Journal of Agricultural and Resource Economics: Preprint ISSN: 1068-5502 (Print); 2327-8285 (Online)

doi: 10.22004/ag.econ.370432

WAEA Presidential Address: JARE Topics, Author Networks, and Citations, 1992 to 2024

Dayton Lambert

This study examines 1,009 articles published in the Journal of Agricultural and Resource Economics (JARE) from 1992 to 2024. It identifies 24 research topics using Latent Dirichlet Allocation and explores how topics and author collaboration networks influence citations. Article citation data from Scopus show that consumer, farm resource, and crop insurance articles lead overall and yearly citation growth. Niche topics demonstrate high citation intensity. A count regression finds that topic choice and certain types of author collaboration significantly predict citation counts. The findings underscore JARE's scholarly influence and highlight its importance as a journal for disseminating impactful research in agricultural and resource economics.

Key words: meta-analysis, textual analysis, network analysis, citations

Introduction

The Journal of Agricultural and Resource Economics (JARE) has served as the flagship publication of the Western Agricultural Economics Association since 1975. What sets JARE apart from other agricultural economics journals is its commitment to publishing innovative and rigorous research related to agriculture, natural resources, and related fields, with a traditional emphasis on the Western United States and the Great Plains region. The Journal emphasizes research on current issues relating to food systems and disruptions, crop and livestock markets, farm resource and environmental challenges, and rural development. JARE fosters ongoing dialogue and advances research and practice in agricultural and resource economics by providing an open platform for research, policy analysis, and scholarly discussion. Over the years, JARE and the Association it represents have built strong connections between students and faculty, supporting the growth of both emerging and experienced scholars in the field.

In celebration of the Journal's 50th anniversary, I was interested in which topics appeared most frequently in JARE and which ones received more or fewer citations. Citation counts reflect research quality and scientific influence (Thelwall & Wilson, 2014; Aistleitner et al., 2019). However, factors other than scientific rigor, such as author or coauthor reputation, journal prestige, topic novelty, and author collaborations, also influence citation patterns (Bornmann & Daniel, 2008; Thelwall & Wilson, 2014; Kumar, 2015; Aistleitner et al., 2019). According to the Scopus Cite-Score database, the corpus of JARE articles from 1992 to 2024 was cited 14,990 times.

First, I performed a topic analysis of articles published in the Journal of Agricultural and Resource Economics (JARE) from 1992 to 2024 to complete this task. The analysis is based on a dataset of 1,009 articles written by over 800 unique contributors. Topics were identified using an unsupervised machine learning method called Latent Dirichlet Allocation (LDA, Blei et al. 2003), which was based on 16,290 unique words found in the article titles and abstracts.

Review coordinated by Vardges Hovhannisyan.

Second, I also wanted to find out if specific types of author collaborations or networks predict citations. I trained ChatGPT-4o (the "AI") to generate code that would connect the 800+ coauthors. The AI was also trained to develop network metrics that characterize author collaboration styles. These variables were used to predict article citation counts, in addition to the latent topics.

The main questions asked are 1) What are the most common research topics in JARE? 2) Which topics receive the most citations? 3) Which topics increase the likelihood of an article being cited? and 4) Do author networks predict citations? To answer these questions, I regressed citation counts on topics and author network variables while controlling for volume-issue fixed effects.

Text and Citation Data

JARE has been archived from 1992 to the present. All articles are archived at AgEcon Search[1]. The data includes author and coauthor names, publication year, volume and issue numbers, page ranges, titles, and abstracts. Amy Bekkerman, the JARE Publications Editor, provided the article data. The dataset for this analysis ends in 2024, covering 32 years. During this period, 1,009 articles were published, spanning 34 volumes and involving over 800 unique authors. Titles and abstracts were used for textual analysis. Combined, abstracts and titles totaled 16,290 words.

Scopus was used to gather JARE article citations. Institutions subscribing to Scopus can access citation data through an application programming interface key. Scopus is a trusted source for citation information due to its strict quality control, extensive disciplinary and international coverage, and large historical dataset (Thelwall & Sud, 2022). It has comprehensive quality assurance measures, including ongoing data cleaning, accurate citation linking, and a systematic review process to exclude underperforming or predatory journals (Baas et al., 2020). Although Google Scholar reports a higher total number of citations, Scopus offers a more reliable database because it minimizes duplicates, non-peer-reviewed journals, and non-academic sources (Gerasimov et al., 2024). One limitation of using Scopus for this analysis is that it has no JARE citation data prior to 1995. From 1995 to 2024, JARE articles were cited 14,990 times.

Methods & Procedures

This section summarizes the procedures used to generate topics and author networks for the citation regression. It first discusses the cleaning and processing steps involved in building a topic model and its calibration and estimation. Next, it explains the procedure used to train AI to create an author network model, followed by the count regression model employed to predict citations.

Topic modeling

JARE topics were identified in the collected articles using Latent Dirichlet Allocation (LDA). LDA is an unsupervised machine learning algorithm that uncovers underlying, latent themes in a document set (Blei et al., 2003). The calibrated LDA topic model assigns articles to topics with probabilities. Matlab's *fitlda* procedure was used to calibrate and estimate the topic model based on the text data in the abstracts and titles of the articles (The Mathworks Inc., 2023).

Text data must be processed and cleaned first before it can be used to train an LDA model. All titles and abstracts were converted to lowercase, and then punctuation and numbers were removed. Next, stopwords were eliminated. Stopwords include articles, conjunctions, adverbs, adjectives, and proper nouns. Sentences were tokenized, meaning they were split into words.

[1] https://ageconsearch.umn.edu/?ln=en

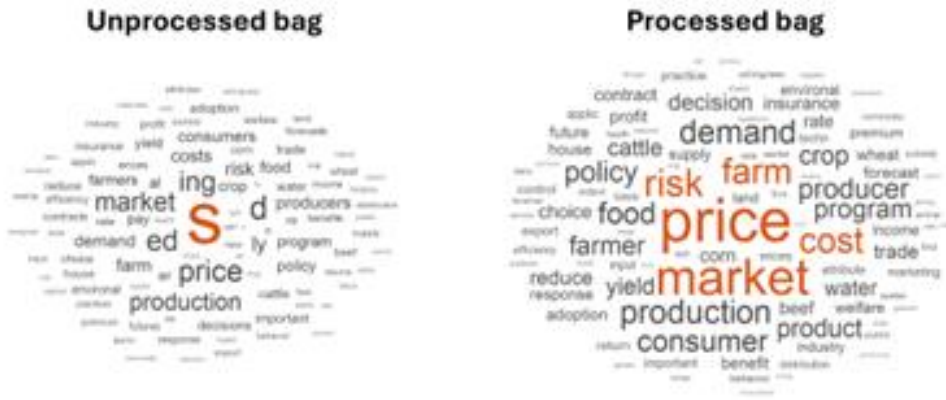

Figure 1. Unprocessed and processed word bag for JARE titles and abstracts, 1992 to 2024.

Words with three or fewer letters were removed, except for common acronyms in agricultural economics, such as CRP (Conservation Reserve Program), EDP (Expected Progeny Difference), and WTP (Willingness to Pay). Words were then lemmatized, reducing them to their base forms. For example, the root for "crops," "crop," and "cropping" is "crop." Next, words appearing less than twice were removed from the text sample. The final cleaned output results in a 'bag of words,' used in the topic modeling step (Figure 1). The bag contained 1,515 unique words.

The LDA categorizes articles into topics, but it cannot determine the best number of topics to use in an analysis. A grid search with cross-validation was used to identify this number. The LDA was fitted across a range of candidate topic counts from 5 to 26. For each fit, 10 percent of the documents were randomly omitted, and their validation perplexity scores were recorded. Lower perplexity scores indicate better model generalization to new, unseen documents (Blei et al., 2003). The number of topics that resulted in the LDA model with the lowest perplexity score was 24.

The LDA is a predictive model that varies in performance, performing poorly in some cases and well in others. It estimates the probability that an article belongs to one of the 24 topics. Naturally, for each article, the sum of these probabilities across all 24 topics equals one. For some articles, the probabilities across topics were nearly uniform, indicating difficulty in categorizing these articles into specific topics. For other articles, the probability of belonging to a particular topic was clearer, as shown by a dominant probability assigned to a single topic. A decision was made to assign each article to a single topic by choosing the one with the highest probability using dummy variables. The LDA only classifies articles into topics; it does not provide descriptions of what theme a topic represents. Topics were named by examining the word clouds for each and then subjectively assigning them thematic labels (Figure 2).

Author network

Articles are cited based on, among other things, factors like author reputation and the culture of academic communities (Bornmann & Daniel, 2008; Thelwall & Wilson, 2014; Aistleitner et al., 2019). Collaborative networks can act as proxies for how individual researchers work with a community of peers, as well as reflect the different styles of collaboration in the research community. In network theory, authors are represented as nodes, and connections to other authors are edges. These unseen relationships and connections across the research community presumably influence citation behavior.

Figure 2. The LDA analysis identified 24 topics.

Network theory uses specific vocabulary to describe network subgraphs, which consist of different node-edge configurations. The network variables used here to predict citations are 'MaxDegree,' 'MaxEigenvector,' 'MaxCloseness,' 'MeanBetweenness,' 'DiversityEntropy,' and 'Density' (Freeman, 1979; Wasserman & Faust, 1994). MaxDegree, which I call "hub author," identifies the most connected author in an article, meaning the author with the most links to other authors. MaxEigenvector, which I call "hub-hub," measures how well connected a hub author is to other hub authors. MaxCloseness, which I label "info. access," indicates an author's ability to access information within their network efficiently. MeanBetweenness shows that an author acts as a collaborative bridge connecting other authors. I refer to this network variable as "facilitator." DiversityEntropy, which I call "connected," tracks an author's coauthors from different previous collaborations. Lastly, Density measures how tightly connected an author's group is.

There are many open-source, free platforms available to conduct network analysis. R and Python support several of these platforms, while others are standalone packages. Instead of choosing one and learning how to use it, I trained ChatGPT-4o to build an author network algorithm. I began with the following prompt.

A dataset contains journal articles. Author names are in the first column. A semicolon separates the authors. The author's names are Lastname, Firstname format. Build an authorship network. This network should display connections between an author and all co-authors. Example: Entry 1: Lambert, Dayton; Boyer, Chris. Entry 2: Boyer, Chris; Brorsen, Wade; Entry 3: Jensen, Kim; Lambert, Dayton. Lambert is connected to Boyer and Jensen. Boyer is connected with Lambert and Brorsen. Brorsen is connected with Boyer. Jensen is connected to Lambert. Display the network as a figure. Display the author's last name on the figure. The font of authors with more connections should be larger than that of authors with fewer connections.

The AI training was completed after three additional prompts that introduced new connections and different authors. Because network analysis of all JARE authors was impractical on the AI platform, the underlying Python code was exported and run on a desktop computer. (NetworkX was the package AI used.) Next, the JARE author network was built using all 800+ authors. AI was used to develop the network variables, which were then added to the JARE article dataset. The author network is shown in Figure 3.

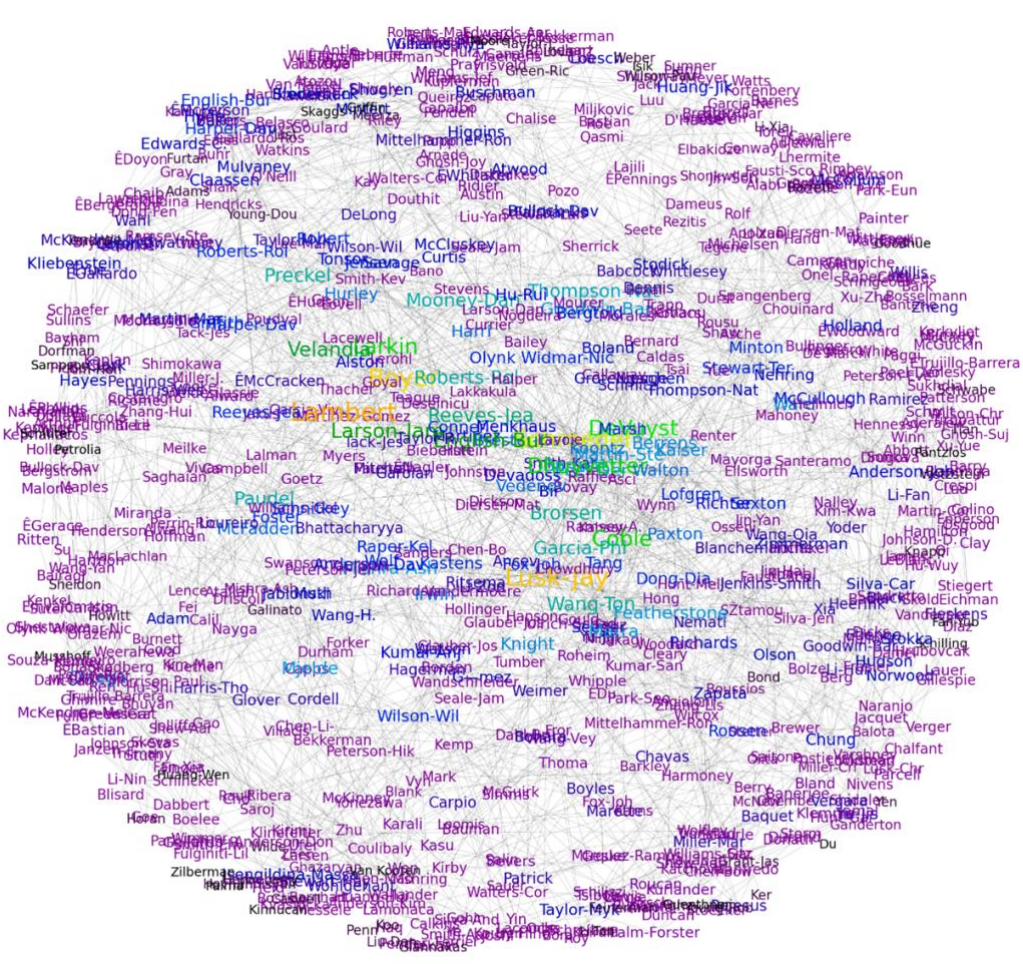

Figure 3. JARE author network, 1992 to 2024.

Citation regression

Citation counts were regressed on the topic dummy variables, the author network variables, volume-issue fixed effects, the article page count, and the author count. I had no prior expectations concerning the relationship between citations and these variables. The summary statistics are in Table 1. On average, journal articles are 15 pages long and have three authors. The author network variables were standardized to a mean of zero and a variance of one. Negative binomial regression was used to estimate the citation model.

Results and Discussion

About 30 percent of the articles published in JARE from 1992 to 2024 focus on crop insurance, consumers, marketing, farm resources, and food economics. These are the top five themes among the 24 identified topics. The five topics with the fewest articles during this period are contracts, producers, livestock production, technology adoption, and industrial organization, which together account for 11 percent of the 1,009 articles.

Table 1. Summary statistics (1,009 articles).

Variable	Mean	Std. dev.	Min	Max
Citations	14.86	26.93	0	324
Hub author	0.00	1	-1.00	3.64
Density	0.00	1	-2.41	0.41
Facilitator	0.00	1	-0.56	7.40
Connection	0.00	1	-0.71	4.84
Info. Access	0.00	1	-1.12	1.58
Hub-hub author	0.00	1	-0.21	7.43
Page count	15.88	4.36	0	38
Author count	2.59	1.17	1	10
Topics				
Consumer	0.06		0	1
Contracts	0.03		0	1
Cropinsurance	0.06		0	1
Decmaking	0.04		0	1
Demandanalysis	0.04		0	1
Envcostben	0.06		0	1
Farm	0.04		0	1
Farmresource	0.06		0	1
Food	0.06		0	1
Grainsforecast	0.03		0	1
Hedonprice	0.03		0	1
Houseecon	0.03		0	1
Industorg	0.02		0	1
Ioprodefficiency	0.04		0	1
Marketing	0.06		0	1
Policy	0.03		0	1
Priceanalysis	0.05		0	1
Producers	0.02		0	1
Livestockprod	0.02		0	1
Programs	0.05		0	1
Publicwelfare	0.03		0	1
Risk	0.05		0	1
Techadoption	0.02		0	1
Trade	0.05		0	1

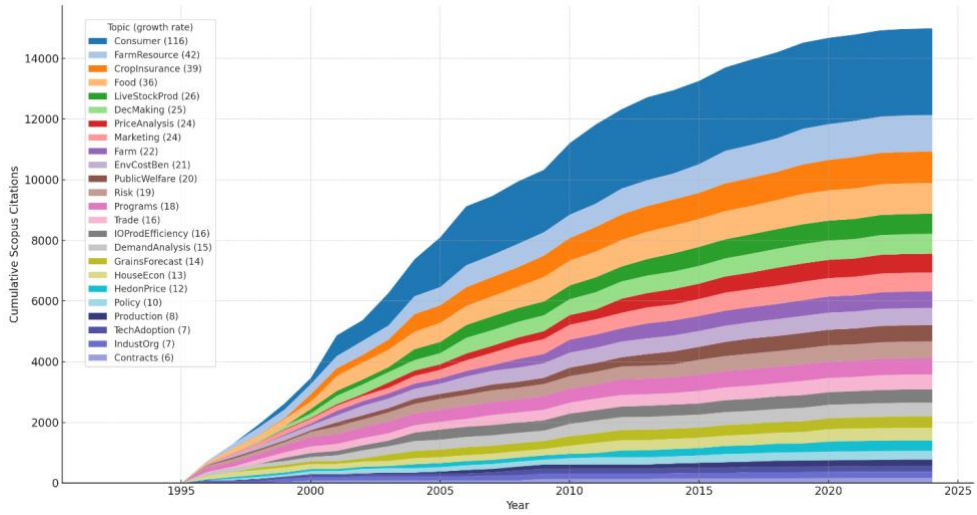

Figure 4. Topic citation evolution, 1992 to 2024.
Notes: numbers next to the topics are the annual growth rate in cumulative citations, and based on the regression, $cumul.cites_t = a + b \times year + u_t$, with b the annual increase in citations.

A different perspective appears when examining the most frequently cited topics (Figure 4). The top five cited topics are consumers (2,851, 19 percent), farm resources (1,207, 8 percent), crop insurance (1,034, 7 percent), food economics (1,015, 7 percent), and livestock production (661, 5 percent). Figure 4 also shows the cumulative citation rate over time. Consumer articles are generally the most cited topic. Its steady rise of 116 citations per year indicates that research into consumer preferences and behavior remains important in JARE and continues to grow. Similarly, farm resource and crop insurance articles are gaining prominence, possibly due to increased interest in uncertainty and risk in agriculture. The moderate growth in citations for articles published in contracts and industrial organization suggests these areas are either a maturing body of research or that the research community's focus has shifted to newer issues.

Prolific or older topics may overstate total citations. Another way to interpret these data is to examine citation intensity. Citation intensity highlights the impact per article rather than total volume. The top ten topics with the highest average citations per article are shown in Figure 5. Consumers, livestock production, and farm resources had the highest citation intensity. These three topics demonstrate a disproportionately high impact per article, likely due to their broad relevance and policy significance. Topics with relatively few articles but many citations, such as grains forecasting and household economics, seem to be selective, high-impact niche topics based on their citation intensity.

The predicted citations by topic are shown in Figure 6. Consumer and livestock production topics dominate in predicted citations. Articles covering consumer economics received the most citations, holding other factors constant. The consumer topic is the reference group. An article on consumer preferences or behavior will, on average, be cited 41 times over the period studied. Predicted citations for livestock production articles were not statistically different from consumer articles (28). Predicted citations were statistically lower than those of consumer and livestock production articles for all other topics. The gap is notable. The next highest topic, public welfare, averages nearly 48% fewer citations than consumer articles. The growth rate estimates indicate that topics like farm resources, food economics, and crop insurance have been gaining citations faster than livestock production or public welfare over time, suggesting a gradual narrowing of the citation gap. This pattern suggests that although established topics have more immediate visibility, emerging research areas may present opportunities for scholarly influence.

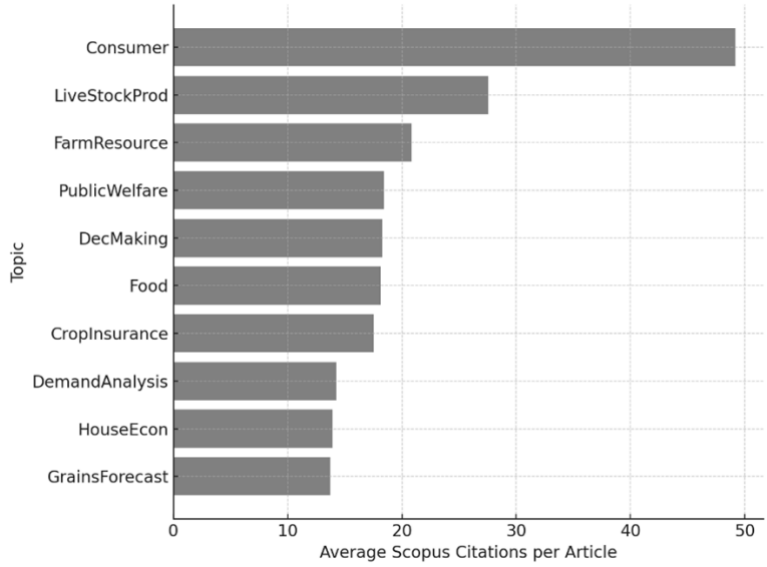

Figure 5. Average number of citations by topic and article.

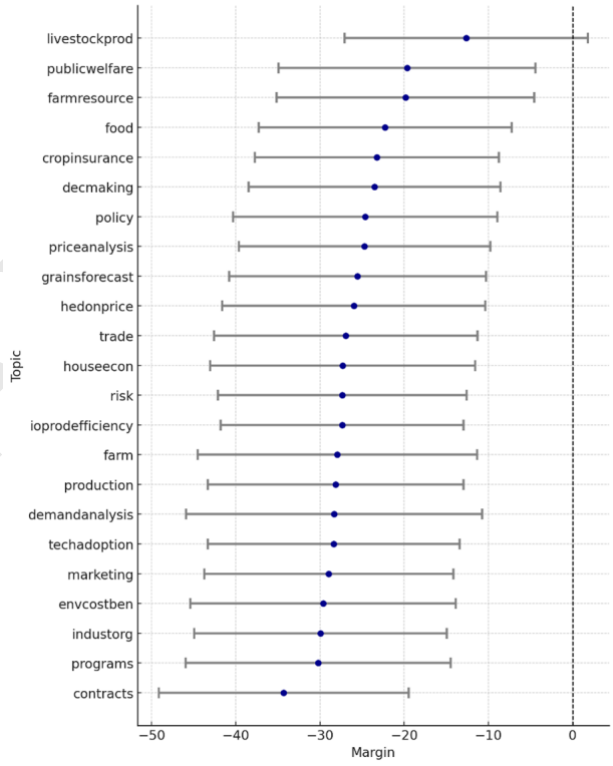

Figure 6. Predicted citations by topic, and marginal effects. Bars are ± 2 standard deviations. *Consumer* is the reference category, with 41 predicted citations.

Table 2. Negative binomial marginal effects (n = 893).

Variable	Margin	z-stat	p-value
Hub author (1)	2.53	2.01	0.04
Hub-hub author	-1.04	-1.28	0.20
Density	1.25	1.45	0.15
Facilitator	-1.38	-1.16	0.25
Connected	1.30	1.74	0.08
Info. Access	-1.04	-1.07	0.29
Page count	0.35	1.93	0.05
Author count	0.57	0.89	0.38

Notes: 1) Network variables are standardized with a mean of '0' and a variance of '1'.

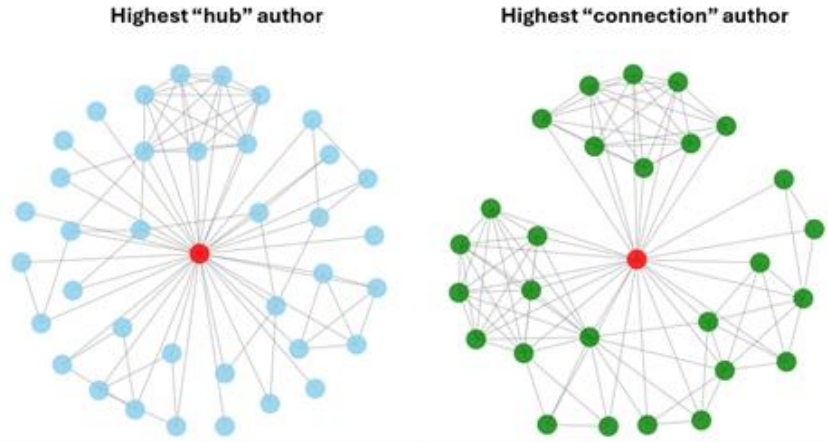

Figure 7. Sub-networks of the highest hub author index and the highest connectivity index.

The marginal effects of the author network variables are in Table 2. Two network variables are significant predictors of article citation counts. The first, hub author, is positively related to the number of citations. Articles with hub authors who collaborate broadly but with less-connected coauthors tend to receive more citations. The second significant network variable is connectivity. Articles authored by teams led by a few highly connected authors tend to attract more citations. Figure 7 illustrates these ideas. It shows the sub-networks of the author with the highest hub author count and the most interconnected authors. The differences are clear when comparing the collaborative styles of hub authors and those working with diverse, experienced, tightly knit teams.

Conclusions

The 50-year history of the Journal of Agricultural and Resource Economics highlights both the breadth and depth of research in agricultural and resource economics, especially related to the Western United States and the Great Plains region. As the leading publication of the Western Agricultural Economics Association, JARE publishes rigorous research and promotes professional connections that shape the careers of early-career scholars and engenders enduring relationships between students and mentors. In doing so, the Journal continues to advance the science, policy, and practice of agricultural and resource economics.

This meta-analysis shows that while some topics, such as consumer economics and livestock production, have historically attracted the most citations, many other fields like farm resources, food economics, and crop insurance also hold opportunities for scholarly influence. Even niche subjects can boost scholarship visibility when they focus on current and relevant issues. The Journal's impact is increased by its diverse group of authors, with collaborations linked to higher citation counts for articles. For the early-career faculty, these findings emphasize JARE's importance as a forum where both established and emerging topics can gain recognition and long-term scholarly acknowledgment.

[First submitted August 2025; accepted for publication September 2025.]

References

Aistleitner, M., J. Kapeller, and S. Steinerberger. 2019. "Citation patterns in economics and beyond." Science in Context 32(4): 361-380. doi: 10.1017/S0269889720000022.

Baas, J., Michiel Schotten, Andrew Plume, Gregoire Cote, and Reza Karimi. 2020. "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies". Quantitative Science Studies 1(1): 377-386. doi: 10.1162/qss_a_00019.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3(Jan), 993–1022.

Bornmann, L., and Hans-Dieter Daniel. 2008. "What do citation counts measure? A review of studies on citing behavior." Journal of Documentation 64(1): 45-80. doi: 10.1108/00220410810844150.

Freeman, L. C. (1979). Centrality in social networks: conceptual clarification. Social Networks, 1(3), 215–239.

Gerasimov, I., Binita KC, Armin Mehrabian, James Acker, and Michael P. McGuire. 2024. "Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Corssref, and DataCite". Scientometrics 129: 3681-3704. doi: 10.1007/s11192-024-05073-5.

Kumar, Sameer. 2015. "Co-authorship networks: a review of the literature." Aslib Journal of Information Management 67(1): 55-73. doi: 10.1108/AJIM-09-2014-0116.

Thelwall, M., and P. Sud. 2022. "Scopus 1900-2020: Growth in articles, abstracts, countries, fields, and journals". Quantitative Science Studies 3(2): 37-50. doi: 10.1162/qss_a_00177.

Thelwall, M., and P. Wilson. 2014. "Regression for citation data: An evaluation of different methods." Journal of Informetrics 8(4): 963-971. doi: 10.1016/j.joi.2014.09.011.

The MathWorks Inc. (2023). MATLAB version: 23.2.0.2428915 (R2023b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.

Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.