WHAT MAKES ECONOMETRIC IDEAS POPULAR: THE ROLE OF CONNECTIVITY

Bertrand Candelon, Marc Joëts, Valérie Mignon

REPRINT • 2024 / 04

UCLouvain

LiDAM
LFIN | Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

Research Policy 53 (2024) 105025

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

What makes econometric ideas popular: The role of connectivity ☆

Bertrand Candelon [a,b], Marc Joëts [c], Valérie Mignon [d,e,*]

[a] Louvain Finance, UCLouvain, Louvain-la-Neuve, Belgium
[b] Maastricht University, Netherlands
[c] IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, France
[d] EconomiX-CNRS, University of Paris Nanterre, Nanterre, France
[e] CEPII, Paris, France

ARTICLE INFO

JEL classification:
C01

Keywords:
Connectivity
Idea diffusion
Econometric publications
Citations
Structural Topic Model
Hurdle count model

ABSTRACT

This paper aims to identify the factors contributing to the diffusion of ideas in econometrics by paying particular attention to connectivity in content and social networks. Considering a sample of 17,260 research papers in econometrics over the 1980-2020 period, we rely on Structural Topic Models to extract and categorize topics relevant to key domains in the discipline. Using a hurdle count model, we show that both content and social connectivity among the authors enhance the likelihood of non-zero citation counts and play a key role in shaping the diffusion of econometric ideas. We also find that high topic connectivity augmented by robust social connectivity among authors or authoring teams further enhances econometric ideas' diffusion success. Finally, our findings unveil an inverted U-shaped relationship between connectivity and the success of idea diffusion; the latter initially escalates but starts to wane upon reaching a certain threshold.

1. Introduction

The skyrocketed amount of academic publications, particularly in the field of econometrics, has resulted in an overwhelming abundance of information, commonly referred to as the 'burden of knowledge' (Jones, 2009). As the academic landscape becomes increasingly saturated, establishing a unique intellectual space and securing peer recognition becomes progressively complex (Jones, 2009; Bloom et al., 2020; Deichmann et al., 2020). While the intrinsic quality of research is undeniably important for an academic career, it alone does not guarantee academic influence. In the field of econometrics, citations serve as a key metric for assessing scholarly impact (Uzzi et al., 2013; Wang, 2016; Archontakis and Mosconi, 2021). Nevertheless, it is worth recognizing that the variables affecting citation counts are multifaceted and do not solely hinge on the quality of the research. The notion of 'research quality' itself is a nuanced and somewhat elusive criterion that is subject to interpretation.

A fundamental concept in academic research is the Matthew effect, as identified by Merton (1968), wherein recognition and visibility tend to accumulate for established scientists, thereby enhancing their ability to garner further success and resources. This phenomenon not only affects the distribution of resources but also extends to the dissemination and reception of scientific contributions, where work presented by esteemed scientists often garners more attention and citation within the scientific community compared to contributions from lesser-known researchers.

Building upon this concept, our study acknowledges the crucial role of social and topic connectivity in the diffusion of ideas within econometrics, especially in the context of the Matthew effect. While Merton's seminal work provides a foundational understanding of how recognition and resources are concentrated among established scholars, it prompts further investigation into the mechanisms through which ideas and innovations spread within academic fields. Our research seeks to extend the narrative by empirically examining the nuanced ways in which social networks and thematic overlaps influence the visibility and impact of econometric research, beyond the general principle of accumulated advantage. By analyzing the interplay between social connectivity and thematic intersections, we aim to offer a comprehensive exploration of academic influence and success, thereby deepening our understanding of the dynamics at play beyond the broad strokes depicted by the Matthew effect.

☆ We would like to thank Emmanuel Flachaire, Francesco Roccazzella, and three anonymous referees for helpful comments and suggestions. Bertrand Candelon has conducted this research as part of the research program entitled "Financial and Extrafinancial Risks Modeling" under the aegis of the Europlace Institute of Finance, a joint initiative with Insti7. The usual disclaimers apply.

* Correspondence to: EconomiX-CNRS, University of Paris Nanterre, 200 avenue de la République, 92001 Nanterre cedex, France.
E-mail addresses: bertrand.candelon@uclouvain.be (B. Candelon), m.joets@ieseg.fr (M. Joëts), valerie.mignon@parisnanterre.fr (V. Mignon).

https://doi.org/10.1016/j.respol.2024.105025
Received 4 December 2023; Received in revised form 1 May 2024; Accepted 7 May 2024
Available online 17 May 2024

Our study relates to the literature on research collaboration, particularly in the context of co-authorship networks. Collaborative efforts among researchers have surged (see, e.g., Cronin et al. (2003, 2004)), forming intricate social networks that facilitate knowledge dissemination (Biscaro and Giupponi, 2014; Köseoglu, 2016; Ji et al., 2022). As recalled by Thelwall et al. (2023), this tendency has two main theoretical foundations: (i) the "historically influential theory" where larger projects and critical societal issues require collaborative efforts (Price, 1963; Gibbons, 1994; Ziman, 2000), and (ii) the existence of a dedicated academic field for research networks, namely team science (Hall et al., 2018).

Empirical evidence suggests that coauthored articles tend to garner more citations, which citation counts often increasing with the number of authors involved, especially in international collaborations (Glänzel et al., 1999; Glänzel, 2001; Glänzel and Schubert, 2001; Hara et al., 2003; Persson et al., 2004; Didegah and Thelwall, 2013; Khor and Yu, 2016; Wagner et al., 2019; Deichmann et al., 2020; Ding et al., 2021; Shen et al., 2021). While the number of coauthors per paper has steadily increased in all disciplines during the 21st century (Fanelli and Lariviere, 2016; Parish et al., 2018), it should, however, be mentioned that some heterogeneity exists across disciplines (see Moed (2005), Parish et al. (2018), Thelwall and Maflahi (2022) and Thelwall et al. (2023) for a brief survey). As noticed by Thelwall and Maflahi (2022), single authorship is most common in mathematics and the arts and humanities, whereas the number of authors is higher in the natural and medical sciences compared to the social sciences. Furthermore, experimental physics is a field in which the association between citation counts and the number of coauthors is stronger due to the projects' large and intensive collaborative nature (Parish et al., 2018).

Our paper falls into this strand of the literature and aims to identify the variables affecting citation counts, elucidating the factors that contribute to the prominence of ideas in econometrics. This specific field deserves particular attention for several reasons. Indeed, econometrics is used in all hard and soft sciences disciplines. For instance, its interdisciplinary nature intersects with economics, finance, statistics, mathematics, and data science. Besides, econometrics covers both theoretical and empirical research, making it widely used in the academic literature. Furthermore, it acts as an empirical foundation for a range of disciplines, from economics and finance to sociology and political science, or even to climatology or neuroscience, underscoring its extensive academic impact.

Prior research across various disciplines suggests that groundbreaking ideas often arise from a blend of pre-existing knowledge. Specialized expertise, if too narrowly focused, can stifle creativity, leading to minor, incremental advances (Uzzi et al., 2013). Conversely, works that integrate diverse areas of knowledge introduce innovative approaches and resonate more broadly within the scientific community (Uzzi et al., 2013; Trapido, 2015; Wagner et al., 2019). Such works often achieve higher citation rates (Kaplan and Vakili, 2015; Deichmann et al., 2020) as they serve as informational shortcuts, connecting disparate research areas through the lens of small-world network theory.

Simultaneously, the social network positioning of authors also influences the acceptance of ideas within academia (McFadyen and Cannella, 2004; Ductor et al., 2014; Wang, 2016). An author's centrality in academic networks bolsters the impact and credibility of their work (Podolny, 2001; Azoulay et al., 2010; Ductor, 2015; Deichmann and Jensen, 2018; Hsieh et al., 2018). Recognizing the importance of diversified perspectives, research in econometrics is increasingly collaborative, fostering interdisciplinary innovation (Andrikopoulos et al., 2016; Jones, 2021). Both ideas and social connectivities at individual and team levels substantially influence the dissemination success of research contributions.

Despite abundant research focusing on the technical aspects of econometrics, its social and relational aspects remain relatively underexplored. A few studies have explored co-authorship patterns in economics (Goyal et al., 2006; Nowell and Grijalva, 2011) and econometrics (Andrikopoulos et al., 2016). Our study seeks to fill this gap by incorporating insights from network theory, social psychology, natural language processing, and data analytics. In line with Deichmann et al. (2020), this paper explores how various forms of connectivity influence the trajectories of econometric theories and practices. We aim to identify the 'hidden bridges' that propel the field forward, thereby providing crucial guidance for scholars navigating in an intricate academic landscape.

Following Deichmann et al. (2020), we argue that the intrinsic quality of an idea is not the sole determinant of its academic dissemination. Instead, 'connectivity' – in its various forms – plays a pivotal role in shaping the diffusion and recognition of econometric ideas. Based on this foundation, we propose the following testable hypotheses:

(i) High thematic/ideas connectivity exerts a positive influence on the successful diffusion of an econometric concept.
(ii) Enhanced social connectivity among the contributing scholars significantly increases the likelihood of successful diffusion.
(iii) An interplay exists between thematic and social connectivity, with well-connected authors or teams more effectively propagating thematically integrated works.
(iv) Idea diffusion success initially increases with both social and content connectivity but declines after hitting a certain threshold.

To empirically assess these hypotheses, we analyze over 17,000 research articles published in leading econometrics journals over the past four decades. Utilizing Structural Topic Models (STM), we categorize themes relevant to key domains in econometrics, including 'structural break', 'factor models', and 'unit root and cointegration'. This enables us to explore both the topical content and temporal evolution of each idea. Consistent with our hypothesis, we posit that publications bridging multiple domains are more likely to gain prominence within the scientific community. To quantify this, we introduce an 'idea connectivity' index for each publication to measure the interlinking of various domains. We rely on measures of betweenness centrality within a two-mode network, as discussed in Borgatti and Everett (1997) and Everett and Borgatti (2005). A high betweenness centrality score signals robust ideas/topics connectivity, serving as an indicator of a publication's role as a significant bridge in the academic landscape. In addition, we examine the 'social connectivity' of individual authors and collaborative teams by using data related to each publication's authorship. In social connectivity, a high betweenness centrality score indicates that an author or a team of authors occupies a central and strategic position within the shortest paths connecting contributors in the academic social network. As a preliminary step, our methodological approach distinguishes itself by being the first, to our knowledge, to systematically identify creatively boundary-crossing publications in the field of econometrics over the last four decades. Furthermore, our study highlights individuals and teams considered to be pivotal idea generators.

To explore how connectivity shapes the success of idea diffusion as gauged by citation counts, we employ a hurdle count model. This model accounts for overdispersion and excess zeros commonly found in scientific citation data (Mullahy, 1986; Cameron and Trivedi, 2005, 2013). Our findings confirm that connectivity significantly enhances the likelihood of non-zero citation counts and is positively correlated with ideas' diffusion, supporting our hypotheses across various time horizons and dimensions. The influence of social connectivity is particularly pronounced at the team level, although the interaction between different types of connectivity varies depending on the empirical framework. Overall, our results confirm that connectivity plays an important role in making econometric ideas popular, these findings being robust to the several robustness checks we run.

In summary, our paper offers several significant contributions to the existing body of literature. First, to the best of our knowledge, we are the pioneers in establishing a nuanced relationship between idea success and connectivity within the specialized domain of econometrics. While previous studies such as Deichmann et al. (2020) have explored this link, they have done so in different discipline – the Semantic Web research community, i.e., a sub-field of computer science – and have employed distinct approaches and perspectives.

Second, we are the first to employ topic modeling techniques to consistently measure the evolution of ideas in econometrics over an extended period. Although this approach has been applied in economics and finance (Larsen and Thorsrud, 2019; Hansen et al., 2018; Brunetti et al., 2024), our study distinguishes itself by adopting a "meta" perspective. Specifically, we consider the field of econometrics in its entirety, synthesizing ideas across a broad range of studies to provide a more comprehensive and holistic understanding of the subject matter. This enables us to identify and analyze the principal ideas that have shaped econometrics over the past 40 years. By leveraging two-mode network centrality metrics, we introduce novel indexes for idea and social connectivity at both individual and team levels. These indexes capture the innovative nature of publications and the extent to which authors and teams are integrated into the scientific community. While our study aligns with the findings of Andrikopoulos et al. (2016) concerning the consequences of scientific collaboration, we approach the topic from a unique angle, focusing on the popularity of ideas.

Finally, we provide an empirical examination of the role of connectivity in shaping the success of econometric ideas. Our results illuminate the multifaceted influences on research impact, extending beyond research quality to include the roles of knowledge and social networks. We offer a clear roadmap for understanding how a publication can gain prominence by bridging different academic domains, particularly when produced by credible and well-connected authors. Overall, our work not only sheds light on the factors contributing to scientific popularity but also paves the way for future research, offering a fresh perspective on scholarly impact in the field of econometrics.

The rest of the paper is organized as follows. Section 2 details the data and metrics used to measure econometric ideas. Section 3 outlines the empirical setup, including the connectivity scores and variables. In Section 4 we analyze the role of connectivity in shaping idea diffusion. Section 5 presents robustness checks and sensitivity analyses, and Section 6 offers concluding remarks.

2. Data and measurements of econometric ideas

This section outlines the database of research publications employed for analyzing the diffusion of ideas, explains the natural language processing approach utilized in measuring econometric ideas, and presents preliminary results from idea estimation.

2.1. Original database

We have constructed a unique database to analyze the evolution of econometric ideas over time, gathering papers published by 11 leading econometric journals over the last 40 years (1980–2020) (see Chang and McAleer (2013) and Appendix A for more details). Using the Web of Sciences Database (WoS), we retrieved 17,260 research publications from these journals after filtering out proceeding papers, editorial notes, and early access papers.

Selecting articles published in the 11 top-tier journals in econometrics may lead to a selection bias problem.[1] Indeed, as noted by Leahey et al. (2017), while the innovation literature suggests that spanning disciplines is beneficial because it allows researchers to establish connections across fields, the literature on categories considers it detrimental in the sense that domain-spanning ideas face a higher acceptance hurdle—the resulting research being potentially viewed as of lower quality. Since our sample is confined to the articles published in 11 top-tier econometric journals, the domain-spanning econometric ideas we observe are only a sub-sample of econometric ideas—those published in top-ranked journals. Consequently, we may encounter a selection bias problem as we undersample unsuccessful domain-spanning ideas.

Such an issue is inherent to all studies dealing with ideas' success and publications. However, it is worth mentioning that this argument can be counterbalanced because our set of articles comes from a coherent body of literature that aligns with our research interests. As emphasized by Biscaro and Giupponi (2014), among others, restricting the set of articles we consider allows us to obtain a quite homogeneous population of papers, fitting with our privileged field of research, better capturing network effects, and limiting variability in the data.[2]

As shown in Table 1, the distribution of the sample is not homogeneous among journals, with JoE (23.1%), REStat (15.4%), and *Econometrica* (14%) accounting for over half of our records. Based on the H-index,[3] these three journals also have the highest impact factor. The extensive time frame enables us to encompass various developments and a wide range of research topics in modern econometrics.

Furthermore, while the number of published papers has been increasing over the years, the level of the H-index peaked between the mid-1990s and mid-2000s and collapsed after (see Figure A1 in Appendix A). This trend indicates a substantial lag between publications and citations. Our large sample provides enough time to consider the diffusion of these ideas. In the empirical Section 4, we investigate this diffusion across various time dimensions (such as two-, six-, and ten-years).

From the 17,260 research papers, we extracted Abstracts, Titles, and Keywords to form a corpus for our econometric ideas estimation. We also collected metadata for each paper, including journals and authors. These variables, described in Section 3.2, are used as control variables in the empirical model of idea diffusion. Overall, our data include contributions from 13,852 individual authors from 1926 institutions across 87 countries.

We transformed the 17,260 research articles into a document-term matrix, describing the frequency of terms within the collection of documents. This matrix, although high-dimensional and sparse (17,260 documents and 80,024 terms with 95% scarcity), is preprocessed by reducing dimensionality. Words and characters with little topical content (e.g., stopwords,[4] numbers, mathematical formulas, and punctuation) were removed, and the remaining terms were stemmed,[5] leaving a $(17{,}260 \times 45{,}714)$ document-term matrix for our topic model estimation.

2.2. Estimation of econometric ideas through topic modeling

We measure the evolution of econometric ideas within our dataset of research publications using probabilistic topic models. Specifically, we employ a mixed-membership approach to extract topics from papers, postulating that econometric ideas are well-represented by these topics. This approach allows each research paper to encompass multiple topics, each characterized by a collection of words.

Latent Dirichlet Allocation (LDA), introduced by Blei et al. (2003), is a widely-used technique for topic modeling. It assumes each document is a mixture of topics, where each topic is a distribution of words, both generated from Dirichlet distributions. Key parameters include the document-topic proportions, denoted as θ_d^k, for documents d across

[1] We would like to thank a referee for pointing out this issue.

[2] This is accentuated by the fact that the domains we defined are within the discipline, as we aim to focus on economics.

[3] Recall that a journal's H-index is a metric that measures the productivity and citation impact of the publications within that journal (Hirsch, 2005). It is defined as the number of articles in a journal that have received at least H citations each over a given period.

[4] The stopword list we used is from http://snowball.tartarus.org/algorithms/english/stop.txt, and is available upon request to the authors.

[5] The stemming algorithm is the Porter stemmer implemented in R.

Table 1
Top econometric journals (1980–2020).

Journals	Period	N. of articles (% of total)	H-index
Econometric Reviews (ER)	Jan. 05–Dec. 20	597 (3.4%)	40
Econometric Theory (ET)	Apr. 88–Dec. 20	1412 (8.5%)	86
Econometrica	Jan. 80–Nov. 20	2427 (14%)	279
Econometrics Journal (EJ)	Jan. 05–Sept. 20	363 (2%)	38
Journal of Applied Econometrics (JAE)	Jan. 87–Dec. 20	1455 (8.5%)	109
Journal of Business Economic Statistics (JBES)	Jan. 85–Dec. 20	1593 (9.6%)	131
Journal of Econometrics (JoE)	Jan. 80–Dec. 20	4034 (23.1%)	207
Journal of Financial Econometrics (JFE)	Mar. 07–Dec. 20	287 (1.6%)	36
Journal of Time Series Analysis (JTSA)	Sep. 00–Dec. 20	865 (5.3%)	45
Oxford Bulletin of Economics and Statistics (OBES)	Feb. 80–Dec. 20	1437 (8.3%)	86
Review of Economics and Statistics (REStat)	Feb. 80–Dec. 20	2660 (15.4%)	187

topics k, and the topic-word proportions, denoted as β_k^w, for topics k across words w. There parameters approximate posterior distributions, facilitating the assignment of topics to documents and words to topics.[6] However, LDA's limitation is its assumption of independent topics within a document, which may not hold in complex research contexts.

To address this issue, we utilize the STM developed by Roberts et al. (2013). STM estimates topic and word distributions similarly to LDA but differs by drawing θ_d from a Logistic-Normal distribution and modeling β_k using multinomial logit. This method accounts for dependence between topic distributions and allows for endogeneity with respect to certain factors.[7] Our focus is on the role of connectivity in shaping econometric ideas, leaving the origins of these ideas for future research.

The STM algorithm, considering known topics within a document, proceeds as follows:

1. Draw the document-topic distribution for a given research paper randomly from a Logistic-Normal distribution as:

   $$\theta_d \mid X_{d\gamma}, \Sigma \sim \text{Logistic-Normal}(\mu = X_{d\gamma}, \Sigma)$$

   where X_d stands for a vector of covariates, $\gamma \sim N(0, \sigma_k^2)$ is a matrix of coefficients, and Σ is the covariance matrix. Note that in our approach we do not consider exogenous factors in modeling topics' distributions.
2. For each word in the research paper:
   - Select one topic from the distributions obtained in Step 1.
   - Using multinomial logit, choose a word corresponding to the selected topic as:

     $$\beta_{d,k} \propto \exp(m + \kappa_v^k + \kappa_v^y + \kappa_v^{y,k})$$

     where m is the baseline word frequency, and $(\kappa_v^k + \kappa_v^y + \kappa_v^{y,k})$ is a collection of coefficients.
3. Repeat Steps 1 and 2 iteratively to generate a set of research papers, each defined by a set of topics that best describe the document.

The key quantities are estimated using a semi-collapsed variational EM algorithm, setting the number of topics K at 60. This choice balances the model's accuracy and interpretability. To determine the optimal dimension, we tested the model with K ranging from 20 to 80, evaluating four statistical metrics:[8] held-out likelihood (Wallach et al., 2009), residual checks (Taddy, 2012), the lower bound, and semantic coherence (Mimno et al., 2011), as shown in Figure B1 in Appendix B.1. These criteria suggested an optimal K between 60 and 80. Following Roberts et al. (2014), Figure B2 in Appendix B.1 combines semantic coherence and word exclusivity to compare models with K = 60, 70, and 80 topics, leading to the selection of K = 60 for its interpretability and statistical robustness.

2.3. Preliminary analysis of econometric ideas

We conducted a 60-topics STM analysis on the document-term matrix comprising econometric publications from January 1980 to December 2020. The model yields topics and word proportions that map the development of econometric ideas throughout the period. Since STM does not assign specific labels to each topic, we followed the methodology of Brunetti et al. (2024) and labeled the topics using two approaches: (i) the top 10 FREX (FRequency and EXclusivity) terms; and (ii) the most probable bigrams.[9] Details on the labeling methods are in Appendix B.2, and Tables B1 to B4 in the same Appendix report the labels.

Among the 60 topics, we selected several to showcase the variety of econometric ideas. The tables in Appendix B.2 identify topics beyond the scope of econometrics, reflecting the broad spectrum covered by the sampled journals. To sharpen our focus on econometric-oriented discussions, we refined our selection to 27 pivotal topics, leading to a (17,260 × 27) document-topic distribution matrix. This refinement was meticulously validated through multiple analyses. Initially, we scrutinized the top words and most probable bigrams for each topic, as delineated in Appendix B.2, to identify those inherently related to econometrics. Subsequently, our preliminary selection was rigorously tested by constructing network diagrams and applying Louvain community clustering. These methods affirmed the strong interconnectivity among the chosen econometric topics, substantiating their relevance and centrality to our analysis.[10]

Using a 5-year rolling window to examine the monthly aggregated time evolution of selected topics, we find cyclical patterns, marked responsiveness to significant economic and financial events, and discernible long-term trends.[11] Specifically, "Topic 8: Structural Break" and "Topic 49: ARCH & GARCH Models" have seen spikes in attention, particularly around pivotal moments such as the Black Monday crash in 1987, the Asian crisis in 1996–97, the Dot-com bubble in 1999–2000, and the Global Financial Crisis in 2007–09. However, they have not garnered much attention recently. Conversely, certain topics like "Topic 47: Unit Root and Cointegration" and "Topic 59: Statistical Inference" held importance over extended periods but have gradually waned in interest, becoming more of forgotten paradigms. Additionally, the rise of the big data era has lent prominence to "Topic 41: Factor Model",

[6] For applications in economics and finance, see Hansen and McMahon (2016), Hansen et al. (2018), and Larsen and Thorsrud (2019).

[7] See Brunetti et al. (2024) for further discussion.

[8] See Roberts et al. (2019) for more details.

[9] Topic labeling facilitates the discussion but does not materially affect the analysis.

[10] Detailed results of these validation steps are available upon request.

[11] See Figure B3 in Appendix B.2. The 5-year window size accommodates the fact that it may take time for a topic to emerge, thus capturing both the timing and trends of each topic. Aggregation is computed by calculating the average proportion for each topic over the months.

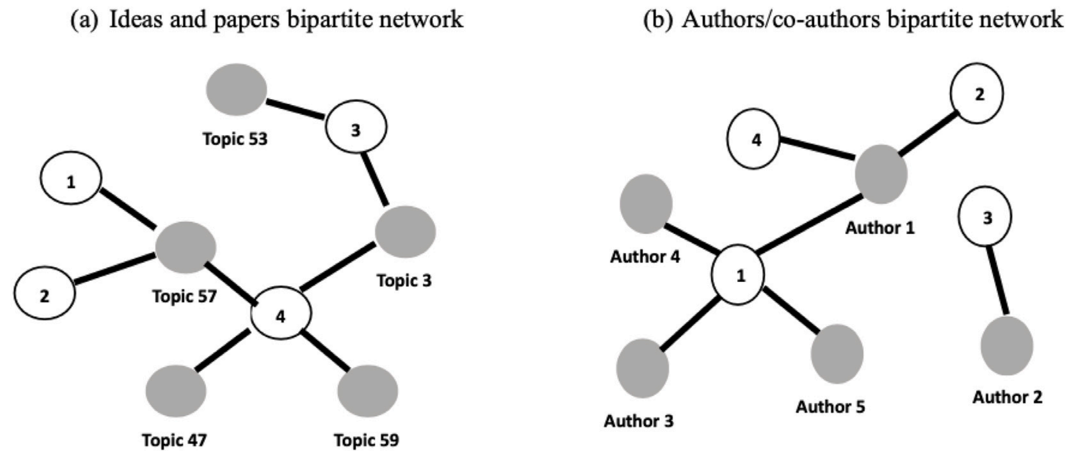

Fig. 1. Two-mode network examples in research publications. Note: The figure reports illustration of two-mode network for topics connectivity (Panel (a)), and social connectivity (Panel (b)).

while growing spatial interdependence has elevated "Topic 5: Spatial Autoregressive Model". Both have evolved from niche areas to growing fields, as documented by Stock and Watson (2017) and Sarafoglou and Paelinck (2008).

This preliminary analysis offers a detailed and multifaceted view of the evolving landscape of econometric ideas, illustrating that econometrics is a diverse field made up of various sub-disciplines, each defined by its unique terminology and conceptual framework. The idea of connectivity between topics, as well as social ties within the scientific community, may influence the diffusion of these ideas. In the next section, we will delve deeper into these connections to understand how they could potentially shape the success of different ideas within the field.

3. Empirical design

This section describes our set of variables, discusses our measures of topics and social connectivity, and presents our empirical setting to analyze econometric ideas' diffusion.

3.1. Dependent variables: ideas' success

In this study, we examine the ingredients of success of econometric ideas in the form of scientific publications. Scientific publications codify knowledge, provide certification, and are a broadcast medium of ideas in the community.[12] In the academic world, the popularity of an idea can be defined by the diffusion across the scientific community through the amount of citations a publication received from peers. Following previous works, we therefore consider the amount of citations extracted for each paper from WoS database as a proxy for ideas' success (see, Uzzi et al. (2013), Magerman et al. (2015), and Deichmann et al. (2020)). We adopt static and dynamic perspectives by considering the total amount of citations over the period (static) and a two-, six-, and ten-year moving window counts (dynamic).[13] It is crucial to clarify that although the terms "scholarly impact", "ideas' success", "popularity", and "diffusion" may carry distinct conceptual meanings, they are used interchangeably in this paper. Collectively, these terms aim to capture the academic recognition and dissemination influence of publications, as demonstrated by citation counts.

As shown in Figure B4 in Appendix B.3, while the amount of total citations is mainly concentrated at the beginning of the sample (between 1980 to 1999), for both 6- and 10-years windows a larger proportion (with few exceptions) is clustered at the end of the sample (from 2005 onward). This illustrates that it takes time for ideas to diffuse in the scientific community and the importance to consider different time windows. In line with this observation, frequency plots show that citations are over-dispersed with an important concentration of zeros. Last published papers therefore need time to attract attention. Over-dispersion and excess zeros are important properties of ideas' success that we try to capture in Section 3.3.

3.2. Independent variables: the role of connectivity

As independent variables, we hypothesize that the role of connectivity plays a significant role in the diffusion of ideas (Uzzi et al., 2013; Trapido, 2015; Deichmann et al., 2020). We consider both topics and social connectivity as explanatory variables. Connectivity, though an abstract concept, is concretely operationalized in our framework through the lens of academic influence. We posit that publications serve as critical nodes when they bridge multiple ideas within the scholarly network. Topic connectivity refers to a publication's ability to connect these different ideas, serving as a conceptual bond. Social connectivity, on the other hand, is defined by a publication's role in linking authors, who themselves may be connected to other works, thus weaving a more intricate web of scholarly communication. Through this dual lens, publications that act as both conceptual and social bridges are anticipated to achieve heightened visibility and citation within the academic landscape.

The two connectivity measures are derived from two-mode networks (i.e., bipartite networks) created from our research publications database. A two-mode network consists of two types of nodes and ties that only belong to nodes of different sets (Borgatti and Everett, 1997; Opsahl, 2013). Fig. 1 reports two visual examples of bipartite networks to measure connectivity in our context.[14] For topics connectivity (Panel (a)), the first set consists of papers (white nodes), the second set is the topics proportion for each paper (grey nodes), and the connections between papers and topics show how each publication acts as a bridge. For instance, we expect Publication 4 to have a high topics connectivity

[12] See Deichmann et al. (2020).

[13] While considering dynamics through moving window counts of citations, we do not integrate formal time dynamic with lags. It is worth mentioning that our dynamic approach allows us to guard against the potential bias associated with the date at which the article is published during a year—i.e., the fact that papers published early in a year may be more cited than articles published later. As Biscaro and Giupponi (2014) noticed, considering fixed time windows to measure citations has several advantages. In particular, it is an adequate measure to assess the scientific impact within a discipline and is not biased toward old articles that can be cited for a more extended period.

[14] We borrow the intuition from Deichmann et al. (2020).

since it acts as a bridge between several ideas and make the link between Publications 1, 2, and 3. Social connectivity (Panel (b)) works in the same way, the more a publication is connected to other papers and authors the higher is the score. We describe further how these measures are computed considering all authors and co-authorships respectively.

As a measure of connectivity, centrality scores provide a consistent approach to analyzing networks. The literature introduces various metrics for this purpose, including degree, eigenvector, local clustering, k-core, and betweenness centrality, each offering a distinct perspective on network connectivity.[15] Degree centrality evaluates the number of direct connections a node has, shedding light on its local activity and influence. In contrast, eigenvector centrality assesses the quality of these connections, emphasizing the significance of being linked to influential nodes. k-core decomposition exposes the network's hierarchical structure by identifying densely interconnected groups, thus underscoring the network's robust areas. The local clustering coefficient measures how likely a node's neighbors are to form a tight-knit community, reflecting localized cohesion. Together, these metrics provide comprehensive insights into network dynamics, from the importance of individual nodes to the overall structure and robustness of the network.

Our paper specifically advocates for using betweenness centrality as a measure of connectivity, due to its ability to identify nodes that act as crucial bridges within the network. By focusing on betweenness centrality, we can recognize the existence of connections and comprehend the strategic role of nodes in facilitating the flow of information and integration across various segments of the network. Consequently, betweenness centrality emerges as a crucial metric for analyzing topics and social connectivity and for understanding the dissemination of ideas within complex systems, offering profound insights into the operational framework of scientific networks. Further analysis, which compares network centralities using the cumulative distribution function (CDF) and correlation as reported in Figure C3 in Appendix C.2 and discussed in Section 5, confirms the rationale behind our approach to measuring topic and social connectivity.

3.2.1. Topics connectivity

To measure topics connectivity, we start from the estimated $(17{,}260 \times 27)$ documents-topics distribution matrix discussed in Section 2.2, which reports for each publication the proportion of each econometric idea (i.e., topics).[16] We then derived a weighted two-mode network in which each publication is related to a set of ideas.

As bipartite networks are rarely analyzed in their original form for convenience, we apply a projection method by compressing the two-mode structure into a one-mode format. This procedure is performed by defining the set of nodes X (either publications or topics) and linking two nodes from X if they were connected to the same node (see, Newman (2001), Seierstad and Opsahl (2011), and Opsahl (2013)). Appendix C.1 provides an illustration of bipartite projection. We use the overlap count method for compression, which consists of counting the number of nodes in the first mode that each pair in the second mode has in common. Section 5 and Appendix C.1 provide a robustness check.

After publication projection, we calculate for each of them the betweenness centrality score, measuring how often a node is a bridge between other nodes on all shortest paths. We follow Borgatti and Everett (1997) and Deichmann et al. (2020) and define the betweenness centrality of node i as:

$$b_i = \frac{1}{2} \sum_{k \neq i}^{n} \sum_{\neq i \neq k}^{n} \frac{p_{kj(i)}}{p_{kj}}, \tag{1}$$

where p_{kj} denotes the total number of shortest paths from node k to j (geodesic path), and $p_{kj(i)}$ is the number of geodesic paths from k to j passing through i (where i is not an endpoint). We use the Brandes algorithm for the computation of the betweenness centrality (Brandes, 2001).

Building on the insights from Panel (a) of Fig. 1, the topic connectivity score quantifies how a publication serves as a bridge within the scientific community, linking disparate econometric ideas.[17] Fig. 2 presents the topic connectivity in the form of a two-mode network between publications and ideas in Panel (a). In the network, the vertices represent two different types of entities: 'ideas', whose size is proportional to their prominence, and 'publications', which are scaled according to their betweenness centrality scores. Larger nodes for ideas signify a more prominent role of the corresponding econometric idea throughout the entire sample period. Nodes representing publications with higher topic connectivity scores are depicted distinctly, in red, within the figure. For the sake of simplicity, the figure includes only publications with high betweenness centrality scores.[18] As the figure emphasizes, publications with high betweenness centrality are pivotal in connecting different econometric ideas, highlighting their influential role in the flow and dissemination of knowledge within the field. For example, Publications

- 8411_ID: "Realized Variance and Market Microstructure Noise" by P.R. Hansen and A. Lunde, *Journal of Business & Economic Statistics* 24(2) 2006.
- 15818_ID: "Estimating the Integrated Volatility with Tick Observations" by J.Jacod, Y. Li and X. Zheng, *Journal of Econometrics* 208(1) 2019.
- 16731_ID: "On the Estimation of Integrated Volatility in the Presence of Jumps and Microstructure Noise" by C. Brownlees, E. Nualart and Y. Sun, *Econometric Review* 39(10) 2020.

exhibit the highest topic connectivity, acting as conduits among T37, T45, T49, and T55, and forming connections with a multitude of other publications. Additionally, ideas that are prominently featured in econometrics often engage widely across various publications, underscoring their central role in the scholarly discourse of the field. Interestingly, the publications mentioned above are primarily oriented toward finance-related themes. For a broader perspective, Panel (b) of Fig. 2 illustrates the betweenness centrality for each publication, highlighting the most pivotal papers in red and depicting the average betweenness centrality with a black dotted line. It is noteworthy that multiple publications achieve a significant degree of topic connectivity.

3.2.2. Social connectivity

For social connectivity, two measures are considered and generated from binary bipartite network. First, we consider all authorship roster and created a two-mode network from a $(17{,}260 \times 33{,}792)$ matrix where publications are connected to each author. Second, to account for the rise of teams and the role of co-authorships in ideas' diffusion, we constructed a two-mode network from a $(11{,}230 \times 12{,}393)$ matrix where papers are connected by shared 'teams' (see Deichmann et al. (2020), and Jones (2021)). 'Teams' within the context of social connectivity are

[15] Burt's constraint is omitted due to its narrow focus on positional advantage, missing broader aspects like connectivity and influence, which are crucial for our complex network analysis.

[16] Structural topic models inherently estimate topic distributions. To focus on the most salient topics for each publication, we consider only topics whose representation exceeds the publication's mean topic weight.

[17] For further details, Section 5 and Appendix C.3 provide sensitivity analyses on betweenness centrality measures using various algorithms.

[18] As a result, certain topics with significant proportions, such as Topic 59: Statistical Inference & UR, may appear to be unrelated to any publications in the figure. However, they are, in fact, well connected within the broader network context.

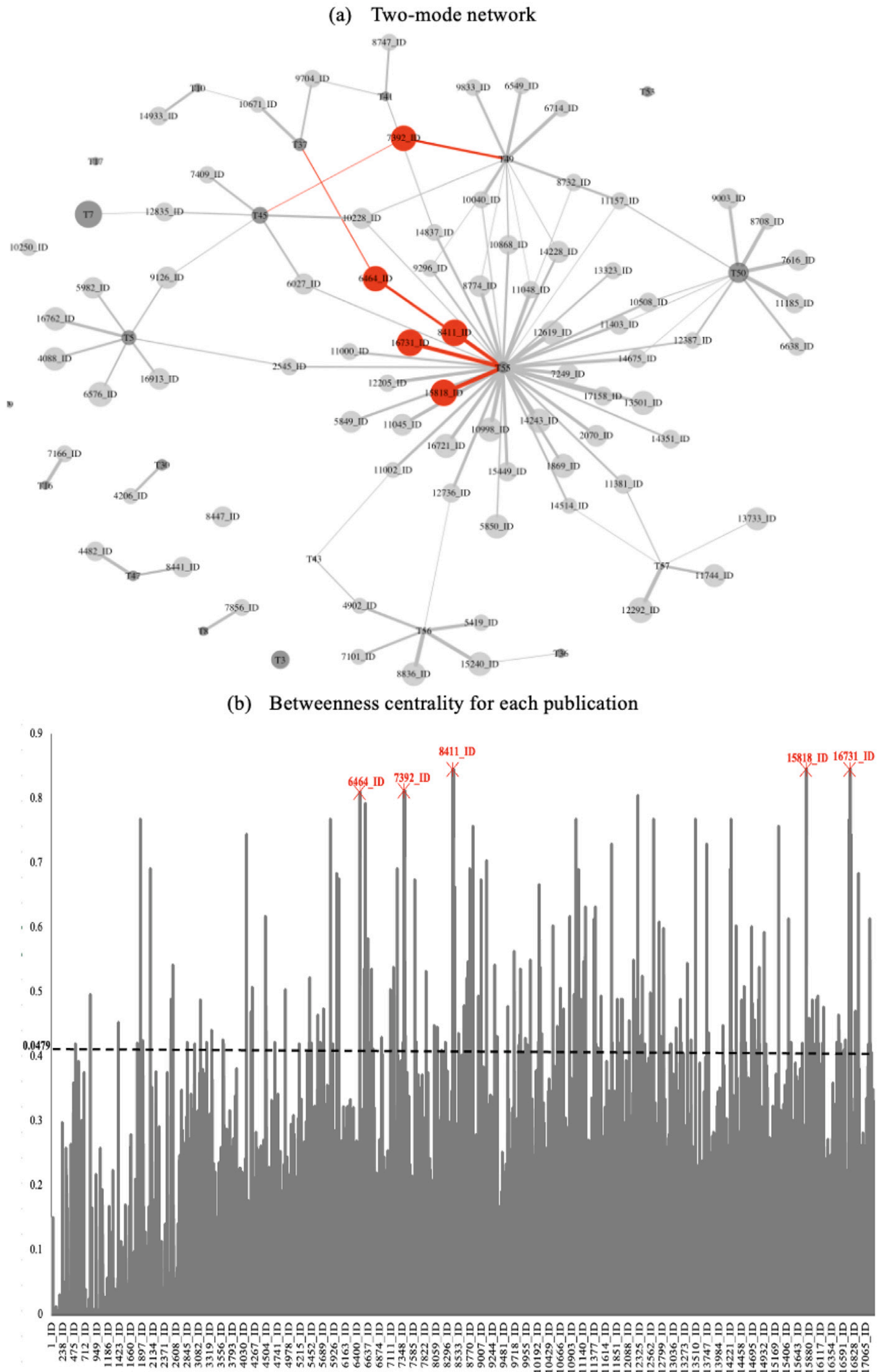

Fig. 2. Topic connectivity (authorship roster). Note: In Panel (a), the figure depicts a weighted two-mode network between publications (in light gray) and econometric ideas (in dark gray). The size of the vertices indicates the publications' betweenness centrality and the total proportion of ideas, respectively. Vertices colored in red highlight the highest topic connectivity score and its associated edges. For clarity, vertices with a betweenness centrality below 0.500 (out of a maximum of 0.844) and edges with topic probabilities below 0.1 have been omitted. In Panel (b), the graph illustrates the betweenness centrality for each publication, with the highest centrality marked in red and the average centrality represented by a black dotted line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

defined by emphasizing the collaborative efforts of authors on research papers, specifically through the lens of co-authored publications. This approach underlines the importance of collective academic endeavors in fostering social networks within the research community.[19] As for topic connectivity, after overlap count publications projection we computed betweenness centrality score using Eq. (1) to capture social connection between authors and teams respectively (see Appendix C.1 for robustness checks). A publication exhibiting high social connectivity, as measured by betweenness centrality, signifies its importance as a juncture within the network of scholars. Such a publication is likely to be linked to a diverse array of authors, acting as a conduit for scholarly interaction and contributing to the interconnectedness of the academic community.

[19] To have a consistent framework, we also compute topic connectivity for co-authored publications separately.

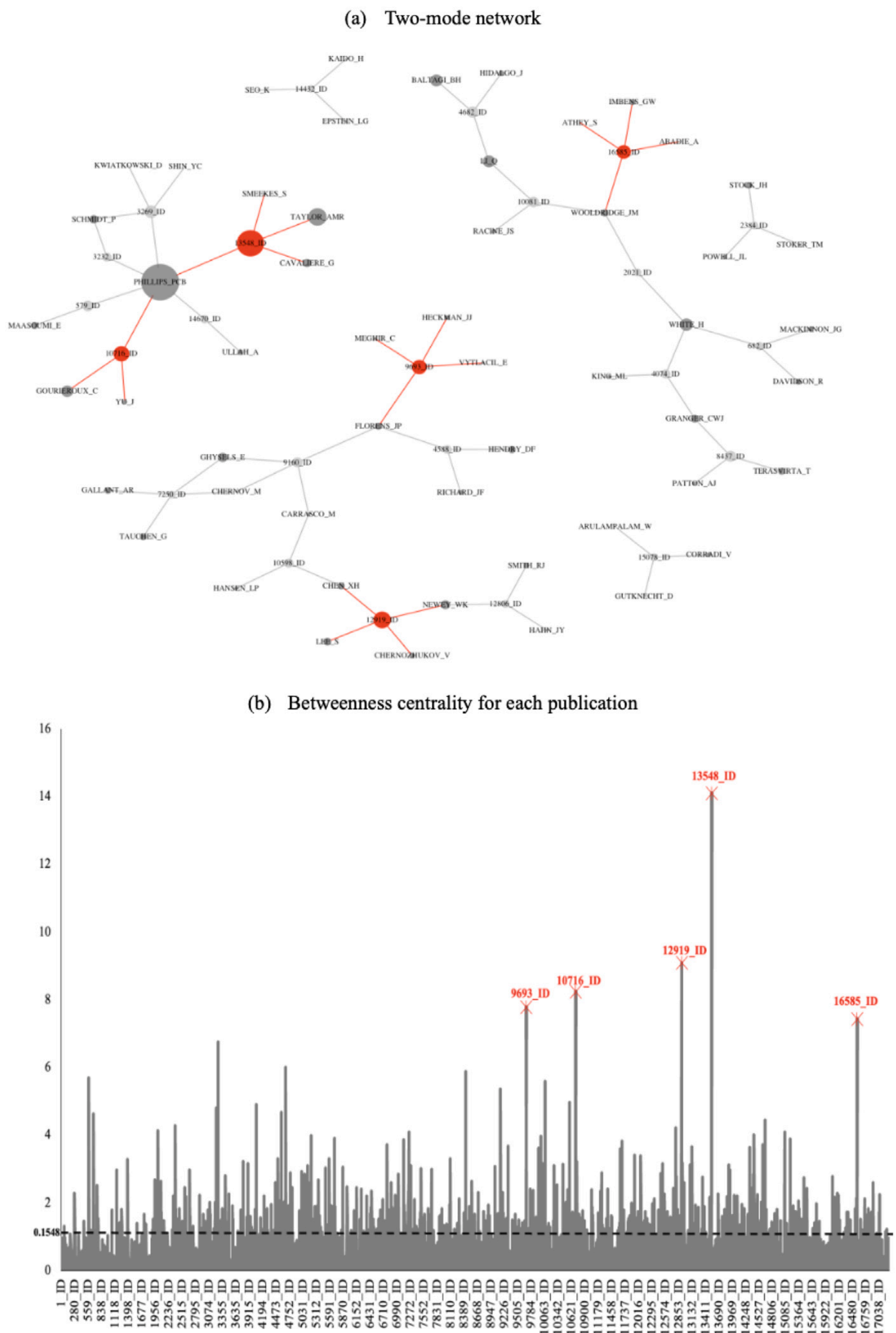

Fig. 3. Social connectivity (authorship roster). Note: In Panel (a), the figure depicts a weighted two-mode network between publications (in light gray) and authors (in dark gray). The size of the vertices indicates the publications' betweenness centrality and the authors' total contribution over the period, respectively. Red vertices color indicates the highest social connectivity score and its connected edges. For clarity, vertices with a betweenness centrality below 4.000 (out of a maximum of 14.104), and isolated nodes have been omitted. In Panel (b), the graph illustrates the social connectivity for each publication, with the highest centrality marked in red and the average centrality represented by a black dotted line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Panel (a) in Fig. 3 illustrates the betweenness centrality in networks of authorship roster, represented as binary two-mode networks.[20] This figure conveys two distinct layers of information. First, the size of the authors' nodes is proportional to their overall contributions to the scientific community, which is measured by the number of publications they have authored.[21] Notably, P.C.B. Phillips, A.M.R. Taylor, H. White, B.H. Baltagi, and C. Gouriéroux are identified as the most significant contributors to highly socially connected publications. This is evident at both the full authorship and co-authorship levels. Second, the size of the publications' nodes represents the varying levels of social connectivity associated with each paper. Publications with higher betweenness centrality are depicted with larger vertices, indicating that they act as important bridges in the network, connecting various

[20] Using networks of co-authorship instead of authorship roaster leads to similar results, as displayed in Figure B5 in Appendix B.3.

[21] In these figures, we include only publications with a betweenness centrality score above 4.000. It is important to note that several other prominent authors with substantial total contributions are present in our sample, but they are not displayed in the network due to this threshold.

authors and teams. For example, highest betweenness publications in red in Fig. 3, 13548_ID (with a score of 14.104) exhibits the highest social connectivity score. This publication serves as a nexus, connecting P.C.B. Phillips with several other authors (S. Smeekes, A.M.R. Taylor, and G. Cavaliere), and is linked to other publications (579_ID, 3229_ID, 3232_ID, 10716_ID, and 14670_ID). The critical role of connectivity becomes even more pronounced when considering teams of authors (co-authored publications only), as demonstrated by a score of 16.237 for publication 13548_ID in Figure B5, which is also highlighted in red. As a comparison, Panel (b) in Fig. 3 reports the betweenness centrality for each publication, emphasizing the publications with the highest social connectivity in red and the average connectivity in black.

3.2.3. Control variables

In testing our hypotheses, we considered a number of control variables to alleviate for possible confounding factors and omitted biases. In order to capture a possible journal effect, we first created a set of dummy variables for each journal as:

$$J_u = \begin{cases} 1 & \text{if the publication } p \text{ is published by the journal } u \\ 0 & \text{otherwise.} \end{cases}$$

where J_u denotes the journal u among the list of econometrics journals reported in Table 1.

Second, the size of the team may be an important factor to attract attention as suggested by Reagans and Zuckerman (2001) and Deichmann et al. (2020). We included, as a count variable, the number of authors per publication. Third, a publication citing many of his peers can be seen as a well-established work in the scientific community (Uzzi et al., 2013). We added the number of cited references for each publication to control for this effect.

Finally, prolific and skilled authors – benefiting from the Matthew effect – can enhance the visibility of specific ideas or topics discussed in a given publication (Wang, 2016). To control for this effect, we focus on the top 20 most prolific authors over the period based on the H-index (see Table A2 in Appendix A), and constructed a dummy variable as:

$$A_p = \begin{cases} 1 & \text{if at least one author of the publication } p \text{ is a top author} \\ 0 & \text{otherwise.} \end{cases}$$

where A_p is the author(s) of the publication p.

The absence of multicollinearity among all variables was confirmed using a pairwise correlation matrix, as shown in Figure A2 in Appendix A.

3.3. Model description

The endogenous variables of citations taking integer values, to test for our hypotheses we therefore rely on (static) count data models. Still, the distribution of the count variables can be apprehended in several ways, and multiple forms of specifications are possible.

The initial point is the Poisson regression.[22] While it is over-restrictive, such an assumption presents the advantage of simplicity. It takes the following form:

$$E[y_p|x_p] = \lambda_p = \exp(x_p^{\mathrm{T}}\beta) = \exp(\beta_1 + \beta_2 x_{2p} + \beta_3 x_{3p} + \beta_4 Z_{4p}), \tag{2}$$

where y_p stands for total, two-, six-, and ten-years moving window citations counts respectively over $p = 1, \ldots, P$ publications. x_{2p} is for topics connectivity, x_{3p} is for social connectivity, and Z_{4p} is for the set of control variables. $E[.|.]$ corresponds to the traditional conditional expectation operator. Using the generalized linear model framework, Eq. (2) can be re-written as a log-linear representation:

$$\ln(E[y_p|x_p]) = \ln(\lambda_p) = x_p^{\mathrm{T}}\beta = \beta_1 + \beta_2 x_{2p} + \beta_3 x_{3p} + \beta_4 Z_{4p}. \tag{3}$$

[22] See Cameron and Trivedi (2005).

The Poisson model is a restricted case of generalized linear models which imposed the conditional variance to be equal to the conditional mean. Thus, it imposes the dispersion to be fixed to 1. As shown by Figure B4 in Appendix B.3, ideas' success is however mainly over-dispersed with excess zeros, invalidating the restriction imposed by the Poisson model. We therefore consider several alternative models to tackle this issue. First, we assume a negative binomial distribution for $y_p|x_p$ which can arise as a gamma mixture of Poisson distributions.

Nevertheless, even if the negative binomial deals with over-dispersion, it does not model excess zero counts that are heavily present for the latest publications. This necessitates models capable of capturing both the count aspect and the zero component. A suitable candidate is the two-component hurdle model (see Mullahy (1986), Cameron and Trivedi (2005) and Cameron and Trivedi (2013)), which integrates a left-truncated count data model ($f_{count}(y|x,\beta)$) for positive counts with a right-censored zero hurdle model ($f_{zero}(y|x',\beta')$) for zero counts as:

$$f_{hurdle}(y|x,x',\beta,\beta') = \begin{cases} f_{zero}(0|x',\beta') & \text{if } y = 0 \\ \frac{1 - f_{zero}(0|x',\beta').f_{count}(y|x,\beta)}{1 - f_{count}(0|x,\beta)} & \text{if } y > 0. \end{cases} \tag{4}$$

From expression (4) and Eq. (3), the corresponding mean regression is

$$\ln(E[y_p|x_p]) = \mathrm{x}_p^{\mathrm{T}}\beta + \ln(1 - f_{zero}(0|x_p',\beta')) - \ln(1 - f_{count}(0|x_p,\beta)), \tag{5}$$

where y_p is the dependent variable (citations), x_p and x_p' are the independent variables for each component respectively (topic connectivity, social connectivity, and control variables) over $p = 1, \ldots, P$ publications. β and β' are parameters for each model. Negative binomial distribution is considered for the count component.

The implicit assumption underlying the use of the hurdle model in our context is the belief that zeros primarily result from the same process, specifically the insufficient time for recent publications to be cited. However, the occurrence of zero citations can stem from various factors other than just a lack of time, such as the absence of quality or relevance in the publications. In this context, the zero-inflated model may offer a compelling alternative to consider. Lastly, to address dispersion in count data potentially unrelated to an excess of zeros, exploring the Conway–Maxwell–Poisson (COM-Poisson) model might be beneficial.

To ensure the statistical robustness of our analysis, we have implemented and compared the models previously discussed, namely the Poisson, COM-Poisson, Negative Binomial, Zero-inflated Negative Binomial, and Hurdle Negative Binomial. A detailed comparison and sensitivity analysis across all considered citation windows are presented in Tables C1 to C4, and Figures C5 to C7 in Appendix C.4.1. This analysis encompasses: (i) dispersion tests; (ii) comparisons of log-likelihood and Bayesian Information Criteria (BIC); (iii) Raftery's BIC probabilities; (iv) Vuong's test; and (v) rootogram analysis. The findings consistently affirm the Hurdle model's superiority in modeling citation counts, which is why it is emphasized as a core result of the paper.

4. The role of connectivity in shaping ideas' diffusion success

Using citation counts, we investigate how connectivity shapes the diffusion of econometric ideas. Initially, we examine the impact of all authors, after which we focus on the role of co-authorship. To ensure comparability, we standardized our variables before incorporating them into the hurdle-negative binomial regression models. The coefficients for the count (non-zero) model were estimated using the negative binomial distribution and are reported as exponential transformations. The zero (hurdle) coefficients were estimated using logistic regression and represent the probability of non-zero citations. To clarify and enhance the interpretability of our model, we opted for a global interpretation framework that does not rely on any reference category. This goal was accomplished by excluding the intercept and maintaining all dummy variables in our regression analyses. This adjustment allows for a direct interpretation of each dummy variable's impact, free from the baseline comparisons typically necessitated by an intercept.[23] It is also important to stress that the estimated models are static, and therefore it is not possible to include time-dependent factors such as greater team volatility. These longitudinal aspects are left for future research.

4.1. Authorship roster

The results of our hurdle-negative binomial regression analysis for the complete list of authors are presented in Table 2.[24] The hurdle component (zero count) is reported in the lower table. Positive (negative) coefficients indicate that an increase in the regressor increases (decreases) the probability of a non-zero count. The count component (non-zero) is reported in the upper table. We interpret coefficients as percentage changes, calculated as $(\exp^{\beta_k} - 1) \times 100$ (reported coefficients are already expressed in exponential form).

Overall, our findings affirm that connectivity significantly enhances the probability of non-zero counts and is positively correlated with ideas' diffusion, corroborating Hypotheses 1 and 2. However, this effect exhibits variation between static and dynamic analyses and is dependent on the chosen time window.
From a static perspective (total citations in column (1)), the hurdle component reveals that social connectivity has an insignificant role, while topic connectivity notably amplifies the probability of having non-zero total citation counts by 51.8%. This indicates that publications bridging various knowledge domains are more likely to be cited. The control variables, such as cited count and prolific authors, also contribute positively to non-zero citations, with probabilities of 49.6% and 55.0%, respectively. Moreover, the type of journal is a significant factor in augmenting the probability of non-zero counts, with REStat (69.1%), *Econometrica* (66.5%), and JoE (65.4%) being the front-runners.

Regarding the count component, publications integrating diverse econometric ideas positively contribute to diffusion by 7.2%. As expected, mainstream econometrics journals with high impact factors (refer to Table 1) are positively associated with total citations, as evidenced by percentages for *Econometrica* (137.9%), JoE (84.6%), REStat (89.8%), JBES (50.6%), and JAE (41.3%). Publications by highly prolific authors also significantly bolster the popularity of ideas, averaging a 14.7% increase.

When considering different time windows (columns (2) to (4)) to analyze the temporal stability of our findings, the results are more nuanced. While the coefficients from the hurdle component remain fairly stable over time, emphasizing the importance of topics connectivity, the effects from the count component oscillate as the time window adjusts. Within a short time horizon (two-year window in column (2)), the topic connectivity score and all control variables positively contribute to ideas' success. In contrast, during longer time frames (six- and ten-year windows in columns (3) and (4)), social connectivity becomes a significant and positive factor in ideas' diffusion, taking about six years to be acknowledged as an influential factor in citations.

In our findings, we note a nuanced effect of social connectivity on citation counts. While the total citation analysis suggests a non-significant and potentially negative influence, specific time windows reveal an opposite effect. This discrepancy underscores the complex dynamics of social connectivity's impact over time, suggesting that its influence on scholarly dissemination may vary across different citation accumulation stages. Such findings prompt a deeper examination of how social ties affect citation trajectories, reflecting the intricate relationship between social connectivity and academic impact.

Another intriguing observation is that the number of authors, which initially contributed negatively to ideas' diffusion (−3.80%), turns out to be positively correlated with citations across different time windows (17.5%, 18.8%, and 18.4% for two-, six-, and ten-year periods, respectively). This observation sets the stage for a deeper examination of the role of co-authorship in ideas' diffusion.

4.2. The role of co-authorship

The influence of co-authorship on ideas' diffusion is explored in Table 3, utilizing the same interpretation grid as employed earlier. For this analysis, we have specifically considered co-authored publications, necessitating the recalculation of both topic and social connectivity scores.

Our analysis reaffirms the hypothesis that connectivity has a positive correlation with ideas' diffusion at both static and dynamic levels. Diverging from prior observations, the concept of team connectivity is unveiled as an indispensable contributor to this phenomenon. Alongside the recognized influences of topic connectivity, journal selection, and individual author contributions, the collaboration within a team emerges as a catalyst for idea propagation. The evidence supporting this conclusion spans multiple metrics, revealing a consistent pattern. For total citations, the hurdle and count models demonstrate increases of 57.2% and 22.1%, respectively. This trend continues over varied time frames, with a two-year period showing 51.7% and 4.6%, a six-year period at 98.0% and 6.4%, and a ten-year period yielding 61.0% and 10.5%. These figures not only substantiate the general hypothesis but also illuminate the nuanced way that co-authorship fosters intellectual cross-pollination.

Furthermore, the results elucidate that ideas conceived and nurtured by teams with robust connections within the co-authorship network exhibit a heightened propensity to attract citations. This effect is not merely incremental; it is accentuated at the team level. This emphasizes the collective intellectual capital and collaborative synergy within a team, which appears to be a driving force in achieving greater academic resonance. In essence, the data paints a compelling portrait of co-authorship as not just a peripheral factor, but a core mechanism in the dissemination and recognition of scholarly ideas as confirmed by the literature (see Introduction).

4.3. Topics and social connectivity interaction

Hypothesis 3 suggests that high topic connectivity, when coupled with strong social connectivity among authors or teams, leads to greater dissemination success. To evaluate this, we incorporated an interaction term (social × topic connectivity) in our models at both the authorship and co-authorship levels, integrating it with the main effects of topic and social connectivity.

Tables B5 and B6 in Appendix B.3.1 present the estimated coefficients for the interaction terms, consistent with our findings in Tables 2 and 3. Notably, the interaction effect varies by citation count and authorship level, significantly boosting six- and ten-year citation counts at the authorship roster level. This demonstrates the synergistic potential of blending topic and social connectivity to enhance ideas' spread. This synergy, as further supported by statistical comparisons between Tables B5 and Table 2, suggests that combining these dimensions can significantly increase ideas' dissemination. At the co-authorship level, the interaction term shows significance for total citations, thereby increasing the model's accuracy compared to the estimations in Table 3. However, the impact of the interaction term is not uniformly observed, indicating its effects are contingent on the specific context.

Our findings offer a detailed exploration of how social and topic connectivity contribute to disseminating scholarly work. The analysis

[23] Analyses incorporating the intercept, albeit excluding one dummy variable, have been conducted and led to the same conclusion. These results are available upon request from the authors.

[24] Sample sizes for each regression are reported after outliers in citations have been excluded.

Table 2
Connectivity and ideas' diffusion (authorship roster).

	(1) Total	(2) 2-years	(3) 6-years	(4) 10-years
	Count model coefficients			
Connectivity measures				
Social connectivity	0.998	1.012	1.021*	1.025*
Topics connectivity	1.072***	1.067***	1.089***	1.135***
Control variables				
# Authors	0.962**	1.175***	1.188***	1.184***
Cited count	1.098***	1.368***	1.456***	1.563***
Econometrica	2.379***	1.639***	1.583***	1.609***
JoE	1.846***	1.376***	1.321***	1.350***
REStat	1.898***	1.497***	1.400***	1.384***
OBES	1.313***	1.131***	1.089***	1.077***
JBES	1.506***	1.282***	1.176***	1.159***
JAE	1.413***	1.234***	1.168***	1.177***
ER	0.986	1.118***	1.025	0.999
ET	1.273***	1.128***	1.079***	1.112***
EJ	1.050**	1.093***	1.033*	1.022
JFE	1.033	1.108***	1.056***	1.043*
JTSA	0.991	1.040***	1.004	0.993
Prolific authors	1.147***	1.098**	1.121***	1.113***
	Hurdle model coefficients			
Connectivity measures				
Social connectivity	0.494	0.507	0.491	0.506
Topics connectivity	0.518**	0.519***	0.543***	0.556***
Control variables				
# Authors	0.496	0.569***	0.608***	0.611***
Cited count	0.496***	0.648***	0.736***	0.771***
Econometrica	0.665***	0.639***	0.670***	0.697***
JoE	0.654***	0.547***	0.567*	0.583*
REStat	0.691***	0.545***	0.573***	0.620***
OBES	0.545***	0.492	0.498	0.510
JBES	0.556***	0.518	0.524	0.547
JAE	0.584***	0.514	0.531	0.553
ER	0.485	0.497	0.471*	0.485
ET	0.537*	0.495	0.479	0.483
EJ	0.507	0.494	0.492	0.513
JFE	0.512	0.496	0.492	0.505
JTSA	0.506	0.500	0.500	0.525
Prolific authors	0.550***	0.526***	0.541***	0.567***
Sample size	17,245	17,259	14,614	10,968
Log-likelihood	-78885	-40230	-52855	-45543
BIC	158100	80802	106046	91413

Note: This table reports estimations of hurdle-negative binomial model. The exponential function is applied to coefficients of the count component, and the Plogis function is applied to coefficients of the hurdle component to convert log-odds into probabilities. ***, **, and * denote significance at 1%, 5% and 10% levels, respectively.

underscores that the combined effect of social connections and thematic focus – captured by the interaction term – can significantly influence citation counts, though this impact is not consistent across all scenarios. This observation highlights the variable strategic role of well-integrated authorship networks in enhancing academic influence. It underscores the complexities inherent in academic collaboration, illustrating how co-authorship can be pivotal in extending the reach and effect of research. This analysis advocates for a comprehensive understanding of how effective collaboration networks, coupled with thematic alignment, can drive scholarly dissemination.

4.4. Nonlinear effect of connectivity

To test our fourth hypothesis, which posits that both topic and social connectivity can exert a nonlinear effect on idea diffusion, we augment our hurdle negative binomial model with a squared term for connectivity.[25] Table B7 in Appendix B.3.2 presents the results for the authorship roster.[26] Consistent with our previous findings, both topic and social connectivity significantly increase the likelihood of receiving non-zero citation counts, as well as contribute to the overall success of idea diffusion—most notably at two, six-, and ten-year window sizes. Intriguingly, the nonlinear component unveils a downward-concave relationship between connectivity and idea success. As both topic and social connectivity increase, the success of idea diffusion initially rises but ultimately starts to decline upon reaching a certain threshold.

[25] Higher-order polynomial transformations were found to be insignificant. While we also considered Generalized Additive Models, we opted for the hurdle model due to the interpretability of its coefficients.

[26] Results for co-authorship are very similar and are available upon request from the authors.

Table 3
Connectivity and ideas' diffusion (co-authorship).

	(1) Total	(2) 2-years	(3) 6-years	(4) 10-years
	Count model coefficients			
Connectivity measures				
Social connectivity	1.221***	1.046***	1.064***	1.105***
Topics connectivity	1.071***	1.062***	1.086***	1.136***
Control variables				
# Authors	0.911***	1.077***	1.073***	1.052***
Cited count	1.053**	1.355***	1.428***	1.550***
Econometrica	1.800***	1.471***	1.503***	1.434***
JoE	1.326***	1.217***	1.248***	1.157***
REStat	1.490***	1.345***	1.368***	1.258***
OBES	1.047	1.045	1.013	0.995
JBES	1.205***	1.165***	1.132***	1.047
JAE	1.170***	1.153***	1.154***	1.085*
ER	0.868***	1.079***	1.001	0.944
ET	1.029	1.044	1.032	0.996
EJ	0.956***	1.053***	1.014	0.986
JFE	0.938***	1.064***	1.051***	0.954
JTSA	0.520	0.931	0.924	0.757
Prolific authors	1.022***	1.004	1.012	1.010
	Hurdle model coefficients			
Connectivity measures				
Social connectivity	0.572***	0.517***	0.980***	0.610***
Topics connectivity	0.513***	0.518***	0.605***	0.553***
Control variables				
# Authors	0.450***	0.542***	0.528*	0.526
Cited count	0.447***	0.639***	0.735***	0.762***
Econometrica	0.648***	0.654***	0.720***	0.741***
JoE	0.651***	0.558***	0.540	0.451
REStat	0.707***	0.558***	0.546	0.509
OBES	0.545**	0.502	0.483	0.436
JBES	0.547**	0.519	0.498	0.458
JAE	0.592***	0.529*	0.526	0.481
ER	0.485	0.501	0.451*	0.415***
ET	0.546***	0.498	0.459	0.411***
EJ	0.502	0.494	0.481	0.469
JFE	0.520	0.501	0.477	0.465
JTSA	0.522	0.522	0.471	0.356
Prolific authors	0.553	0.504	0.558	0.561
Sample size	11,613	11,621	9,461	6,622
Log-likelihood	-52944	-28453	-35344	-28420
BIC	106216	57234	70009	57148

Note: This table reports estimations of the hurdle-negative binomial model. The exponential function is applied to coefficients of the count component, and the Plogis function is applied to coefficients of the zero component. ***, **, and * denote significance at 1%, 5% and 10% levels, respectively.

5. Additional results and robustness checks

We conducted a comprehensive set of robustness checks, which are detailed in Appendix C, to assess the sensitivity of our results to different model specifications.

First, we scrutinize the sensitivity arising from bipartite projection, a key factor that transforms a two-mode network into a one-mode network for calculating ideas and social connectivity variables. To this end, we utilized various projection methods – including matching, Jaccard, and Pearson – and compared their similarity scores to the method employed in this study. These comparative analyses are depicted in Figure C2 in Appendix C.1. Our results consistently demonstrate a high degree of concordance with the methodology adopted in this paper.

Second, we evaluate the robustness of our betweenness centrality measure, which adeptly captures the intricacies of ideas and social connectivity in multiple dimensions. We have evaluated the behavior of our connectivity measure against alternative centrality scores commonly referenced in the literature, including degree, eigenvector, k-core, and local clustering. Figure C3 in Appendix C.2 presents a comparison of each measure using CDFs and Pearson correlation matrices for both topic and social connectivities. Our findings validate the rationale behind adopting the betweenness centrality score as a measure of connectivity, specifically defined by brokerage. In addition to using the Brandes algorithm (Brandes, 2001, 2008) for our study, we have also explored alternative algorithms, such as the approximate betweenness algorithm (Geisberger et al., 2008). Figures illustrating normalized betweenness centrality measures derived from each algorithm are provided in Figure C4 in Appendix C.3. Our analyses confirm the reliability and consistency of the betweenness centrality measure utilized in our study.

Third, we evaluate the superiority of our hurdle model over the Poisson, COM-Poisson, Negative Binomial, and Zero-inflated models in accounting for overdispersion and the excess of zeros present in our citation count data. Detailed results are presented in Appendix C.4.1. Initially, a dispersion test, as reported in Table C1, confirms the rejection of the equidispersion hypothesis pertaining to our citation count variables. The efficacy of the hurdle model in capturing the nuances of citation counts is further compared using log-likelihood, BIC (with Raftery's probability), and Vuong's non-nested hypothesis test (Vuong, 1989). These comparisons are documented in Tables C2 through C4 and visualized via rootogram plots (Kleiber and Zeileis, 2016) in Figures C5 to C7. Across all metrics, our analyses consistently affirm the superiority and robustness of the hurdle model in capturing the influence of connectivity on the success of ideas in econometrics.

Lastly, to examine model specification, we performed various estimations with different configurations: (i) using only connectivity variables; (ii) using only control variables; and (iii) a full model incorporating both control and connectivity variables. Table C5 in Appendix C.4.2 presents the BIC and Raftery probability, comparing (i) and (ii) with (iii), which confirms the superiority of our comprehensive approach. Furthermore, to address potential uncertainties arising from the calculation of ideas and social connections, we conducted a bootstrap analysis. The results, detailed in Table C6 in Appendix C.4.2, reinforce the paper's principal findings.

6. Conclusion

This paper aims to explore the factors influencing the dissemination of ideas in econometrics, a discipline uniquely positioned at the intersection of economics, finance, and statistics. Although the field's interdisciplinary nature suggests a high potential for rapid theory diffusion within the scientific community, it remains relatively underexplored, presenting a promising avenue for innovative discoveries.

Drawing from a comprehensive dataset of more than 17,000 research articles collected over four decades, our empirical analysis employs a sophisticated blend of natural language processing and network analysis techniques. This approach serves to illuminate the critical role that connectivity plays in the dissemination and recognition of ideas in econometrics. While the intrinsic quality of research undeniably remains a fundamental determinant of its impact, we demonstrate that it is intricately interlaced with additional factors—specifically, thematic and social connectivities. Thematic connectivity enhances a paper's scholarly relevance by bridging multiple academic domains, thereby increasing its citation potential. Similarly, social connectivity – defined by the strength and breadth of an author's or team's academic network – amplifies a paper's visibility and credibility within the scientific community. Furthermore, our findings indicate that as both idea and social connectivity increase, the success of idea diffusion initially rises but eventually begins to decline after reaching a certain threshold. The robustness of our findings is empirically grounded using several models including the hurdle negative binomial, Poisson, COM-Poisson, Zero-inflated, and negative binomial, along with multiple network measures.

A key takeaway from our research is that, all else being equal, a publication aiming to be impactful should bridge various knowledge domains and be produced by scholars who are well-recognized and connected within the scientific community. Overall, our findings provide actionable insights for scholars navigating the complex world of academic publishing, while also establishing a foundational framework for further investigation of this intricate ecosystem. This finding is particularly insightful for young scholars entering the domain and aiming at maximizing the impact of their research. It also could be used by research (public or private) agencies to select and fund projects, having high potential of dissemination. Finally, whereas our findings are in line with the literature showing that increased coauthorship enhances citation counts, we find that a threshold exists. This result is reassuring as it highlights that pure opportunistic behaviors consisting of increasing the number of coauthors only to gain citations face a limit that must not be crossed.

Our findings reinforce and extend the analysis of the Matthew effect in the econometrics domain, showing that beyond the accumulation of advantages for established scholars, the proliferation of ideas and academic influence are intricately tied to social and thematic connectivity. This underscores a more dynamic interplay between an author's historical success and the visibility and impact of their current research.

Our research not only sheds light on the multifaceted elements contributing to scholarly impact but also serves as a catalyst for future explorations into the dynamics of academic influence within econometrics and beyond. Given the ever-evolving landscape of scientific research, understanding the role of connectivity in the dissemination of ideas has never been more relevant.

CRediT authorship contribution statement

Bertrand Candelon: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Marc Joëts:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Valérie Mignon:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.respol.2024.105025.

References

Andrikopoulos, A., Samitas, A., Kostaris, K., 2016. Four decades of the Journal of Econometrics: Coauthorship patterns and networks. J. Econometrics 195 (1), 23–32.
Archontakis, F., Mosconi, R., 2021. Søren Johansen and Katarina Juselius: A bibliometric analysis of citations through multivariate bass models. Econometrics 9 (3), 30.
Azoulay, P., Graff Zivin, J.S., Wang, J., 2010. Superstar extinction. Q. J. Econ. 125 (2), 549–589.
Biscaro, C., Giupponi, C., 2014. Co-authorship and bibliographic coupling network effects on citations. PLoS One 9 (6).
Blei, D., Ng, A., Jordan, M., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (Jan), 993–1022.
Bloom, N., Jones, C.I., Van Reenen, J., Webb, M., 2020. Are ideas getting harder to find? Amer. Econ. Rev. 110 (4), 1104–1144.
Borgatti, S.P., Everett, M.G., 1997. Network analysis of 2-mode data. Social Networks 19 (3), 243–269.
Brandes, U., 2001. A faster algorithm for betweenness centrality. J. Math. Sociol. 25 (2), 163–177.
Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. Social Networks 30 (2), 136–145.
Brunetti, C., Joëts, M., Mignon, V., 2024. Reasons Behind Words: OPEC Narratives and the Oil Market. Working Papers 2024-03, Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System.
Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press.
Cameron, A.C., Trivedi, P.K., 2013. Regression Analysis of Count Data, Vol. 53, Cambridge University Press.
Chang, C.-L., McAleer, M., 2013. Ranking leading econometrics journals using citations data from ISI and RePEc. Econometrics 1 (3), 217–235.
Cronin, B., Shaw, D., La Barre, K., 2003. A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. J. Am. Soc. Inf. Sci. Technol. 54 (9), 855–871.
Cronin, B., Shaw, D., La Barre, K., 2004. Visible, less visible, and invisible work: Patterns of collaboration in 20th century chemistry. J. Assoc. Inf. Sci. Technol. 55, 160–168.
Deichmann, D., Jensen, M., 2018. I can do that alone or not? How idea generators juggle between the pros and cons of teamwork. Strateg. Manag. J. 39 (2), 458–475.

Deichmann, D., Moser, C., Birkholz, J.M., Nerghes, A., Groenewegen, P., Wang, S., 2020. Ideas with impact: How connectivity shapes idea diffusion. Res. Policy 49 (1), 103881.
Didegah, F., Thelwall, M., 2013. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. J. Informetr. 7 (4), 861–873.
Ding, J., Shen, Z., Ahlgren, P., Jeppsson, T., Minguillo, D., Lyhagen, J., 2021. The link between ethnic diversity and scientific impact: the mediating effect of novelty and audience diversity. Scientometrics 126 (9), 7759–7810.
Ductor, L., 2015. Does co-authorship lead to higher academic productivity? Oxf. Bull. Econ. Stat. 77 (3), 385–407.
Ductor, L., Fafchamps, M., Goyal, S., Van der Leij, M.J., 2014. Social networks and research output. Rev. Econ. Stat. 96 (5), 936–948.
Everett, M., Borgatti, S.P., 2005. Ego network betweenness. Social Networks 27 (1), 31–38.
Fanelli, D., Lariviere, V., 2016. Researchers' individual publication rate has not increased in a century. Plos One 11, e0149504.
Geisberger, R., Sanders, P., Schultes, D., 2008. Better approximation of betweenness centrality. In: 2008 Proceedings of the Tenth Workshop on Algorithm Engineering and Experiments. ALENEX, SIAM, pp. 90–100.
Gibbons, M. (Ed.), 1994. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. SAGE Publications, Thousand Oaks, Calif..
Glänzel, W., 2001. National characteristics in international scientific co-authorship relations. Scientometrics 51, 69–115.
Glänzel, W., Schubert, A., 2001. Double effort=Double impact? A critical view at international co-authorship in chemistry. Scientometrics 50, 199–214.
Glänzel, W., Schubert, A., Czerwon, H.-J., 1999. A Bibliometric Analysis of International Scientific Cooperation of the European Union (1985–1995). Vol. 45, Katholieke Universiteit Leuven, Open Access publications from Katholieke Universiteit Leuven.
Goyal, S., Van Der Leij, M.J., Moraga-González, J.L., 2006. Economics: An emerging small world. J. Polit. Econ. 114 (2), 403–412.
Hall, K., Vogel, A., Huang, G., Serrano, K., Rice, E., Tsakraklides, S., Fiore, S., 2018. The science of team science: A review of the empirical evidence and research gaps on collaboration in science. Am. Psychol. 73, 532–548.
Hansen, S., McMahon, M., 2016. Shocking language: Understanding the macroeconomic effects of central bank communication. J. Int. Econ. 99, S114–S133.
Hansen, S., McMahon, M., Prat, A., 2018. Transparency and deliberation within the FOMC: a computational linguistics approach. Q. J. Econ. 133 (2), 801–870.
Hara, N., Solomon, P., Sonnenwald, D., Kim, S.-l., 2003. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. J. Am. Soc. Inf. Sci. Technol. 54.
Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. 102 (46), 16569–16572.
Hsieh, C.-S., Konig, M.D., Liu, X., Zimmermann, C., 2018. Superstar Economists: Coauthorship networks and research output.
Ji, P., Jin, J., Ke, Z.T., Li, W., 2022. Co-citation and Co-authorship networks of statisticians. J. Bus. Econom. Statist. 40 (2), 469–485.
Jones, B.F., 2009. The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? Rev. Econ. Stud. 76 (1), 283–317.
Jones, B.F., 2021. The rise of research teams: Benefits and costs in economics. J. Econ. Perspect. 35 (2), 191–216.
Kaplan, S., Vakili, K., 2015. The double-edged sword of recombination in breakthrough innovation. Strateg. Manag. J. 36 (10), 1435–1457.
Khor, K., Yu, L., 2016. Influence of international co-authorship on the research citation impact of young universities. Scientometrics 107.
Kleiber, C., Zeileis, A., 2016. Visualizing count data regressions using rootograms. Amer. Statist. 70 (3), 296–303.
Köseoglu, M.A., 2016. Growth and structure of authorship and co-authorship network in the strategic management realm: Evidence from the Strategic Management Journal. Bus. Res. Q. 19.
Larsen, V.H., Thorsrud, L.A., 2019. The value of news for economic developments. J. Econometrics 210 (1), 203–218.
Leahey, E., Beckman, C.M., Stanko, T.L., 2017. Prominent but less productive: The impact of interdisciplinarity on scientists' research. Adm. Sci. Q. 62 (1), 105–139.
Magerman, T., Van Looy, B., Debackere, K., 2015. Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology. Res. Policy 44 (9), 1702–1713.
McFadyen, M.A., Cannella, Jr., A.A., 2004. Social capital and knowledge creation: Diminishing returns of the number and strength of exchange relationships. Acad. Manag. J. 47 (5), 735–746.
Merton, R.K., 1968. The Matthew effect in science. Science 159 (3810), 56–63.
Mimno, D., Wallach, H., Talley, E., Leenders, M., McCallum, A., 2011. Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. pp. 262–272.
Moed, H., 2005. Citation Analysis in Research Evaluation. Springer.
Mullahy, J., 1986. Specification and testing of some modified count data models. J. Econometrics 33 (3), 341–365.
Newman, M.E., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys. Rev. E 64 (1), 016132.
Nowell, C., Grijalva, T., 2011. Trends in co-authorship in economics since 1985. Appl. Econ. 43 (28), 4369–4375.
Opsahl, T., 2013. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Soc. Netw. 35 (2), 159–167.
Parish, A., Boyack, K., Ioannidis, J., 2018. Dynamics of co-authorship and productivity across different fields of scientific research. Plos ONE 13, e0189742.
Persson, O., Glänzel, W., Danell, R., 2004. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. Scientometrics 60 (3), 421–432.
Podolny, J.M., 2001. Networks as the pipes and prisms of the market. Am. J. Sociol. 107 (1), 33–60.
Price, D.J.D.S., 1963. Little Science, Big Science. Columbia University Press, New York Chichester, West Sussex.
Reagans, R., Zuckerman, E.W., 2001. Networks, diversity, and productivity: The social capital of corporate R&D teams. Organ. Sci. 12 (4), 502–517.
Roberts, M.E., Stewart, B.M., Tingley, D., 2019. Stm: An R package for structural topic models. J. Stat. Softw. 91 (1), 1–40.
Roberts, M.E., Stewart, B.M., Tingley, D., Airoldi, E.M., et al., 2013. The structural topic model and applied social science. In: Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation. Vol. 4, Harrahs and Harveys, Lake Tahoe, pp. 1–20.
Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., Rand, D.G., 2014. Structural topic models for open-ended survey responses. Am. J. Political Sci. 58 (4), 1064–1082.
Sarafoglou, N., Paelinck, J.H., 2008. On diffusion of ideas in the academic world: the case of spatial econometrics. Ann. Reg. Sci. 42 (2), 487–500.
Seierstad, C., Opsahl, T., 2011. For the few not the many? The effects of affirmative action on presence, prominence, and social capital of women directors in Norway. Scand. J. Manag. 27 (1), 44–54.
Shen, H., Xie, J., Li, J., Cheng, Y., 2021. The correlation between scientific collaboration and citation count at the paper level: a meta-analysis. Scientometrics 126 (4), 3443–3470.
Stock, J.H., Watson, M.W., 2017. Twenty years of time series econometrics in ten pictures. J. Econ. Perspect. 31 (2), 59–86.
Taddy, M., 2012. On estimation and selection for topic models. In: Artificial Intelligence and Statistics. PMLR, pp. 1184–1193.
Thelwall, M., Kousha, K., Abdoli, M., Stuart, E., Makita, M., Wilson, P., Levitt, J., 2023. Why are coauthored academic articles more cited: Higher quality or larger audience? J. Assoc. Inf. Sci. Technol. 74 (7), 791–810.
Thelwall, M., Maflahi, N., 2022. Research Co-authorship 1900–2020: Continuous, universal, and ongoing expansion. Quant. Sci. Stud. 3, 1–15.
Trapido, D., 2015. How novelty in knowledge earns recognition: The role of consistent identities. Res. Policy 44 (8), 1488–1500.
Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical combinations and scientific impact. Science 342 (6157), 468–472.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.
Wagner, C.S., Whetsell, T.A., Mukherjee, S., 2019. International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. Res. Policy 48 (5), 1260–1270.
Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D., 2009. Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 1105–1112.
Wang, J., 2016. Knowledge creation in collaboration networks: Effects of tie configuration. Res. Policy 45 (1), 68–80.
Ziman, J., 2000. Real Science: What it Is and What it Means. Cambridge University Press.