count effects. *PS: Polit. Sci. & Polit.* **52**, 684–690 (2019).
33. APSA, Gender, race, and ethnicity within apsa organized sections (2023).
34. C Sugimoto, V Larivière, *Equity for Women in Science: Dismantling Systemic Barriers to Advancement.* (Harvard University Press), (2023).
35. DL Teele, K Thelen, Gender in the journals: Publication patterns in political science. *Polit. Sci. & Polit.* **50**, 433–447 (2017).
36. ML Dion, JL Sumner, SM Mitchell, Gendered citation patterns across political science and social science methodology fields. *Polit. Analysis* **26**, 312–327 (2018).
37. KQ Hill, Citations to the publications of male and female political scientists revisited. *PS: Polit. Sci. & Polit.* **57**, 552–557 (2024).
38. PA Djupe, AE Smith, AE Sokhey, Explaining gender in the journals: How submission practices affect publication patterns in political science. *PS: Polit. Sci. & Polit.* **52**, 71–77 (2019).
39. J Barcelo, C Paik, P van der Windt, H Zhai, A global ranking of research productivity of political science departments. *PS: Polit. Sci. & Polit.* pp. 1–12 (2025).
40. JE Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci.* **102**, 16569–16572 (2005).
41. D Wang, AL Barabási, *The Science of Science.* (Cambridge University Press, New York, NY), (2021).
42. M Perry, PJ Reny, How to count citations if you must. *Am. Econ. Rev.* **106**, 2722–41 (2016).
43. AD VanHelene, et al., Inferring gender from first names: Comparing the accuracy of genderize, gender api, and the gender r package on authors of diverse nationality. *PLOS Digit. Heal.* **3**, 1–15 (2024).
44. JW Lockhart, MM King, C Munsch, Name-based demographic inference and the unequal distribution of misrecognition. *Nat. Hum. Behav.* **7**, 1084–1095 (2023).
45. S Shugars, A Quintana-Mathé, R Lange, D Lazer, Categorizing the non-categorical: the challenges of studying gendered phenomena online. *J. Comput. Commun.* **29** (2024).
46. 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).
47. 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).
48. V Larivière, Y Gingras, On the relationship between interdisciplinarity and scientific impact. *J. Am. Soc. for Inf. Sci. Technol.* **61**, 126–131 (2010).
49. SM Iacus, G King, G Porro, Causal inference without balance checking: Coarsened exact matching. *Polit. Analysis* **20**, 1–24 (2012).
50. S Iacus, G King, G Porro, cem: Software for coarsened exact matching. *J. Stat. Softw.* **30**, 1 – 27 (2009).
51. L Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
52. A Liaw, M Wiener, Classification and regression by randomforest. *R News* **2**, 18–22 (2002).