



OPEN ACCESS

EDITED BY

Jesús Marolla-Gajardo,
Universidad Bernardo O'Higgins, Chile

REVIEWED BY

Ariadne de Andrade Costa,
Universidade Federal de Jataí, Brazil
Ufuk Turen,
OSTİM Technical University, Türkiye
Hamide Özyürek,
OSTİM Technical University, Türkiye

*CORRESPONDENCE

Turgut Karaköse
✉ turgut.karakose@dpu.edu.tr

RECEIVED 10 May 2026

REVISED 31 July 2026

ACCEPTED 10 August 2026

PUBLISHED 27 August 2026

CITATION

Karaköse T, Kaptan O, Kılıç GN, Yirci R
and Kocabaş I (2026) The network
structure of research on artificial
intelligence in school leadership and
management: insights from a multi-
method approach.
Front. Educ. 11:1877606.
doi: 10.3389/feduc.2026.1877606

COPYRIGHT

© 2026 Karaköse, Kaptan, Kılıç, Yirci and
Kocabaş. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these
terms.

The network structure of research on artificial intelligence in school leadership and management: insights from a multi-method approach

Turgut Karaköse^{1*}, Orhun Kaptan², Gülenay Nagihan Kılıç³,
Ramazan Yirci⁴ and Ibrahim Kocabaş⁵

¹Faculty of Education, Kütahya Dumlupınar University, Kütahya, Türkiye, ²Bakırköy Science and Art Center, MoNE, Istanbul, Türkiye, ³Faculty of Education, Fatih Sultan Mehmet Vakıf University, Istanbul, Türkiye, ⁴Faculty of Education, Sütçüimam University, Kahramanmaraş, Türkiye, ⁵Faculty of Education, Yıldız Technical University, Istanbul, Türkiye

Even though Artificial Intelligence has become a rapidly growing research focus in the field of school leadership and management, it has not been statistically examined how the knowledge production structure and citation networks of this exponentially developing research field are constituted. Previous research employing techniques such as thematic analysis, co-word, or descriptive mapping has yielded themes and trends; however, limited attention has been paid to thematic and regional clustering of citation ties and to the cumulative advantage of citation counts for the further visibility of scientific works. Using a systematic-search-based scientometric and citation-network design, this study examines the citation network structure of a constructed corpus of 105 peer-reviewed studies on AI in school leadership and management through Exponential Random Graph Models (ERGMs), power-law analysis, and negative binomial regression. The analysis revealed three key findings: First, the observed citation network is sparse and centralized, with citations concentrated around a limited number of high-impact core studies. Second, intermediary studies that connect central and peripheral parts of the network are limited, suggesting weak brokerage and only emerging thematic sub-clustering within the corpus. Third, citation accumulation is highly uneven, with citations concentrated around a small number of highly visible studies, indicating a heavy-tailed and visibility-linked citation structure rather than a confirmed temporal preferential-attachment mechanism. Negative binomial regression showed that broader Google Scholar citation recognition was positively associated with within-corpus citation counts in the primary model, whereas the remaining coefficients were sensitive to the exclusion of this variable. This study contributes to discussions on the epistemic structure of AI research in school leadership and management by providing corpus-based evidence that the observed citation network is sparse, centralized, and uneven.

KEYWORDS

artificial intelligence, citation networks, educational management, exponential random graph models, preferential attachment, school leadership

1 Introduction

With the dawn of the Fourth Industrial Revolution, the emergence of artificial intelligence in the educational ecosystem has triggered a technological tsunami that challenges conventional paradigms of school leadership and management (Fullan et al., 2024). Thanks to this paradigmatic shift, school leaders are no longer assumed to be figures carrying out administrative tasks but are considered strategic actors who are expected to manage data-driven decision-making processes with the help of intelligent partners that mimic human cognition (Karaköse and Tulubas, 2025; Wang, 2021b). Whereas this technological transformation provides opportunities to ease administrative burdens, it also poses risks, including ethical responsibilities, algorithmic bias, and data privacy concerns (Igbokwe, 2024; Wang, 2021a). While debates about whether AI will render school principals obsolete are ongoing (Adams and Thompson, 2025; Karaköse, 2024), it is necessary to understand how the body of knowledge regarding AI in school leadership and management is structurally formed to see the direction in which the topic of AI will evolve in the field.

The literature on AI in school leadership and management demonstrates that the impact of AI in the field first appeared in 2018 and increased sharply in subsequent years (Arar et al., 2025; Richardson et al., 2025). Recent research indicates that this development is primarily concentrated in regions such as the US and China, and the most commonly inquired elements are ethics, performance, and human-machine symbiosis (Arar et al., 2024; Polat et al., 2025). While some studies suggest that school leaders' mindsets, self-efficacy, and readiness levels play a critical role in technology adoption (Anysiadou and Gkiliati, 2025; Kılınç et al., 2025), others put forward that the value middle managers attribute to AI is shaped by organizational support (Berkovich and Eyal, 2025). In addition, the role of generative artificial intelligence (GenAI) in instructional leadership, along with the comprehensive taxonomies developed to integrate educational institutions into this process, strengthens the field's theoretical foundation (Berkovich, 2025; Bixler and Ceballos, 2025; Sposato, 2025). On the other hand, it is observed that a significant portion of the existing literature is based on conceptual discussions and descriptive bibliometric analyses, and the ways in which school leaders use AI tools in their daily practices still remain at the experimental level (Berkovich, 2025; Ghamrawi et al., 2024; Kafa, 2025a, 2025b).

Although research on AI in school leadership and management has yielded important findings, these results have not yet firmly established the field's intellectual foundation. Given the lack of clear consensus on the advantages and challenges of AI use (Arar et al., 2025), the integration of AI with management and leadership (Kafa, 2025a), the ethical risks of AI use (Aldighrir, 2024), and the key themes associated with AI, this research aims to reveal the direction of academic interaction in the field of school leadership and management and the dynamics that control the intellectual infrastructure of the field.

Descriptive thematic mapping and bibliometric summaries can identify which topics, countries, journals, or authors are visible in the literature, but they provide limited evidence about the structural associations underlying observed citation patterns

or whether visible clusters correspond to statistically discernible knowledge communities. This distinction is important for AI research in school leadership and management because the field is still emerging and its intellectual boundaries are not yet fully stabilized. Network modeling can help determine whether observed citation ties are associated with thematic similarity, regional proximity, journal location, research design, and degree concentration. Accordingly, ERGM-based evidence is used here to examine whether the corpus displays attribute-based clustering or a centralized citation structure.

Unlike previous descriptive bibliometric analyses and literature reviews, this study statistically examines citation-network structure, citation-tie patterns, and article-level citation accumulation. It is therefore designed as a systematic-search-based scientometric and citation-network analysis rather than as a conventional systematic review. The PRISMA-guided search and screening procedure was used to construct a transparent corpus, while the subsequent analyses focused on citation relationships and citation concentration within that corpus.

In line with the exploratory purpose of the study, the following research questions were formulated:

Research Question 1: What are the structural characteristics of the citation network in research on Artificial Intelligence in school leadership and management?

Research Question 2: To what extent are observed citation ties associated with similarity in country, geographic region, journal of publication, research design, and dominant theme?

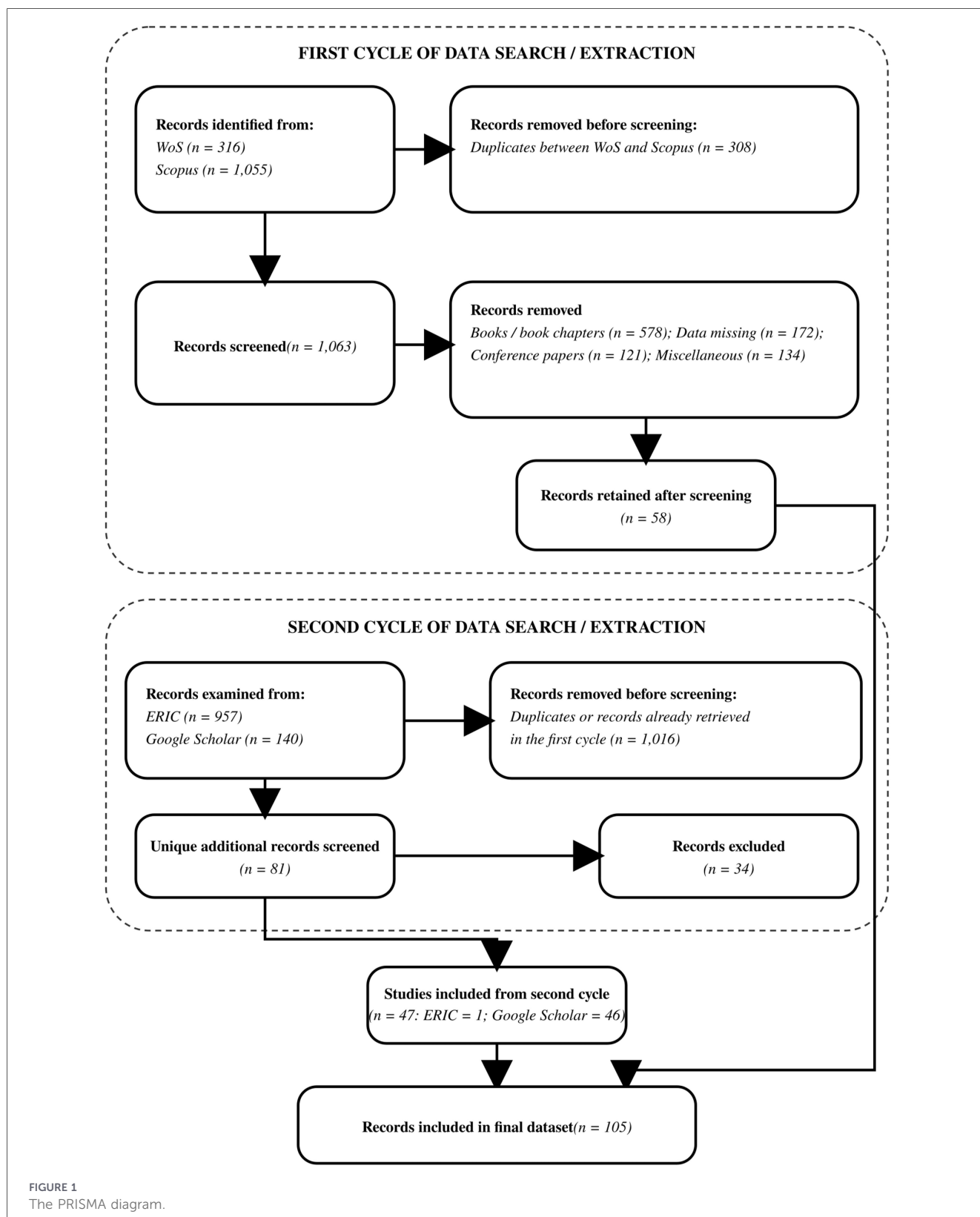
Research Question 3: To what extent does the observed within-corpus citation distribution exhibit heavy-tailed concentration consistent with, but not sufficient to establish, preferential attachment?

Research Question 4: Which article characteristics are associated with higher within-corpus citation counts?

2 Artificial intelligence in school leadership and management

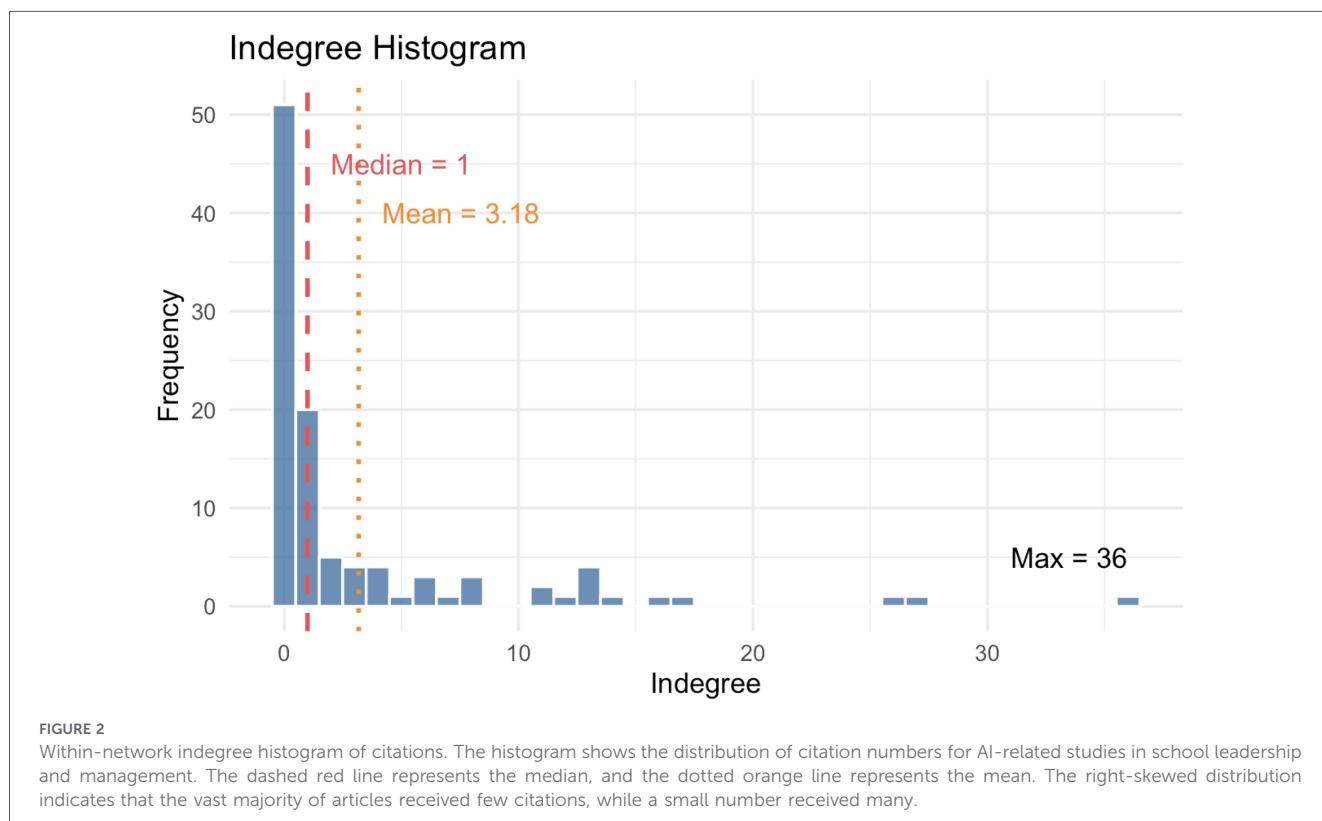
Discussions about artificial intelligence (AI) in school leadership and management build on the broader AI literature, whose origins date back to the 1950s, in the context of education (Arar et al., 2024; Igbokwe, 2024; Wang, 2021a). Although the first empirical evidence in the literature appeared in the works of Halff (1986) and Nelson (1989), the relationship between AI and school leadership has long been overshadowed by discussions of instructional technologies (Arar et al., 2024; Berkovich, 2025). However, advancements in machine learning, big data, and algorithmic analytics capabilities since 2017–2018 have propelled this field into a new phase of growth. In this regard, with the emergence of generative artificial intelligence, AI's peripheral position in school leadership and management discussions has shifted to a central role (Arar et al., 2025; Fullan et al., 2024; Karaköse and Tulubas, 2024; Richardson et al., 2025). This chronological evolution suggests that the field has begun to generate intense academic interaction and points to a tendency towards centralization around specific authoritative works.

This intellectual accumulation also indicates that knowledge production is clustered around certain driving themes. Notably,



data-driven decision-making (DIDM), human-machine symbiosis, and AI's potential for managerial automation form the fundamental pathways along which school leaders consider AI as an intelligent partner (Arar et al., 2024; Dai et al., 2025; Wang, 2021a, 2021b). Similarly, operational efficiency, the

automation of POSDCORB tasks, and mechanisms that reduce the workload of leaders and managers, thereby enabling greater focus on instructional leadership, are other emerging research lines (Adams and Thompson, 2025; Berkovich, 2025; Bixler and Ceballos, 2025; Kafa, 2025b; Karakose, 2024). On the other



hand, the rise of themes such as ethics, trust, transparency, and algorithmic bias (Igbokwe, 2024; Polat et al., 2025; Wang, 2021a) attests to the fact that the field is not only concentrated around the logic of technological optimization but also affected by normative and governance debates. This potential for thematic clustering provides a theoretical basis for homophilic links in the citation network structure of AI in school leadership and management research.

Likewise, school leaders' self-efficacy, readiness, and mindsets and attitudes towards AI (Anysiadou and Gkliati, 2025; Bellibaş et al., 2025; Berkovich and Eyal, 2025; Kılınç et al., 2025), and the role of teacher leadership in this process (Ghamrawi et al., 2024; Kesim, 2025) can be considered as emerging clusters in the literature. Additionally, the impact of AI on school culture and organizational structures points to a shift towards more distributed and flexible leadership models (Kesim et al., 2025; Sposato, 2025). This shift also paves the way for an expectation of clustering tendencies based not only on thematic similarities but also on homophilies, depending on the research design and theoretical approach.

The regional development pattern of AI in school leadership and management also supports the aforementioned structural dynamics. Previous research demonstrates that knowledge production in the field was initially concentrated in the United States and China, and followed by Australia, Russia, and India (Arar et al., 2024, 2025; Marrone et al., 2025; Polat et al., 2025). Although this structure initially led to knowledge homophily around specific epistemic centers (Arar et al., 2024), in recent years it has become more diversified with research from Turkey (Bellibaş et al., 2025; Kesim et al., 2025; Kılınç et al., 2025), Greece (Anysiadou and Gkliati, 2025), Cyprus (Kafa, 2025a,

2025b), Israel (Berkovich, 2025), Saudi Arabia (Aldighrir, 2024) and the Global South (Abdelazim et al., 2025; Berkat et al., 2024; Gârdan et al., 2025; Göçen and Döğler, 2025; Tyson and Sauers, 2021). This regional concentration and dissemination pattern provides a basis for examining attribute-based clustering and citation concentration around particular central studies within the observed network.

From a citation-network perspective, these thematic and regional concentrations may also be reflected in the patterning of observed citation ties. Studies addressing similar themes may cite overlapping sources because they share conceptual vocabularies, problem definitions, and reference traditions, providing a basis for examining thematic homophily. Country- or region-based concentrations may likewise be associated with citation proximity through shared policy contexts, academic communities, and publication circuits. In addition, studies published in visible outlets or positioned as early reference points may occupy more central positions because later studies use them as epistemic anchors. These expectations provide a theoretical basis for examining whether observed citation patterns are associated with homophily, degree concentration, and publication visibility. They do not, however, establish temporal citation-formation or preferential-attachment mechanisms.

In addition, it is suggested that the current knowledge framework is significantly formed by conceptual debates, bibliometric mappings, and early empirical studies (Berkovich, 2025; Ghamrawi et al., 2025; Kafa, 2025a, 2025b). This situation necessitates testing the intellectual structure of the field not only in terms of content but also in terms of network dynamics. The concentration of citations around highly visible studies, possible attribute-based clustering, heavy-tailed citation inequality, and

disparities associated with publication visibility provide the theoretical basis for the research questions examined in this study. In this context, the present study goes beyond descriptive mapping by examining the intellectual architecture of the field through social network analysis, ERGM, power-law analysis, and negative binomial regression.

3 Methodology

3.1 Research design and methodological approach

This study was designed as a systematic-search-based scientometric and citation-network analysis examining citation relationships among studies on Artificial Intelligence in school leadership and management. Although the corpus was constructed through a PRISMA-guided systematic search process, the study does not aim to synthesize evidence in the conventional systematic-review sense. Instead, it examines citation-network structure, observed citation-tie patterns, citation concentration, and article-level associations with within-corpus citation counts. Traditional bibliometric approaches explain the structure of scientific knowledge using techniques such as co-citation analysis (Small, 1973), bibliographic coupling (Kessler, 1963), and science mapping (Cobo et al., 2011). Building on this tradition, the present study approaches the literature as a directed citation network. Within the framework of social network theory (Wasserman and Faust, 1994), academic citations are represented as directed links, and the included articles constitute the nodes of the network (Newman, 2001).

Exponential Random Graph Models (ERGMs) were used to examine structural dependencies and attribute-based associations in the observed citation network. ERGM represents the probability of the observed network as a function of specified network configurations and node or dyadic attributes (Hunter et al., 2008; Robins, 2011). In this study, degree dependencies and similarity in country, geographic region, journal of publication, dominant theme, and research design were examined. Because the analysis represents an aggregated static citation network and does not restrict the risk set by publication chronology, the ERGM coefficients are interpreted as structural associations within the observed network rather than as direct evidence of temporal citation-formation processes.

Each analytical procedure addressed a distinct research question. Descriptive network measures, component analysis, and k-core analysis were used to examine sparsity, centralization, connectivity, and core-like citation concentration in relation to RQ1. ERGM was used to examine degree dependencies and associations between observed citation ties and article similarity in relation to RQ2. Power-law analysis was used to assess whether the within-corpus indegree distribution exhibited heavy-tailed concentration consistent with, but not sufficient to establish, preferential attachment, thereby addressing RQ3. Finally, negative binomial regression was used to examine associations between article characteristics and within-corpus citation counts in relation to RQ4. Together, these methods provide complementary evidence concerning

structural concentration, attribute-based associations, citation inequality, and article-level citation recognition.

3.2 Research corpus

The research corpus includes not only empirical studies addressing Artificial Intelligence in school leadership and management, but also systematic literature reviews and literature synthesis studies.

The literature search was conducted between 16 and 23 April 2026 using Web of Science Core Collection, Scopus, ERIC, and the standard Google Scholar interface. Because the databases use different search fields and interfaces, the search strategy was adapted separately for each source. Web of Science and Scopus were searched using database-specific title-field codes. ERIC was searched through its default search field, whereas Google Scholar's advanced-search option was used to restrict the searches to article titles. The database-specific search procedures, limits, and retrieval counts are reported in Table 1.

Google Scholar was searched through its standard interface rather than through Publish or Perish or another external retrieval tool. The advanced-search option restricting search terms to article titles was applied, and no publication-year limit was imposed. The same AI-, leadership-, and management-related term combinations used in the ERIC search were entered through multiple searches. Since Google Scholar reports approximate and potentially overlapping result totals, the first 14 pages ranked by relevance were examined, yielding 140 records for comparison and screening. This procedure was used as a supplementary search intended to identify relevant studies not retrieved through Web of Science, Scopus, or ERIC, rather than as an exhaustive screening of all results estimated by Google Scholar.

The retrieved records were examined at the title, abstract, and full-text levels and evaluated according to predetermined inclusion and exclusion criteria. Studies were eligible if they directly addressed the relationship between artificial intelligence and school leadership or management, were published as peer-reviewed journal articles, and employed an empirical, systematic review, literature review, bibliometric, scientometric, or conceptual synthesis design. Studies focusing exclusively on classroom-level instructional technologies without a substantive leadership or management dimension were excluded, as were books, book chapters, conference materials, dissertations, reports, and other non-journal publications. Studies were eligible if they were published in English as peer-reviewed journal articles, directly addressed the relationship between artificial intelligence and school leadership or management, and employed an empirical, systematic review, literature review, bibliometric, scientometric, or conceptual synthesis design.

The search and screening process was conducted in two stages. In the first stage, Web of Science returned 316 records and Scopus returned 1,055 records. After 308 duplicate records were removed, 1,063 unique records remained for screening, of which 58 met the eligibility criteria. In the supplementary stage, ERIC returned 957 results and the first 14 Google Scholar results pages, corresponding to 140 records, were examined. Records

TABLE 1 Database-specific search strategy.

Database	Search strategy	Search field and limits	Records retrieved/ examined
Web of Science Core Collection	TI = ("Artificial Intelligence" OR "AI") AND TI = ("leader*" OR "leadership" OR "school leadership" OR "educational leadership" OR "principal*" OR "school principal*" OR "middle manager*" OR "head*" OR "headship*" OR "headteacher*" OR "head teacher*" OR "school administrator*" OR "school administration" OR "school manager*" OR "school management" OR "administration" OR "management" OR "administrator*" OR "manager*" OR "vice-principal*" OR "vice principal*" OR "assistant principal*" OR "deputy principal*" OR "deputy head*" OR "deputy head teacher" OR "deputy headmaster")	Title field; searched 16–23 April 2026	316
Scopus	TITLE("Artificial Intelligence" OR "AI") AND TITLE("leader*" OR "leadership" OR "school leadership" OR "educational leadership" OR "principal*" OR "school principal*" OR "middle manager*" OR "head*" OR "headship*" OR "headteacher*" OR "head teacher*" OR "school administrator*" OR "school administration" OR "school manager*" OR "school management" OR "administration" OR "management" OR "administrator*" OR "manager*" OR "vice-principal*" OR "vice principal*" OR "assistant principal*" OR "deputy principal*" OR "deputy head*" OR "deputy head teacher" OR "deputy headmaster")	Title field; searched 16–23 April 2026	1,055
ERIC	Multiple combinations of the AI terms ("artificial intelligence" and "AI") with the leadership and management terms listed in the note below	Default search field; "Peer reviewed only" filter; searched 16–23 April 2026	957
Google Scholar	Multiple combinations of the same AI, leadership, and management terms	Advanced search; terms restricted to article titles; no publication-year limit; first 14 results pages examined; searched 16–23 April 2026	140

The ERIC and Google Scholar searches combined "artificial intelligence" and "AI" with the following terms across multiple searches: "school leadership," "educational management," "school management," "school principal," "head teacher," "administration," and "educational administration." Google Scholar displayed approximately 27,800 results across these searches; because this figure is an approximate platform estimate rather than a stable retrieval count, screening was limited to the first 14 relevance-ranked results pages, corresponding to 140 records. No language restriction was applied at the database-search stage; however, only studies published in English were retained during eligibility screening.

considered potentially eligible were compared both across the supplementary sources and with the Web of Science and Scopus records, primarily using titles and DOIs, with authors, publication years, and journal information consulted where necessary. After duplicate checking and eligibility assessment, 47 additional studies were retained: one uniquely identified through ERIC and 46 uniquely identified through Google Scholar. Together, the two search stages produced the final corpus of 105 studies. The complete search and screening process is shown in [Figure 1](#).

The 105 × 105 directed citation matrix was constructed manually by examining the reference list of each article in the final corpus. For every article, the cited references were compared with the titles of the other articles included in the corpus. A directed tie was coded as 1 when an article cited another article within the corpus and as 0 when no such citation was present. Self-citations on the diagonal were coded as 0. Reference matching was based primarily on article titles. Cases involving incomplete references, minor differences in title wording, inconsistent author names, or discrepancies between online-first and issue publication dates were manually checked

against the relevant journal webpages. Where publication dates differed across sources, the dates reported on the journal webpages were used. After the initial matrix had been constructed, all co-authors independently checked the citation matches and matrix entries. Any discrepancies identified during this verification process were reviewed against the original reference lists and journal webpages and resolved through comparison of the source records.

The corpus size was determined by the number of studies meeting the predefined search and eligibility criteria rather than by an *a priori* sample-size target. Accordingly, the 105 articles constitute the complete bounded corpus identified for the explicitly defined intersection of AI and school leadership or management. For ERGM analysis, analytical adequacy is not determined by a universal minimum number of nodes alone but also by network density, model complexity, parameter identifiability, and diagnostic performance. The corpus was therefore considered sufficient for examining the specified structural tendencies, while the inferential constraints associated with its limited size are acknowledged below.

For the analyses, two separate Excel files were prepared. The first file stored the verified 105×105 directed citation adjacency matrix, with rows and columns corresponding to articles and cell values indicating the presence or absence of within-corpus citations. The second file contained article-level attributes, including publication year, journal, country, geographic region, indexation status, research design, dominant theme, and other variables used in the ERGM and negative binomial regression models. After importation into R, the adjacency matrix was used to construct a directed network in which articles constituted nodes and within-corpus citations constituted directed edges. The files were subsequently used for ERGM estimation, power-law analysis, and negative binomial regression. The final corpus consisted of 105 peer-reviewed journal articles. The coded dataset and citation matrix can be provided by the corresponding author upon reasonable request for verification or replication purposes.

Given the recent emergence of AI research in school leadership and management, the final corpus should be interpreted as a relatively small but field-specific body of literature rather than as an exhaustive representation of all AI-in-education research. The inclusion of review-type studies was necessary because reviews, syntheses, and bibliometric studies constitute an important part of knowledge production in this early-stage literature and often function as organizing reference points within citation networks.

3.3 Data analysis

Citation relationships were modeled using directional social network logic; studies were defined as nodes, and citation relationships as edges (Newman, 2001; Wasserman and Faust, 1994). Analyses were conducted at two levels: network level and node level. In the first stage, the structural characteristics of the network were examined. Network density was calculated using the following formula:

$$\text{Density} = \frac{L}{N(N-1)}$$

Here, L represents the number of observed links, and N represents the number of nodes (Wasserman and Faust, 1994). Geometrically weighted degree terms (gwideg and gwodeg) were used to model heterogeneity in citation distributions. These terms model the concentration in citation and citation behavior, respectively. Because the decay parameter was fixed *a priori* and alternative decay values were not examined, the results should be interpreted as conditional on the selected value of 0.5. Future research should compare alternative fixed-decay specifications to assess the sensitivity of the degree-related coefficients. Furthermore, measures of global transitivity and centralization were calculated (Borgatti et al., 2024). Degree, indegree, outdegree, betweenness, and eigenvector centrality were used to evaluate centralization, authority concentration, and core-like organization within the observed network (Brass and Borgatti, 2019; Freeman, 1978). In directed networks, indegree and

outdegree centrality were calculated as follows:

$$C_D^{\text{in}}(i) = \sum_j a_{ji}$$

$$C_D^{\text{out}}(i) = \sum_j a_{ij}$$

Betweenness centrality:

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Eigenvector centrality is defined as follows:

$$x_i = \frac{1}{\lambda} \sum_j a_{ij} x_j$$

Matrix representation:

$$Ax = \lambda x$$

In the second stage, ERGMs were estimated to examine structural dependencies and attribute-based associations in the observed citation network. The general form of ERGM is as follows:

$$P(Y = y) = \frac{1}{\kappa(\theta)} \exp \left(\sum_k \theta_k g_k(y) \right)$$

Here, $g_k(y)$ represents the network statistics, and θ_k represents the parameters of these statistics. Modeling was carried out step-by-step. The first model tests the general tendency towards connection using only the 'edges' term. The terms 'gwideg' and 'gwodeg' have been added to model degree dependencies in structural models. In subsequent models, nodematch terms were used for country, region, journal, theme, index, and research design variables to test for homophily effects (McPherson et al., 2001). Model fit was evaluated using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), with lower values interpreted as a better fit (Sakamoto et al., 1986).

Model fit was evaluated not only through parameter significance but also through Goodness-of-Fit (GoF) analyses (Hunter et al., 2008). In this context, the simulated networks generated by the model were compared with the observed network with respect to degree distributions, common neighborhood structures, and geodesic distances.

The estimation procedure differed according to model complexity. The density-only baseline model was estimated using maximum likelihood estimation. The more complex ERGM specifications, including geometrically weighted in-degree/out-degree terms and homophily terms, were estimated using maximum pseudolikelihood estimation (MPLE) with Godambe-adjusted covariance estimates. Therefore, the reported ERGM coefficients should not be interpreted as MCMC-MLE estimates, and MCMC-specific diagnostics such as trace plots, MCMC standard errors, and random seeds are not applicable to the final reported models. The geometrically weighted in-degree

and out-degree terms were specified with a fixed decay parameter of 0.5. This value was selected *a priori* as a moderate decay setting to capture degree heterogeneity while limiting instability in a relatively small and sparse citation network.

In the final stage, an article-level count model was estimated to examine associations between article characteristics and within-corpus citation counts. Because the citation counts exhibited overdispersion, the mean–variance equality assumption of Poisson regression,

$$\text{Var}(Y) = E(Y),$$

was examined and found to be unsuitable. Negative binomial regression was therefore used (Hilbe, 2011). In this model, the variance is defined as follows:

$$\text{Var}(Y) = \mu + \frac{\mu^2}{\theta},$$

where μ represents the expected number of within-corpus citations and θ is the dispersion parameter. The regression model was specified as follows:

$$\begin{aligned} \log[E(Y_i)] = & \beta_0 + \beta_1 \text{Year}_i + \sum_r \beta_r \text{Region}_{ir} \\ & + \sum_j \beta_j \text{IndexGroup}_{ij} + \sum_d \beta_d \text{Design}_{id} \\ & + \beta_k \text{LogScholarCitation}_i. \end{aligned}$$

Incidence rate ratios (IRRs) were calculated to facilitate coefficient interpretation:

$$\text{IRR} = e^{\beta}.$$

Publication year was mean-centered before its inclusion in the model to improve the interpretability of the intercept and avoid numerical instability associated with raw calendar years. It was also included to account for differences in the time available for articles to accumulate citations. When calculated using a common reference year, publication age is an exact linear transformation of publication year. Consequently, replacing publication year with publication age would not change model fit or statistical significance, although it would reverse the direction of the corresponding coefficient. Publication year was therefore retained and interpreted as a control for citation exposure.

Of the 105 articles, 51 (48.6%) had zero within-corpus indegree. A substantial proportion of zero observations does not, by itself, demonstrate zero inflation or necessitate a zero-inflated or hurdle specification. Negative binomial regression can accommodate zero observations in overdispersed count data, and the inclusion of a separate zero-generating component should be supported by the assumed data-generating process (Bhaskar et al., 2023). In the present corpus, no theoretically distinct structural-zero group could be identified because all included articles were potentially eligible to receive incoming citations. The negative binomial specification was therefore retained.

The negative binomial model was used as a supplementary article-level count model to examine associations between article characteristics and within-corpus citation counts. Google Scholar citation counts capture citations from the wider scholarly literature beyond the 105-article corpus and are therefore not equivalent to within-corpus indegree. Nevertheless, because both are citation-based measures and may partially overlap, their association was interpreted as correspondence between broader scholarly recognition and citation centrality within the constructed corpus rather than as a temporally prior or causal effect.

In addition, the heavy-tailed structure of the within-corpus indegree distribution was examined as a diagnostic assessment of citation concentration. The power-law model $P(k) \sim k^{-\alpha}$ was estimated using the maximum likelihood approach proposed by Clauset et al. (2009). Given the relatively small corpus and the limited number of observations in the estimated upper tail, this analysis was not treated as a direct test of preferential attachment or as conclusive evidence of a power-law-generating mechanism. The log-log plot was used only for visual inspection. Likelihood-ratio comparisons were also conducted against alternative heavy-tailed distributions, including lognormal, exponential, and stretched-exponential models.

The ERGM models were specified for complementary purposes. Model 1, including only the edges term, served as a baseline representation of network sparsity. Model 2 added geometrically weighted in-degree and out-degree terms to examine degree-based structural dependencies. Model 3 combined these structural terms with homophily terms and was used as the main substantive specification for examining whether attribute-based associations remained after accounting for degree concentration. Model 4 included only homophily terms and served as a diagnostic comparison. Model 3 produced the lowest AIC, whereas Model 2 produced the lowest BIC. Model selection was therefore not unambiguous. Model 3 was retained as the main substantive specification because it directly addressed the attribute-based research question while preserving the degree terms, and its interpretation was evaluated alongside the more parsimonious Model 2 and the goodness-of-fit diagnostics.

Because citation networks are temporally constrained, the ERGM results should be interpreted with caution. Unlike ordinary cross-sectional social networks, citation networks do not allow all ordered dyads to form ties: a later-published article may cite an earlier-published article, but an earlier-published article cannot cite a later-published one. Therefore, the ERGM analysis in this study should be understood as a static descriptive model of the observed citation pattern within the constructed corpus rather than as a full temporal model of the citation-generation process. The model is used to examine whether the observed citation matrix displays sparsity, degree concentration, and attribute-based clustering, but it does not claim to capture the complete temporal dynamics through which citations are produced over time.

3.4 Reliability and robustness checks

Various diagnostic checks were performed to assess the robustness of the findings. Overdispersion was tested in count

data models, and Poisson and Negative Binomial models were compared (Hilbe, 2011). To assess whether the estimates for the remaining article characteristics were sensitive to the inclusion of Google Scholar citations, the negative binomial model was re-estimated after removing `log_scholar_citation` while retaining the same 105 articles. The sole experimental study (Anysiadou and Gkliati, 2025) was grouped with the quantitative reference category in both negative binomial specifications because its experimental design was quantitative in form. The Google Scholar-excluded specification was treated as a sensitivity analysis rather than as an alternative primary model. The risk of multicollinearity among independent variables was assessed using the Variance Inflation Factor (VIF), and alternative model specifications were established, particularly for conceptually related variables such as country and region (Kutner et al., 2005). Because some homophily variables were conceptually adjacent, particularly country and geographic region, additional dyad-level collinearity diagnostics were calculated for the ERGM homophily indicators. Detailed dyad-level multicollinearity diagnostics are provided in [Supplementary Appendix B](#). The VIF values for the same-country, same-region, same-journal, same-theme, and same-design indicators ranged from 1.004 to 1.105, with the highest value observed for same-country. The pairwise phi correlation between same-country and same-region was also modest ($\phi = .279$). These results indicate that, despite their conceptual adjacency, the homophily terms did not create problematic empirical multicollinearity. Country and region were therefore retained in the same specification because they represent different levels of citation proximity: country captures national academic, policy, and publication contexts, whereas region captures broader geographical knowledge clusters. ERGM adequacy was evaluated using AIC and BIC comparisons together with goodness-of-fit diagnostics for indegree, outdegree, shared-partner, and geodesic-distance distributions. Detailed goodness-of-fit diagnostics for Models 1–4 are provided in [Supplementary Appendix A](#). These criteria did not identify a uniformly superior specification: Model 3 had the lowest AIC, whereas the more parsimonious Model 2 had the lowest BIC. The simulated distributions indicated that Model 3 generally approximated the observed degree distributions more closely than Models 1 and 4, although discrepancies remained for more complex shared-partner structures. The results were therefore interpreted comparatively rather than as evidence of unequivocal model superiority (Burnham and Anderson, 2004). In addition, the research design variable was included as a covariate in the models to control for the potential dominant effect of systematic reviews and literature searches. All analyses were conducted using the R statistical software environment.

To ensure the accuracy and consistency of the article-level attributes, a predefined data-extraction framework was used for publication year, country, geographic region, journal, indexation status, research design, and dominant theme. These variables were treated as article attributes rather than as the products of an independent interpretive content-coding process. One researcher determined and entered the attributes using the source articles and bibliographic records. A second researcher subsequently checked all entries against the same sources to identify and correct possible classification or manual data-entry errors. Any unclear cases were discussed and resolved by

reference to the predefined variable definitions. Because this procedure involved verification rather than independent double coding, an inter-rater agreement coefficient such as Cohen's kappa or Krippendorff's alpha was not applicable.

4 Results

4.1 Research question 1: What are the structural characteristics of the citation network in research on artificial intelligence in school leadership and management?

To conduct the analysis of the first hypothesis, a directed social network matrix of 105 articles on AI in the field of school leadership and management was constructed. The matrix exhibited that the citation network comprised 105 nodes and 334 directed links. The network density was found to be 0.031. This value indicates that only approximately 3% of all possible direct links ($105 \times 104 = 10,920$) were observed. This finding reveals that the citation network of the research corpus is structurally sparse. It also indicates that the observed citation network is structurally sparse within the constructed corpus. However, because citation ties are temporally constrained, this density value should be interpreted as a descriptive measure of the observed static citation matrix rather than as the probability that any two articles were equally at risk of forming a citation tie. The low density therefore reflects the limited number of observed citation connections in the corpus, while the centralization results provide additional evidence that citations are concentrated around a limited number of nodes.

The global transitivity coefficient of the network was 0.375, suggesting that the citation relationships are not distributed randomly and that there are local clusters and tendencies toward triple closure around specific nodes. In other words, the citation network includes structural concentrations clustered around certain authority nodes even though the network is sparse. The global transitivity coefficient is reported here as a descriptive indicator of local clustering in the observed citation matrix. It is not interpreted as an ERGM-estimated transitivity mechanism. Because citation networks are directed and temporally constrained, shared-partner and triadic structures may reflect common reference traditions, publication timing, or concentration around central studies rather than a direct process of triadic closure. For this reason, transitivity was discussed descriptively, while the final ERGM specification focused on sparsity, degree heterogeneity, and attribute-based homophily.

As shown in [Table 2](#), the density parameter is negative and significant ($\beta = -3.456$, $p < .001$). The significant density parameter estimate indicates that the assumptions were met, thereby justifying the transition to further ERGM analysis with structural terms and homophily attributes. The negative and significant edges parameter indicates that the observed citation matrix is sparse after representing articles and citation relations as a directed network. However, because citation ties are shaped by publication timing, this coefficient should not be interpreted as the probability of citation formation across all theoretically ordered dyads. Rather, it indicates a low baseline tendency for

TABLE 2 Maximum likelihood estimation results of the ERGM first model.

Parameter	Estimate (coefficient)	Std. error	Z value	p
MODEL 1				
Density (edges)	−3.45615	0.05557	−62.2	<1e-04 ***
Null pseudo-deviance:	15,138 on 10,920 degrees of freedom			
Residual pseudo-deviance:	2,987 on 10,919 degrees of freedom			
AIC: 2,989 BIC: 2,996				

*** $p < .001$.

TABLE 3 Centralization values of the citation network.

Measure	Centralization
Degree	0.322
In-degree	0.155
Out-degree	0.155
Betweenness	0.004
Eigenvector	0.514

observed citation ties within the static network representation used in this study. The AIC value of Model 1 was found to be 2,989.

Table 3 presents network centralization patterns. The degree centralization is 0.322. This finding demonstrates that the citation relationships do not follow a completely dispersed pattern. Instead, citations are concentrated around specific studies. It is understood that citations are not homogeneously distributed across the network, and certain nodes occupy more central positions.

The indegree (receiving-citations) centralization value of 0.155 also indicates that citation behavior is concentrated among a subset of nodes. It is understood that some studies have become authoritative cores receiving significantly more references than other works. Likewise, the outdegree centralization across nodes (citing other works) was 0.155. This value also confirms that the authors' citing behavior is concentrated on specific works. This finding suggests that some studies are not only heavily referenced but also positioned as nodes that establish a broader citation repertoire in the literature. The equal levels of in-degree and out-degree centralization are noteworthy because they indicate that citation and citing patterns are not sharply divergent. The betweenness centralization was found to be quite low (0.004). This finding indicates that there are few intermediary studies in the literature that serve as bridges between study clusters, and that information flows largely directly through centralized studies. In other words, structural brokerage within the network appears weak. The limited brokerage structure was also examined through node-level betweenness centrality and cross-theme citation ties. Only 41 of the 105 articles had non-zero betweenness centrality, and the

highest normalized betweenness value was 0.0043. This indicates that very few articles occupy structurally meaningful intermediary positions between different parts of the citation network. Cross-theme citation ties were relatively common among coded ties, but these ties were largely directed toward or organized around central studies rather than distributed through a broad set of bridging articles. Therefore, the findings suggest limited structural brokerage rather than a complete absence of thematic cross-citation. The eigenvector centralization of 0.514 suggests that the network not only contains central nodes but also that these nodes form an interconnected core.

According to the centrality analyses, the study by Fullan et al. (2024) has the highest in-degree, with 36 citations. This is followed by Wang (2021b), with 27 citations, and Tyson and Sauers (2021), with 26 citations. The fact that these studies also have high Google Scholar citation values demonstrates that intra-network centrality correlates with broader academic visibility. This structure supports the formation of a distinct concentration of authority and a core-periphery pattern in the literature.

To provide more direct evidence for the observed core-periphery-like organization, component and k-core analyses were also conducted. Component analysis showed that the network contained 22 weak components, with the largest weak component including 82 of the 105 articles. Nineteen nodes were isolated. In the directed network, 104 strongly connected components were identified, and the largest strongly connected component contained only two nodes. This result is consistent with the temporally constrained nature of citation networks, in which reciprocal citation structures are structurally limited. The k-core analysis, conducted on the undirected projection of the citation network, identified a maximum k-core of ($k = 8$), which included 27 articles. By contrast, 34 articles had coreness values of 0 or 1. In this study, the core is therefore operationally defined as the maximum k-core of the citation network, while the periphery refers to isolated and low-coreness nodes that remain weakly connected to the main citation structure.

Taken together, these findings address RQ1 by showing that the observed citation network is sparse, centralized, and unevenly connected. Citation ties are concentrated around a limited number of highly cited studies, while component and k-core analyses indicate a citation-dense maximum k-core alongside numerous isolated or low-coreness articles. Because no formal core-periphery blockmodel was estimated, these findings are described as evidence of core-like citation concentration or a centralized citation structure rather than as confirmation of a formal core-periphery structure.

4.2 Research question 2: To what extent are observed citation ties associated with similarity in country, geographic region, journal of publication, research design, and dominant theme?

RQ2 examines whether observed citation ties are associated with similarity in country, geographic region, journal of publication, dominant theme, and research design. Three ERGM specifications were compared to distinguish attribute-based associations from degree-based structural dependencies. To test

this hypothesis, three different ERGM models were estimated. The first model considered only the structural dependencies of the network (edges, *gwidegree*, *gwodegree*); the second model included all homophily variables in addition to the structural terms; and the third model tested only the homophily terms by excluding degree dependencies. This multi-stage modeling strategy enables evaluation of whether homophily effects are independent of the network's intrinsic structural properties.

The structural model (Model 2) exhibits that the network has a sparse structure characterized by degree dependencies. The density coefficient is negative and significant ($\beta = -1.5949$, $p < .001$). Similarly, the *gwidegree* term, representing in-degree dependence, produced a strong negative coefficient ($\beta = -4.6010$, $p < .001$). The *gwodegree* term, representing out-degree dependence, was also negative and significant ($\beta = -3.1477$, $p < .001$). These results indicate that observed citation ties are concentrated around certain central nodes and that degree-based structural dependencies are strongly associated with the observed citation pattern. The AIC value of Model 2 is 2,231.

Model 3, including both homophily variables and structural terms, is the fundamental evaluation model for RQ2. In this model, the coefficients for density ($\beta = -1.6032$, $p < .001$), *gwidegree* ($\beta = -4.5027$, $p < .001$), and *gwodegree* ($\beta = -3.1467$, $p < .001$) remained significant, clarifying that structural dependencies contribute consistently to the model. Among the homophily terms, only the same-country coefficient showed a weak marginal positive association with observed citation ties ($\beta = 0.8122$, $p = .078$); the remaining terms were not statistically significant. Model 3 had the lowest AIC (2,216), whereas Model 2 had the lower BIC (2,253 compared with 2,275 for Model 3). Accordingly, the information criteria do not establish unequivocal superiority of Model 3; rather, Model 3 is retained as the main substantive specification because it directly evaluates the attribute-based associations addressed in RQ2 while accounting for degree dependencies.

In Model 4, which includes only homophily terms, same-country ($\beta = 1.1846$, $p < .001$), same-journal ($\beta = 0.6768$, $p = .036$), and same-design ($\beta = 0.3335$, $p = .009$) effects are positive and statistically significant. By contrast, the same-region coefficient is negative and statistically significant ($\beta = -0.4991$, $p < .001$), indicating that citation ties are less likely among studies from the same broad geographic region in this homophily-only specification. This opposite direction should be interpreted cautiously because Model 4 does not control for degree-based structural dependencies. When *gwidegree* and *gwodegree* terms are included in Model 3, the region effect is no longer statistically significant. Therefore, the negative region coefficient in Model 4 is treated as a model-specific diagnostic association rather than as evidence of a stable regional heterophily pattern.

When *gwidegree* and *gwodegree* are included in Model 3, the journal, region, and design coefficients are no longer statistically significant. This indicates that the associations observed in Model 4 are not robust after accounting for degree-based structural dependencies. Accordingly, the Model 4 coefficients should not be interpreted as evidence that shared country, journal, or research-design characteristics independently generate citation ties. They are better understood as model-dependent associations that overlap with the concentration of citations around highly connected studies. Model 3 was

therefore used as the main substantive specification for RQ2, while Model 2 was retained as an important, more parsimonious comparison.

Regarding RQ2, the models provide no consistent evidence that similarity in country, geographic region, journal, dominant theme, or research design is independently associated with observed citation ties after accounting for degree dependencies. Only the same-country term showed a weak marginal positive association. The results therefore indicate that the observed citation pattern is more closely associated with degree concentration than with the homophily attributes included in the analysis. This interpretation is conditional on the static network specification and should not be treated as evidence concerning temporal citation-formation mechanisms.

4.3 Research question 3: To what extent does the observed within-corpus citation distribution exhibit heavy-tailed concentration consistent with, but not sufficient to establish, preferential attachment?

To address RQ3, the within-corpus indegree distribution was examined to determine whether it displayed heavy-tailed citation concentration compatible with, but not sufficient to establish, preferential attachment. The within-corpus indegree distribution is illustrated in Figure 2. The network contains 105 articles with an average indegree of 3.18. However, the distribution is highly asymmetric (Min = 0; Median = 1; Mean = 3.18; Max = 36). Fifty-one of the articles received no citations, and 20 received only one citation (see Table 4). In contrast, a very limited number of articles achieved double-digit citation values. This indicates that citations are largely concentrated in a small number of high-impact articles.

The indegree histogram reveals that the vast majority of articles received between 0 and 3 citations; in contrast, only a very small number of articles received more than 10 citations. This structure indicates a heavy-tailed distribution. In the second step, the log-log-scaled degree distribution graph was analyzed to determine whether it followed a power law.

As shown in Figure 3, while concentration is high at lower citation levels, certain studies in the upper tail achieve higher citation levels. The fact that points in a log-log degree distribution generally show a linear trend suggests that the citation distribution may exhibit a structure similar to a power-law character. Therefore, in the third step, the upper-tail behavior of the citation distribution was tested within the power-law framework, following the approach of Clauset et al. (2009).

The upper-tail behavior of the citation distribution was examined using the maximum likelihood approach proposed by Clauset et al. (2009). The estimated minimum threshold was ($x_{\min} = 11$), indicating that the fitted model applies only to the upper tail of the citation distribution rather than to the full distribution. The number of articles in this upper tail was ($n_{\text{tail}} = 13$). The estimated power-law exponent was ($\alpha = 3.441$), with a 95% confidence interval of [2.111, 4.771]. The fitted upper-tail CCDF is shown in Figure 4. This wide interval indicates substantial uncertainty around the

TABLE 4 Structure and homophily analysis of citation network.

Parameter	Estimate (coefficient)	Std. error	Z value	p
MODEL2				
Density (edges)	−1.5949	0.2474	−6.445	<1e-04***
gwideg.fixed.0.5	−4.6010	0.3949	−11.652	<1e-04***
gwodeg.fixed.0.5	−3.1477	0.4021	−7.829	<1e-04***
Null Deviance:	15,138 on 10,920 degrees of freedom			
Residual Deviance::	2,225 on 10,917 degrees of freedom			
AIC: 2,231 BIC: 2,253 (Smaller is better. MC Std. Err. = 0)				
MODEL3				
Density (edges)	−1.60316	0.31328	−5.117	<1e-04***
gwideg.fixed.0.5	−4.50265	0.44514	−10.115	<1e-04***
gwodeg.fixed.0.5	−3.14669	0.45059	−6.983	<1e-04***
nodematch.country	0.81224	0.46071	1.763	0.0779
nodematch.region	−0.42690	0.31967	−1.335	0.1817
nodematch.journal	−0.25133	0.50292	−0.500	0.6173
nodematch.theme	0.03317	0.36549	0.091	0.9277
nodematch.design	0.39924	0.26920	1.483	0.1381
Null Deviance:	15,138 on 10,920 degrees of freedom			
Residual Deviance::	2,200 on 10,912 degrees of freedom			
AIC: 2,216 BIC: 2,275 (Smaller is better. MC Std. Err. = 0)				
MODEL4				
Density (edges)	−3.45569	0.07969	−43.366	<1e-04***
nodematch.country	1.18463	0.20406	5.805	<1e-04***
nodematch.region	−0.49913	0.13253	−3.766	0.000166***
nodematch.journal	0.67684	0.32258	2.098	0.035886*
nodematch.theme	0.02365	0.16061	0.147	0.882938
nodematch.design	0.33347	0.12723	2.621	0.008766**
Null Deviance:	15,138 on 10,920 degrees of freedom			
Residual Deviance::	2,941 on 10,914 degrees of freedom			
AIC: 2,953 BIC: 2,996 (Smaller is better. MC Std. Err. = 0)				

* $p < .05$; ** $p < .01$; *** $p < .001$.

exponent estimate, which is expected given the small number of observations in the tail.

Additional likelihood-ratio comparisons showed that the power-law model did not significantly outperform alternative heavy-tailed distributions, including the lognormal distribution ($R = -0.098$, $p = .724$), the exponential distribution ($R = 0.122$, $p = .879$), and the stretched-exponential distribution ($R = -0.115$, $p = .725$). Therefore, the power-law fit should not be interpreted as strong evidence that the citation distribution follows a power-law form. Rather, the results indicate that the citation distribution is highly skewed and heavy-tailed, with citation counts concentrated around a small number of highly visible studies. The log-log plot is used only as a visual diagnostic of this upper-tail concentration and not as statistical evidence for a power-law mechanism.

Regarding RQ3, the observed indegree distribution is highly skewed and exhibits upper-tail concentration. Nevertheless, only

13 articles were included in the estimated upper tail, the exponent estimate had a wide confidence interval, and the power-law model did not significantly outperform the alternative heavy-tailed distributions. The findings therefore support citation inequality, heavy-tailed concentration, and visibility-linked centralization within the corpus. They do not establish preferential attachment as a temporal generative mechanism.

4.4 Research question 4: Which article characteristics are associated with higher within-corpus citation counts?

To address RQ4, a negative binomial regression model was estimated to examine associations between article characteristics and within-corpus citation counts. The dependent variable was

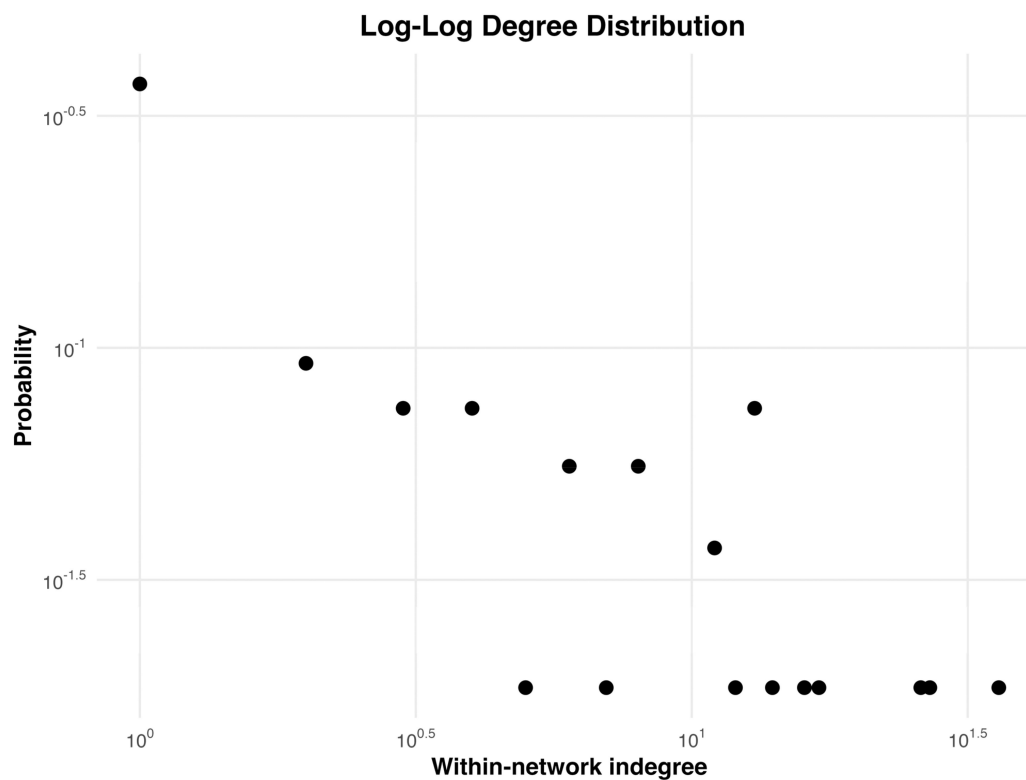


FIGURE 3

Log-Log degree distribution. The graph displays the distribution of within-network citation counts on logarithmic axes. The figure is used only as a visual diagnostic of skewness and upper-tail concentration; it is not treated as statistical evidence for a power-law distribution.

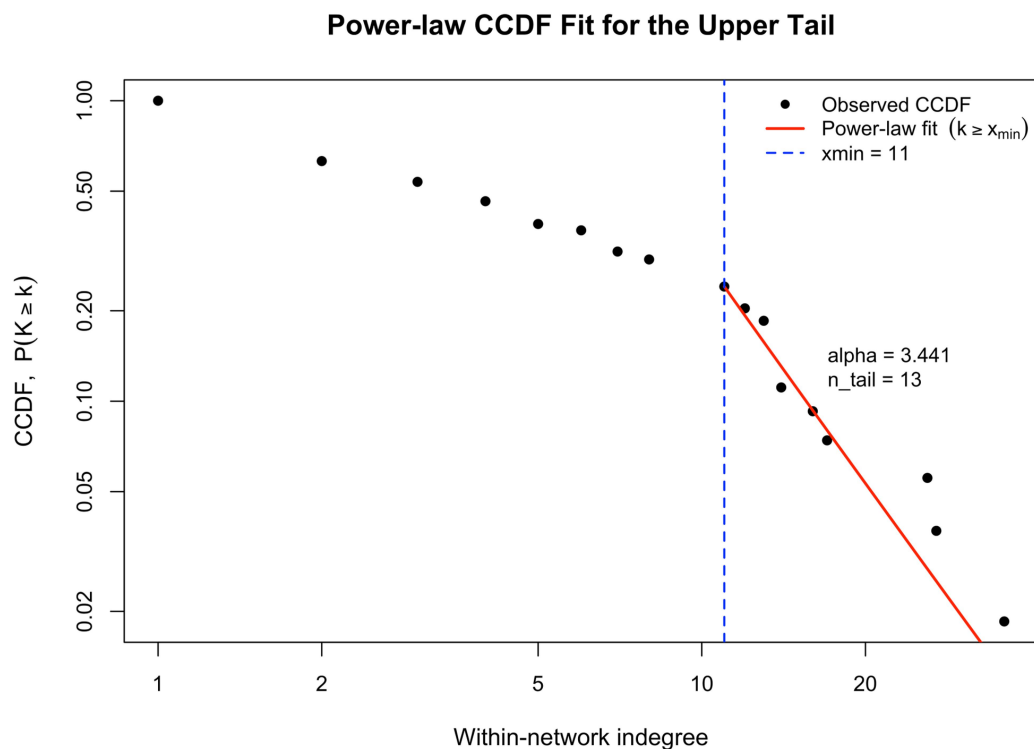


FIGURE 4

Power-law CCDF fit for the upper tail of the within-network indegree distribution. The observed complementary cumulative distribution function is shown together with the fitted power-law curve above the estimated threshold $x_{\min} = 11$. The fitted tail includes $n_{\text{tail}} = 13$ articles, and the estimated exponent is $\alpha = 3.441$. The figure is used as a diagnostic visualization of upper-tail concentration rather than as conclusive evidence of a power-law distribution.

the number of citations (indegree) received by each article within the network. Because the citation counts were overdispersed, negative binomial regression was preferred to Poisson regression. In the primary model, the dispersion parameter was estimated as $\theta = 2.076$ (SE = 0.679). The model produced an AIC of 356.47 and a residual deviance of 95.101.

The descriptive characteristics of the predictor variables included in the model are presented in Table 4. Reference categories were determined for the categorical variables to provide a basis for analytical comparison. In the Region variable, Africa was chosen as the reference category. This choice was preferred because it provides a neutral comparison base that is underrepresented and does not constitute a dominant center within the network. In the index_group variable, the Both category (studies indexed in both WoS and Scopus) has been selected as the reference group. This selection allows for comparison of other index categories based on their strong indexing position. In the design variable, quantitative studies were determined as the reference category. Thus, the effects of other methodological designs were interpreted by taking quantitative research as the primary comparison group.

According to Table 5, the Google Scholar citation variable shows a right-skewed distribution. The fact that the average Google Scholar citation value is 24.12 and the median is only 6 indicates that the limited number of high-visibility articles is pulling the distribution upwards. Besides, out of a total of 105 studies, 26 were literature reviews, and 19 were systematic literature reviews. Accordingly, 42.9% of the studies on AI in school leadership and management were literature reviews, while empirical studies accounted for 57.1%. This relatively strong presence of review-type studies is important for interpreting the network structure, because such studies may attract citations not

only due to their methodological design but also because they summarize, classify, and stabilize an emerging field. Therefore, their prominence may contribute to citation concentration around a limited number of epistemic anchor studies.

Table 6 presents the results of the updated primary negative binomial model estimated with all 105 articles. The estimated dispersion parameter was $\theta = 2.076$ (SE = 0.679), and the model produced an AIC of 356.47. The null deviance decreased from 332.114 to a residual deviance of 95.101.

The mean-centered publication-year coefficient was not statistically significant ($\beta = 0.0233$, SE = 0.1440, $z = 0.162$, $p = .872$). None of the regional categories differed significantly from the African reference category: Asia ($\beta = 0.2448$, $p = .625$), Australia ($\beta = 0.3428$, $p = .623$), Europe ($\beta = 0.2056$, $p = .750$), North America ($\beta = 0.3258$, $p = .570$), and South America ($\beta = -2.6428$, $p = .083$).

None of the indexation or research-design contrasts reached the conventional .05 significance level in the primary model. The coefficients were $\beta = -0.0245$ for Scopus ($p = .936$), $\beta = 1.5162$ for WoS ($p = .084$), and $\beta = -0.7823$ for the Others indexation group ($p = .074$). Literature review design showed a positive but statistically nonsignificant coefficient relative to quantitative studies ($\beta = 0.7383$, $p = .072$; IRR = 2.09). Mixed ($\beta = -0.8520$, $p = .091$), qualitative ($\beta = 0.2038$, $p = .592$), and systematic literature review designs ($\beta = 0.2176$, $p = .611$) were also not statistically significant. Thus, the primary model did not provide statistically significant evidence of research-design differences in within-corpus citation counts.

The strongest and only statistically significant association among the substantive predictors in the primary model was observed for log_scholar_citation ($\beta = 0.7852$, SE = 0.1157, $z = 6.788$, $p < .001$; IRR = 2.19, 95% CI [1.75, 2.75]). A one-unit increase in log(1 + Google Scholar citations) was associated with an expected within-corpus citation rate approximately 2.19 times higher. Google Scholar counts capture citations from a much broader scholarly literature and are therefore not equivalent to within-corpus indegree. Nevertheless, because both are citation-based measures and may partially overlap, the coefficient should be interpreted as an association between broader academic citation recognition and centrality within the constructed corpus rather than as a temporally prior or causal effect.

To assess sensitivity to the inclusion of Google Scholar citations, the model was re-estimated without log_scholar_citation using the same 105 articles. In this specification, publication year was negatively associated with within-corpus citation counts ($\beta = -0.7214$, $p < .001$), the Others indexation category had a negative coefficient ($\beta = -2.2926$, $p < .001$), and literature review ($\beta = 1.9202$, $p < .001$) and qualitative designs ($\beta = 1.0784$, $p = .019$) had positive coefficients. The systematic literature review coefficient was marginal ($\beta = 0.9509$, $p = .059$). The direction or significance of several estimates therefore differed between the two specifications, indicating that the remaining coefficients were not stable when Google Scholar citations were excluded. The complete sensitivity-model results are reported in Supplementary Appendix C.

Regarding RQ4, broader Google Scholar citation recognition was positively associated with within-corpus citation counts in the primary model. The coefficients for publication year, indexation, and research design were sensitive to whether this citation-based variable was included; consequently, they should

TABLE 5 Descriptive and categorical information on the research corpus.

Predictor	Categories/ Summary	Reference category
Year	2021–2026 (Median = 2025)	Continuous
Region	Africa, Asia, Australia, Europe, North America, South America	Africa
Index Group	Both ($n = 49$), Scopus ($n = 18$), WoS ($n = 3$), Others ($n = 35$)	Both
Design	Mixed ($n = 11$), Quantitative ($n = 28$), Experimental ($n = 1$), Qualitative ($n = 20$), Literature Review ($n = 26$), Systematic Review ($n = 19$)	Quantitative/experimental (combined $n = 29$ in the negative binomial models)
Google Scholar Citation	Min = 0; Median = 6; Mean = 24.12; Max = 321	Continuous
$n = 105$		

For the negative binomial analyses, the sole experimental study was grouped with the quantitative studies, producing a reference category of $n = 29$. The original research-design coding was retained for the descriptive presentation of the corpus.

TABLE 6 Negative binomial regression analysis.

Parameter	Estimate (coefficient)	Std. error	z value	p
(Intercept)	−1.54423	0.60732	−2.543	.0110*
Year-centered	0.02328	0.14402	0.162	.8716
regionAsia	0.24480	0.50119	0.488	.6252
regionAustralia	0.34281	0.69801	0.491	.6233
regionEurope	0.20558	0.64448	0.319	.7497
regionNorth America	0.32584	0.57379	0.568	.5701
regionSouth America	−2.64285	1.52405	−1.734	.0829
index_groupScopus	−0.02454	0.30557	−0.080	.9360
index_groupWoS	1.51624	0.87665	1.730	.0837
index_groupOthers	−0.78230	0.43838	−1.785	.0743
designliterature review	0.73834	0.41056	1.798	.0721
designmixed	−0.85198	0.50398	−1.691	.0909
designqualitative	0.20376	0.37998	0.536	.5918
designsystematic literature review	0.21763	0.42837	0.508	.6114
log_scholar_citation	0.78517	0.11567	6.788	<.001***
Model information				
Number of observations	105			
AIC	356.474			
Number of Fisher scoring iterations	1			
Dispersion parameter (θ)	2.076	0.679		
Log-likelihood	−162.237			
2 × log-likelihood	−324.474			
Null deviance	332.114		df = 104	
Residual deviance	95.101		df = 90	

The dependent variable is the within-corpus citation count (indegree). Reference categories are Africa for region, Both for indexation group, and quantitative/experimental for research design. Year is mean-centered, and Google Scholar citations are transformed as log(1 + citations). $p < .05$; * $p < .05$; *** $p < .001$.

be interpreted as model-dependent associations rather than stable article-level differences.

Figure 5 summarizes the incidence rate ratios and their 95% confidence intervals for the updated primary negative binomial model. The confidence intervals for publication year, all regional and indexation categories, and all research-design categories intersect the IRR = 1 reference line. Although the point estimates for WoS indexation and literature review design are above 1, their confidence intervals include 1 and therefore do not provide statistically significant evidence of an association in the primary model. The confidence interval for log_scholar_citation does not intersect 1, indicating the strongest association with within-corpus citation counts. This association should be interpreted as correspondence between broader citation recognition and within-corpus centrality rather than as a causal visibility effect.

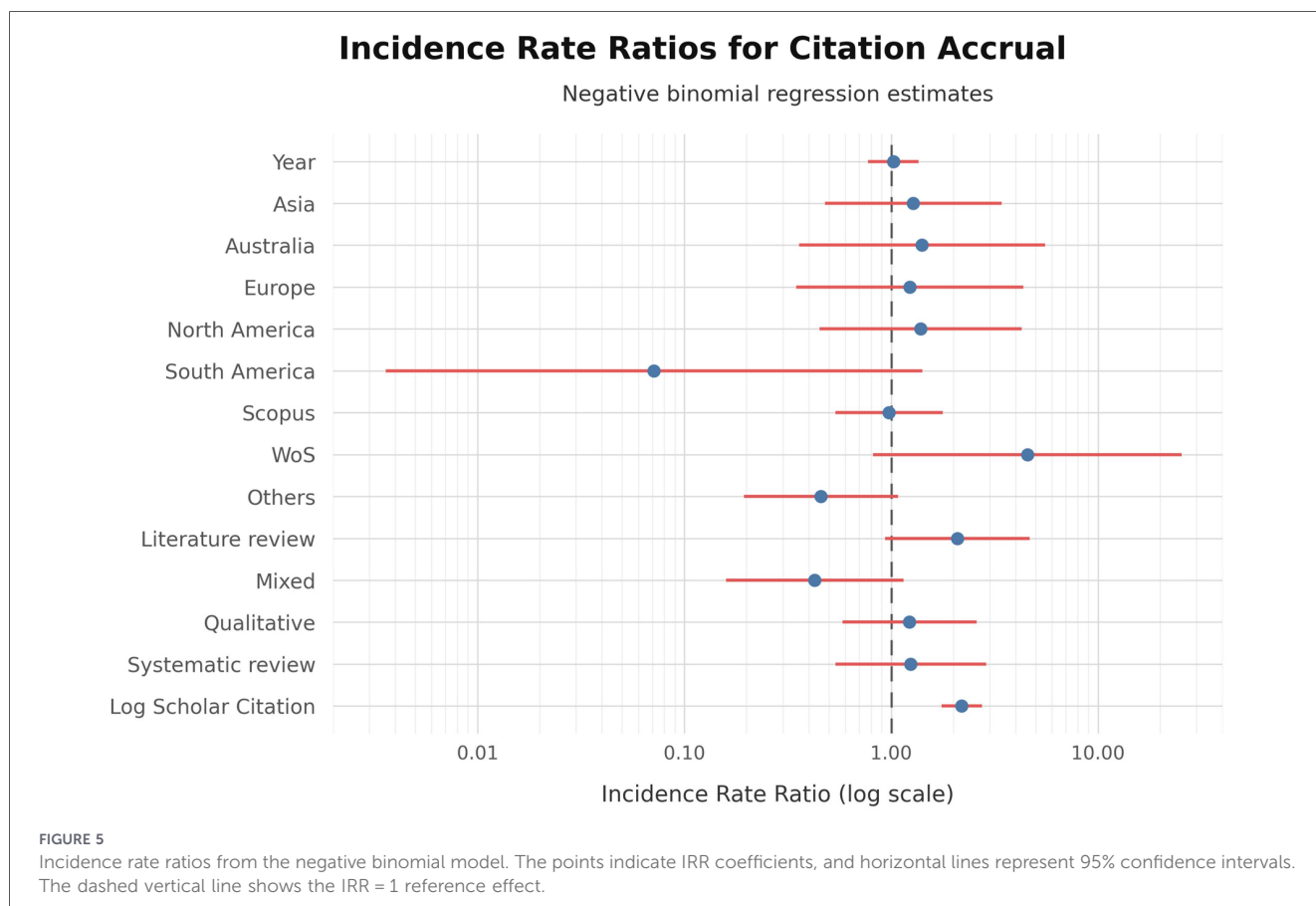
Figure 6 shows that articles with higher within-network citation counts are concentrated in the core region of the network. This pattern indicates that the observed citation structure is organized around a limited number of highly cited studies, while many peripheral studies remain weakly connected to the central part of the corpus. This visual pattern demonstrates the coexistence of a

citation-dense core and a sparsely connected periphery within the observed network. The relatively high indegree levels of studies in the first and second rings surrounding the center suggest that the field’s knowledge base is beginning to take shape around a limited set of central references. At the same time, the lack of studies serving as bridges between the center and the periphery is also evident in Figure 6.

Taken together, the findings addressing RQ4 indicate that broader Google Scholar citation recognition was positively associated with within-corpus citation counts in the primary model. Publication year, regional affiliation, indexation, and research design did not show statistically significant associations in that specification. Several of these coefficients changed when Google Scholar citations were excluded, indicating that the article-level findings beyond broader citation recognition are sensitive to model specification.

5 Discussion

The findings indicate that the literature on AI in school leadership and management is organized around a relatively

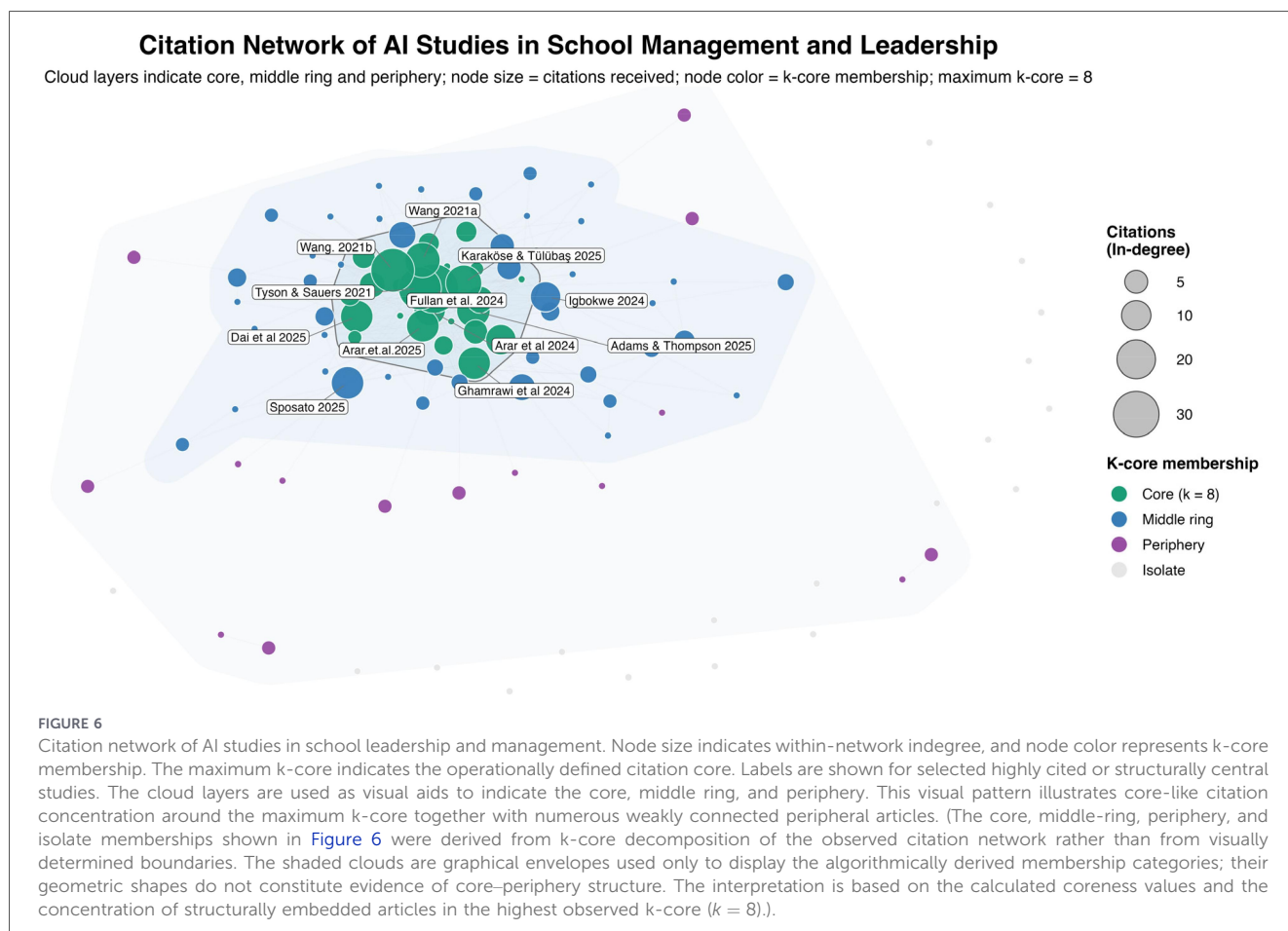


small number of influential studies and displays a sparse network structure. This can be taken as evidence for the suggestion that the field is in an early stage of development, and that the observed knowledge structure is organized around a limited set of central studies and research clusters. This finding aligns with previous studies depicting AI in school leadership and management literature as a limited, emerging, and insufficiently explored research domain (Adams and Thompson, 2025; Berkovich and Eyal, 2025; Bixler and Ceballos, 2025; Fullan et al., 2024; Kafa, 2025a, 2025b; Kılınç et al., 2025; Richardson et al., 2025; Tyson and Sauers, 2021; Wang, 2021a). Similarly, Polat et al. (2025) demonstrate strong geographical and institutional centralization in this literature, noting that studies are concentrated in specific regions and around elite institutions, such as the United States, Canada, and Australia. Arar et al. (2025) likewise showed that the United States occupies a central position in terms of both single-country broadcasting and multi-country collaborations. Hence, the centralization identified in this study is not merely a technical consequence of network topology but also a structural indicator of the geography of knowledge production, the academic visibility regime, and the distribution of epistemic power within the field.

The concentration of citations around a small number of highly visible studies is compatible with cumulative citation advantage, but it does not establish preferential attachment as a temporal generative mechanism. Studies such as Wang (2021a), Tyson and Sauers (2021), Arar et al. (2024, 2025), Fullan et al. (2024), Berkovich and Eyal (2025), Berkovich (2025), Dai et al.

(2025), Adams and Thompson (2025), Karaköse et al. (2023), and Richardson et al. (2025) occupy prominent positions as reference points within the observed corpus. Tyson and Sauers' (2021) discussion of diffusion networks, trusted colleagues, repeated exposure, and opinion leaders provides one possible theoretical interpretation of this concentration. Nevertheless, because the present analysis is based on an aggregated citation network rather than longitudinal citation trajectories, it cannot determine whether these studies accumulated citations through preferential attachment over time. The findings therefore indicate citation concentration and visibility-based inequality rather than confirmation of a preferential-attachment mechanism.

The relationship between peripheral studies that do not receive citations and central studies that do not cite them is one of the most original contributions this study makes to the literature. While the current literature often focuses on the visibility of core studies, the constructive effects of review articles, or thematic gaps, it fails to address the invisibility of peripheral studies alongside their active citation behavior. This finding shows that being on the periphery is not just a matter of not receiving citations, but also a matter of not being able to connect, or not wanting to connect, to the central knowledge base. In other words, peripheral studies struggle to circulate within the literature and gain visibility when they fail to connect with central theoretical and methodological lines. This situation can be linked to the problem of the global South, non-Western societies, and underrepresented contexts not being sufficiently visible in the literature, as highlighted by Arar et al. (2024,



2025). Therefore, the periphery of the citation network is not merely a quantitative visibility problem, but also a matter of information inclusion, representation, and epistemological access.

The lack of studies bridging the gap between the periphery and the center is also related to the fragmented nature of the literature. While Kılınç et al. (2025) point to the descriptive and fragmented nature of the literature, Kafa (2025a), Kafa (2025b), Bixler and Ceballos (2025), and Berkovich and Eyal (2025) emphasize that empirical gaps, conceptual ambiguities, and the need for an integrative model persist. This explains the weakness of bridging positions at the level of the structural citation network. However, some recent studies suggest that this finding should be interpreted with caution. Dai et al. (2025) build conceptual bridges between AI-assisted administrative decision-making and classical decision theory; Adams and Thompson (2025) build bridges between AI, ChatGPT, school administration, and instructional leadership; Richardson et al. (2025) build bridges between practitioner literature, ethics, equity, and school leadership; and Sposato (2025) integrates scattered literature into a ten-domain taxonomy. Therefore, it can be concluded that, while relationships are beginning to emerge between the sub-dimensions of school leadership and management and AI, peripheral and low-citation studies are not interacting sufficiently with high-citation studies at the center. This situation may be due to the field's fragmented thematic structure.

The weak marginal same-country association suggests that contextual proximity and national information-dissemination

patterns may be related to observed citation ties. Studies such as Kılınç et al. (2025), Kafa (2025a, 2025b), Berkovich and Eyal (2025), and Berkovich (2025) indicate that relationships between AI and leadership are embedded in national education systems and context-specific implementation conditions. Nevertheless, the region term was not significantly associated with observed citation ties after degree dependencies were included. The citation pattern therefore cannot be attributed to regional affiliation and appears more closely associated with degree concentration, broader academic recognition, and proximity to central literature.

The finding concerning thematic clustering requires careful refinement in light of existing literature. The results of this study demonstrate that the AI literature in school leadership and management has not yet undergone a crystallized thematic differentiation. This result is consistent with studies indicating that the field is limited, early-stage, fragmented, and largely theoretical (Anysiadou and Gkliati, 2025; Berkovich and Eyal, 2025; Bixler and Ceballos, 2025; Kafa, 2025a, 2025b; Kılınç et al., 2025). However, current scientific mapping and taxonomy studies necessitate a significant softening of this interpretation. Arar et al. (2025) show that themes such as performance, management, work, ethics, skills, decision-making, collaboration, and framework become prominent in co-occurrence network and thematic map analyses. Polat et al. (2025) reveal the formation of distinct thematic clusters through analyses of bibliographic coupling, co-citation, and keyword clustering.

Sposato (2025) developed a comprehensive taxonomy that classifies AI applications into ten distinct areas, including administrative efficiency, personalized learning, governance, ethical AI leadership, DEI initiatives, and strategic planning. Richardson et al. (2025) identify the themes of equity/ethics, teaching-learning enhancement, stakeholder collaboration, and exploration-reflection; Ghamrawi et al. (2024) propose competency clusters such as technological literacy, adaptability, coaching, data-informed decision-making, and human-centeredness; and Karaköse et al. (2023) reveal the sub-dimensions of digital leadership, technology integration, school culture, professional development, data-driven decision-making, and digital citizenship. Therefore, a more defensible interpretation is that thematic differentiation has begun in the literature, but it has not yet fully matured into stable, institutionalized citation-based communities. In other words, the field is not entirely amorphous in its current state, but thematic crystallization, the process of transforming into collections of references, is in its early stages.

The absence of significant homophily effects suggests that citation formation in this corpus is not primarily organized around shared country, region, journal, theme, or research design. One possible explanation is that research on AI in school leadership and management remains an emerging and interdisciplinary field whose knowledge base is still being assembled across otherwise distinct scholarly communities. Researchers may therefore cite broadly visible or conceptually foundational studies across categorical boundaries rather than preferentially citing articles that resemble their own. This pattern may indicate that the field has not yet developed stable and internally cohesive citation clusters based on these attributes and that its intellectual organization is currently shaped more by cross-cutting references than by attribute-based citation communities. Nevertheless, given the size and composition of the corpus, the nonsignificant estimates should be interpreted as an absence of detected homophily within this bounded network rather than definitive evidence that such mechanisms do not exist.

The marginally positive coefficient for articles indexed only in WoS suggests a possible association between indexation, academic visibility, and within-corpus citation counts. Polat et al. (2025) and Arar et al. (2025) emphasize the role of WoS and Scopus in academic visibility and scientific mapping. Nevertheless, the WoS coefficient was only marginally significant and should not be interpreted as evidence of a causal indexation or prestige mechanism. It may instead indicate a relationship between discoverability and within-corpus citation counts that requires confirmation in larger and temporally structured datasets.

The primary model provided no statistically significant evidence of indexation differences in within-corpus citation counts. The WoS and Others coefficients were marginal, but neither reached the conventional significance threshold. Moreover, the WoS coefficient changed direction and remained nonsignificant when Google Scholar citations were excluded, whereas the Others category became significantly negative. This divergence indicates that the indexation coefficients are sensitive to model specification and should not be interpreted as evidence of a stable prestige or discoverability mechanism.

The inclusion of empirical studies, literature reviews, systematic reviews, bibliometric mappings, conceptual syntheses,

and taxonomies within the same citation network remains important for interpreting citation concentration. The literature-review coefficient was positive but nonsignificant in the primary model and became larger when Google Scholar citations were excluded. Qualitative design also became significant only in the sensitivity specification. These changes suggest that research-design associations are intertwined with broader citation recognition and do not constitute a stable independent citation advantage. Nevertheless, review-type studies may still serve as organizing references by synthesizing dispersed knowledge and providing conceptual entry points, while early empirical studies may function as influential reference points. The observed authority structure may therefore reflect the combined organizing role of review publications and pioneering empirical studies rather than a uniform advantage associated with any single research design. The observed citation concentration should not, however, be interpreted as evidence that the field is organized around a small number of dominant research groups. The present analysis models citation relations among articles rather than authors, institutions, or collaborative teams. Determining whether intellectual authority is also concentrated within particular research groups would require a separate co-authorship or institutional network analysis.

Finally, Google Scholar citation counts and within-corpus indegree represent related but non-equivalent measures. Google Scholar captures citations originating from a much wider scholarly literature, including publications outside the constructed corpus, whereas within-corpus indegree records only citation ties among the 105 included articles. Some overlap may nevertheless remain because both variables are citation-based. The sensitivity analysis further showed that excluding Google Scholar citations altered the direction or statistical significance of several remaining coefficients. Accordingly, the primary association should be interpreted as cross-sectional correspondence between broader citation recognition and centrality within this bounded network, while the other article-level associations should be regarded as specification-dependent.

To sum up, this study demonstrates that the artificial intelligence literature in school leadership and management has evolved through three simultaneous processes. First, the observed citation structure is concentrated around a limited number of central studies, displaying a pattern compatible with cumulative citation advantage but not confirming preferential attachment as a temporal mechanism. Second, the knowledge structure is becoming increasingly thematically differentiated, displaying early crystallization around focal areas such as instructional leadership, digital leadership, school management, administrative decision-making, ethics, equity, teacher leadership, and human-AI collaboration. Thirdly, while structural citation brokerage remains weak, conceptual, intellectual, ethical, and human-AI brokerage forms have begun to develop. Therefore, the field exhibits not a tension but a simultaneity between a centralized citation structure and evolving thematic diversification. The AI literature in school leadership and management does not stand as a completely mature body of knowledge yet; however, it presents a dynamic research field progressing towards institutionalization through core authorities, thematic focuses, and new forms of mediation.

For scholars and practitioners of school leadership and educational management, these findings suggest that the development of AI-related knowledge should not be guided only by highly visible review studies or a small group of central reference works. While such studies help organize an emerging field, an overreliance on them may narrow the range of educational contexts, leadership practices, and empirical problems that inform AI-related decision-making in schools. A more balanced knowledge base requires studies that connect central debates on ethics, data-informed decision-making, instructional leadership, and human–AI collaboration with diverse empirical settings and underrepresented educational systems. For school leaders and educational management researchers, this means that AI leadership should be approached not merely as a technical or administrative innovation, but as a developing educational field that requires contextual evidence, critical engagement with dominant sources, and stronger bridges between conceptual models and school-level practice.

6 Conclusion

The findings indicate a centralized citation structure and core-like concentration around the maximum k-core rather than a formally established core–periphery structure. Outdegree and indegree centrality values support this finding. In addition, the eigenvector centrality value indicates that studies that become central and receive high citations in the network are not only visible but also strongly interconnected. The fact that a very large proportion of studies conducted on the periphery have received no citations suggests a lack of interaction between these studies and the central research area. Likewise, the relatively low betweenness centrality value reveals a lack of intermediary studies that could bridge the gap between studies at the center and those at the periphery. These findings suggest that the flow of information in the literature is not widely and horizontally distributed, but rather concentrated around a limited number of core studies.

The ERGM results showed only a weak marginal same-country association. Similarity in geographic region, journal, dominant theme, and research design was not consistently associated with observed citation ties after degree dependencies were included. Accordingly, the citation pattern appears more closely associated with degree concentration than with the homophily attributes included in the model. This result concerns structural associations within the static network and does not identify temporal citation-formation mechanisms.

The analysis also indicates that citation counts are highly uneven and concentrated around a small number of highly visible studies. Although the upper tail of the distribution is compatible with a heavy-tailed citation structure, the power-law model did not significantly outperform alternative heavy-tailed distributions. Therefore, the findings should be interpreted as evidence of citation inequality, visibility-based concentration, and upper-tail skewness rather than as conclusive evidence of a power-law distribution or a temporal preferential-attachment mechanism.

In the primary negative binomial model, broader Google Scholar citation recognition was the only article-level characteristic significantly associated with within-corpus citation counts. Publication year, regional affiliation, indexation, and

research design were not statistically significant in that specification. When Google Scholar citations were excluded, the coefficients for publication year, the Others indexation category, literature review design, and qualitative design changed in magnitude or statistical significance. The article-level findings beyond broader citation recognition are therefore sensitive to model specification. Moreover, because Google Scholar citations and within-corpus indegree are related citation-based measures, their association should be interpreted as alignment between broader academic recognition and centrality within the corpus rather than as evidence that visibility independently causes subsequent citation accumulation.

In summary, the findings indicate that, within the AI literature in school leadership and management, studies that do not cite the highly cited publications located at the center of the network tend to remain peripheral and attract relatively lower citations. Moreover, the absence of studies functioning as bridges between peripheral and central studies suggests a weakly connected knowledge structure. This pattern indicates that the field is characterized simultaneously by citation concentration, limited diffusion, and insufficient connective capacity across different segments of the literature. In this context, it seems critical for future studies to focus on integrative and bridging research that connects peripheral studies to the center, rather than simply reproducing the core literature, to transform the field into a more balanced and inclusive knowledge network.

7 Limitations and future directions

This study has some limitations. First, since the analysis was limited to the school leadership and management corpus that was created and met specific inclusion criteria, sources that are not systematically represented in books, policy reports, and citation databases, as well as the use of AI in education in general, could not be included in the network structure. Secondly, the citation network was treated as static, and the evolution of citation relationships over time was not examined using dynamic network models. Thirdly, the variables used in the ERGM and negative binomial models are limited to specific structural and bibliometric factors; alternative mechanisms, such as author collaboration, institutional networks, and content similarity, are not included in the models. Future studies are recommended to enhance these findings by incorporating dynamic ERGMs, multilayer network analyses, and content-based clustering approaches, and by conducting robustness tests across different data sources. Another limitation of this study is that the themes, attributes, and references identified in the homophily and negative binomial tests were based on prior literature. It is suggested that future studies adopt diverse perspectives and hypotheses to identify themes, attributes, and references.

A related limitation concerns the relatively small size of the corpus and its implications for the statistical analyses. A network of 105 articles is not inherently too small for ERGM estimation, because parameter adequacy depends on network structure, model specification, parameter identifiability, and diagnostic performance rather than on a universal minimum node threshold. Nevertheless, the limited corpus may reduce the

precision of estimates, particularly for infrequent article attributes and sparsely represented categories. Accordingly, statistically nonsignificant ERGM and negative binomial coefficients should not be interpreted as definitive evidence that the corresponding relationships are absent. The power-law findings require particular caution because only 13 observations were located in the fitted upper tail. This limited tail size reduces the discriminatory power of goodness-of-fit comparisons and increases uncertainty around the estimated scaling parameter. The power-law analysis should therefore be regarded as a diagnostic assessment of citation concentration rather than conclusive evidence of a scale-free distribution or preferential-attachment process. Larger corpora and longitudinal citation data would permit more precise parameter estimation and stronger comparisons among competing distributional and network models. Although the title-based search strategy increased topical precision, it may also have narrowed the scope of the corpus. Because terminology in this emerging field remains unstable, some relevant studies may refer to machine learning, generative AI, ChatGPT, algorithmic decision-making, intelligent systems, data-driven leadership, or intelligent automation without using “Artificial Intelligence” or “AI” in the title. Therefore, the corpus should be interpreted as a focused citation-network representation of studies that explicitly foreground AI and school leadership/management, rather than as an exhaustive map of all possible AI-related leadership studies. Another limitation concerns the potential influence of author-group self-citations on the observed citation-network structure. The present study did not separately identify, remove, or test the effect of self-citation ties. Therefore, if a substantial number of citation ties involve studies authored or co-authored by the same research group, this may have contributed to the centrality of particular nodes and to the observed concentration of citations around the core of the network. Future studies should examine self-citation patterns more explicitly, compare citation-network structures with and without self-citations, and assess whether the removal of such ties changes centrality rankings, core-periphery patterns, or ERGM parameter estimates.

The substantial share of review-type studies within the corpus presents an additional interpretive constraint. Since the literature on AI in school leadership and management is still emerging, review articles, syntheses, and bibliometric studies may occupy disproportionately central positions by serving as early organizing references. This may partly intensify citation concentration and should be considered when interpreting the observed core-periphery structure and cumulative visibility effects.

A further limitation concerns the temporal structure of citation networks. Citation networks are temporally constrained: later-published articles can cite earlier-published articles, whereas earlier-published articles cannot cite later ones. Therefore, not all ordered dyads among the articles are equally at risk of forming citation ties. The present analysis does not restrict the ERGM risk set to temporally feasible dyads and does not employ a temporal ERGM, citation-age model, or dynamic network framework. Consequently, density values and ERGM coefficients should be interpreted as descriptive indicators of the observed static citation structure rather than as estimates of the full generative process of citation formation. In addition, the geometrically weighted degree terms were estimated using a fixed decay value of 0.5, and

alternative decay values were not examined. The degree-related findings should therefore be interpreted as conditional on this specification. Future studies should restrict the citation risk set to temporally feasible dyads, incorporate citation age or publication timing into the model, compare static and temporal network specifications, and evaluate the sensitivity of degree-related findings to alternative decay values.

The variation in publication timing across the corpus also introduces an important constraint for interpreting citation exposure and the article-level count model. Because the corpus includes articles published in different years, newer publications had less time to receive citations within the network. Although publication year was included in the negative binomial model and mean-centered for interpretability, the model does not fully adjust for citation exposure time through a longitudinal citation-age structure. Moreover, the negative binomial regression treats article-level citation counts as count outcomes, although these counts are derived from the same interdependent citation network. Therefore, the model should be interpreted as a supplementary associational analysis rather than as a network-aware or causal model of citation accumulation. Future studies should use longitudinal citation data, exposure offsets, age-adjusted citation indicators, hurdle or zero-inflated specifications where theoretically justified, and network-aware count models to examine citation accumulation more directly.

A further limitation concerns the use of Google Scholar citation count as an explanatory variable for within-corpus indegree. Google Scholar counts capture citations from a substantially broader scholarly literature and are therefore not identical to citations received from the other 104 articles in the constructed corpus. Nevertheless, because both measures are citation-based, some overlap may remain, and Google Scholar citation counts were not measured at a temporally prior point. The sensitivity analysis excluding this variable produced notable changes in the direction or statistical significance of several remaining coefficients. This does not demonstrate circularity, but it indicates that the article-level estimates are sensitive to model specification. Accordingly, the Google Scholar coefficient should be interpreted as cross-sectional correspondence between broader citation recognition and within-corpus centrality rather than as a causal visibility effect. Future studies should use temporally prior or more independent visibility indicators, longitudinal citation data, author- or institution-level measures, publication-venue characteristics, and age-adjusted citation indicators.

Data availability statement

The final list of included studies, data-extraction framework and variable definitions, citation matrix or edge list, article-level analytical variables, and R scripts used in the analyses are available from the corresponding author upon reasonable request.

Author contributions

TK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation,

Visualization, Writing – original draft, Writing – review & editing. OK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. GK: Data curation, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. RY: Data curation, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. IK: Conceptualization, Investigation, Resources, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

References

- Abdelazim, A., Al Breiki, M., and Khlaif, Z. N. (2025). AI In education: the mediating role of perceived trust in adoption decisions of school leaders. *Educ. Inf. Technol.* 30 (15), 20943–20975. doi: 10.1007/s10639-025-13596-4
- Adams, D., and Thompson, P. (2025). Transforming school leadership with artificial intelligence: applications, implications, and future directions. *Leadersh. Policy. Sch.* 24 (1), 77–89. doi: 10.1080/15700763.2024.2411295
- Aldighrir, W. M. (2024). Impact of AI ethics on school administrators' decision-making: the role of sustainable leadership behaviors and diversity management skills. *Curr. Psychol.* 43 (41), 32451–32469. doi: 10.1007/s12144-024-06862-0
- Anysiadou, M., and Gkiliati, A. (2025). An experimental analysis of Greek school leaders' readiness for AI integration. *Leadersh. Policy. Sch.* 24 (1), 155–177. doi: 10.1080/15700763.2024.2439833
- Arar, K., Tlili, A., and Salha, S. (2024). Human-machine symbiosis in educational leadership in the era of artificial intelligence (AI): where are we heading? *Educ. Manag. Adm. Leadersh.* 54: 17411432251292295. doi: 10.1177/17411432251292295
- Arar, K., Tlili, A., Schunka, L., Salha, S., and Saiti, A. (2025). Reimagining educational leadership and management through artificial intelligence: an integrative systematic review. *Leadersh. Policy. Sch.* 24, 1–23. doi: 10.1080/15700763.2025.2451982
- Bhaskar, A., Thennarasu, K., Philip, M., and Jaisooray, T. S. (2023). Regression models for count data with excess zeros: a comparison using survey data. *Quant. Methods Psychol.* 19 (1), 1–13. doi: 10.20982/tqmp.19.1.p001
- Bellibaş, MŞ, Özgenel, M., and Bozkurt, A. (2025). Principals' artificial intelligence literacy and leadership self-efficacy: the technology acceptance model. *Asia-Pac. Educ. Res.* 35 (3), 631–641. doi: 10.1007/s40299-025-01059-w
- Berkat, B., Alexandro, R., and Basrowi, B. (2024). Utilization of big data and artificial intelligence on quality education management and its implications on school sustainability. *Int. J. Data Netw. Sci.* 8 (3), 1895–1906. doi: 10.5267/j.ijdns.2024.1.023
- Berkovich, I. (2025). The rise of AI-assisted instructional leadership: empirical survey of generative AI integration in school leadership and management work. *Front. Educ.* 10, 1643023. doi: 10.3389/feduc.2025.1643023
- Berkovich, I., and Eyal, O. (2025). Support for generative artificial intelligence as a predictor of middle leaders' generative artificial intelligence self-efficacy, valuing, and integration in school leadership work. *Educ. Manag. Adm. Leadersh.*, 17411432251361251. doi: 10.1177/17411432251361251
- Bixler, K., and Ceballos, M. (2025). Principals leading AI in schools for instructional leadership: a conceptual model for principal AI use. *Leadersh. Policy. Sch.* 24 (1), 137–154. doi: 10.1080/15700763.2024.2428297
- Borgatti, S. P., Agneessens, F., Johnson, J. C., and Everett, M. G. (2024). Analyzing social networks. Available online at: <https://www.torrossa.com/gs/resourceProxy?an=5730558&publisher=FZ7200> (Accessed April 02, 2026).
- Brass, D. J., and Borgatti, S. P. (2019). "A brief primer on social network analysis," in *Social Networks at Work*, eds. D. J. Brass and S. P. Borgatti (New York, NY: Routledge), 1–8.
- Burnham, K. P., and Anderson, D. R. (2004). "Advanced issues and deeper insights," in *Model Selection and Multimodel Inference* (New York: Springer), 267–351. doi: 10.1007/978-0-387-22456-5_6
- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* 51 (4), 661–703. doi: 10.1137/070710111
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., and Herrera, F. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *J. Am. Soc. Inf. Sci. Technol.* 62 (7), 1382–1402. doi: 10.1002/asi.21525
- Dai, R., Thomas, M. K. E., and Rawolle, S. (2025). The roles of AI and educational leaders in AI-assisted administrative decision-making: a proposed framework for symbiotic collaboration. *Aust. Educ. Res.* 52 (2), 1471–1487. doi: 10.1007/s13384-024-00771-8
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Soc. Networks.* 1 (3), 215–239. doi: 10.1016/0378-8733(78)90021-7
- Fullan, M., Azorin, C., Harris, A., and Jones, M. (2024). Artificial intelligence and school leadership: challenges, opportunities and implications. *Sch. Lead. Manag.* 44 (4), 339–346. doi: 10.1080/13632434.2023.2246856
- Gârdan, I. P., Manu, M. B., Gârdan, D. A., Negoită, L. D. L., Paștiu, C. A., Ghiță, E., et al. (2025). Adopting AI in education: optimizing human resource management

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2026.1877606/full#supplementary-material>

- considering teacher perceptions. *Front. Educ.* 10, 1488147. doi: 10.3389/feduc.2025.1488147
- Ghamrawi, N., Shal, T., and Ghamrawi, N. A. R. (2024). Exploring the impact of AI on teacher leadership: regressing or expanding? *Educ. Inf. Technol.* 29 (7), 8415–8433. doi: 10.1007/s10639-023-12174-w
- Göçen, A., and Döğler, M. F. (2025). A global perspective on artificial intelligence in educational leadership. *J. Educ. Res.* 118 (6), 752–770. doi: 10.1080/00220671.2025.2510397
- Half, H. M. (1986). Instructional applications of artificial intelligence. *Educ. Leadersh.* 43 (6), 24–31.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge: Cambridge University Press.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). Ergm: a package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* 24 (3), nihpa54860. doi: 10.18637/jss.v024.i03
- Igbokwe, I. C. (2024). Artificial intelligence in educational leadership: risks and responsibilities. *Eur. J. Arts Humanit. Soc. Sci.* 1 (6), 3–10. doi: 10.59324/ejass.2024.1(6).01
- Kafa, A. (2025a). Exploring integration aspects of school leadership in the context of digitalization and artificial intelligence. *Int. J. Educ. Manag.* 39 (8), 98–115. doi: 10.1108/IJEM-11-2024-0703
- Kafa, A. (2025b). Exploring school leaders' perceptions of using an artificial intelligence tool for management tasks: empirical evidence from Cyprus. *Manag. Educ.*, 08920206251394137. doi: 10.1177/08920206251394137
- Karaköse, T. (2024). Will artificial intelligence (AI) make the school principal redundant? A preliminary discussion and future prospects. *Educ. Process: Int. J.* 13 (2), 7–14. doi: 10.22521/edupij.2024.132.1
- Karaköse, T., Demirkol, M., Yirci, R., Polat, H., Özdemir, T. Y., and Tülübaşı, T. (2023). A conversation with ChatGPT about digital leadership and technology integration: comparative analysis based on human–AI collaboration. *Adm. Sci.* 13 (7), 157. doi: 10.3390/admsci13070157
- Karaköse, T., and Tülübaşı, T. (2024). School leadership and management in the age of artificial intelligence (AI): recent developments and future prospects. *Educ. Process: Int. J.* 13 (1), 7–14. doi: 10.22521/edupij.2024.131.1
- Karaköse, T., and Tulubası, T. (2025). The role of educational leaders in the age of artificial intelligence (AI). *Educ. Process: Int. J.* 16, e2025267. doi: 10.22521/edupij.2025.16.267
- Kesim, E. (2025). Transformation of school leadership in the AI era: requisite leadership roles of teachers. *J. Qual. Res. Educ.* 44, 188–211. doi: 10.14689/enad.44.2261
- Kesim, E., Atmaca, T., and Turan, S. (2025). Reshaping school cultures: AI's influence on organizational dynamics and leadership behaviors. *Leadersh. Policy. Sch.* 24 (1), 117–136. doi: 10.1080/15700763.2024.2413703
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *Am. Doc.* 14 (1), 10–25. doi: 10.1002/asi.5090140103
- Kılınç, AÇ, Özdemir, S., Sezgin, F., Savaş, G., and Dolapçı, E. (2025). Are Turkish school principals ready to use artificial intelligence? Exploring the roles of growth mindset, self-efficacy and openness to change. *Educ. Manag. Adm. Leadersh.*, 17411432251393449. doi: 10.1177/17411432251393449
- Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. Boston, MA: McGraw, Irwin.
- Marrone, R., Fowler, S., Bathakur, A., Dawson, S., Siemens, G., and Singh, C. (2025). Perceptions and perspectives of Australian school leaders on the integration of artificial intelligence in schools. *Sch. Lead. Manag.* 45 (1), 30–52. doi: 10.1080/13632434.2024.2425019
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* 27 (1), 415–444. doi: 10.1146/annurev.soc.27.1.415
- Nelson, W. A. (1989). Artificial intelligence knowledge acquisition techniques for instructional development. *Educ. Technol. Res. Dev.* 37 (3), 81–94. doi: 10.1007/BF02299060
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U.S.A.* 98 (2), 404–409. doi: 10.1073/pnas.98.2.404
- Polat, M., Karataş, İH, and Varol, N. (2025). Ethical artificial intelligence (AI) in educational leadership: literature review and bibliometric analysis. *Leadersh. Policy. Sch.* 24 (1), 46–76. doi: 10.1080/15700763.2024.2412204
- Richardson, J. W., Vedder, B. C., Roberts, A. B., and McLeod, S. (2025). What's the chatter about AI and school leaders? *Leadersh. Policy. Sch.* 24 (1), 103–116. doi: 10.1080/15700763.2024.2409307
- Robins, G. (2011). “Exponential random graph models for social networks,” in *The SAGE Handbook of Social Network Analysis*, eds. J. Scott and P. J. Carrington (London: SAGE Publications), 484–500.
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike information criterion statistics. *D. Reidel*, 81(10.5555), 26853.
- Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* 24 (4), 265–269. doi: 10.1002/asi.4630240406
- Sposato, M. (2025). Artificial intelligence in educational leadership: a comprehensive taxonomy and future directions. *Int. J. Educ. Technol. High. Educ.* 22 (1), 20. doi: 10.1186/s41239-025-00517-1
- Tyson, M. M., and Sauers, N. J. (2021). School leaders' adoption and implementation of artificial intelligence. *J. Educ. Adm.* 59 (3), 271–285. doi: 10.1108/JEA-10-2020-0221
- Wang, Y. (2021a). Artificial intelligence in educational leadership: a symbiotic role of human-artificial intelligence decision-making. *J. Educ. Adm.* 59 (3), 256–270. doi: 10.1108/JEA-10-2020-0216
- Wang, Y. (2021b). When artificial intelligence meets educational leaders' data-informed decision-making: a cautionary tale. *Stud. Educ. Eval.* 69, 100872. doi: 10.1016/j.stueduc.2020.100872
- Wasserman, S., and Faust, K. (1994). *Social network analysis: Methods and applications*.