IJWR

A Network Analysis of Retracted Citations by Iranian Computer Scientists

Parsa Zahedi*, Sadegh Sulaimany

Social & Biological Network Analysis Laboratory (SBNA), Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran; Parsa.Zahedi@UoK.ac.ir, S.Sulaimany@UoK.ac.ir

ABSTRACT

Retracted publications continue to influence scholarship long after withdrawal. This study assembles a curated set of 169 Iran-affiliated retractions in computer science, data science, and electrical engineering from 2008 to 2024, links them to citing and cited works through two complementary retrieval pipelines, and constructs an expanded citation network of 1'694 nodes and 1,703 edges. We quantify retraction reasons and timing, community structure, node centrality, self-citation patterns, author and institutional concentration, international co-authorship, and a field-adjusted national benchmark. Misconduct-related causes predominate. The average interval from publication to retraction increased into 2021 and has since begun to shorten. The citation network exhibits strong community structure with three major thematic clusters. Centrality profiling isolates five retracted works that function as hubs, often reinforced by self-citation loops. Contribution is highly concentrated among a small set of authors and institutions, while collaboration extends across multiple regions beyond Iran. A field-adjusted retraction rate places the national record among mid-tier producers. These results identify practical leverage points to reduce downstream spread of invalidated findings: persistent indexing flags on hub retractions, routine screening of citations to retracted work, and focused attention on repeat patterns in self-citation and institutional clusters. The study offers a reproducible dual-pipeline approach, a full centrality profile of an enlarged network, and actor-level diagnostics that support targeted integrity interventions.

Keywords— ***Retraction, Computer Science, Iran, Citation Analysis.***

1. Introduction

The credibility of the scholarly record is a prerequisite for cumulative scientific progress, yet the growing incidence of article retractions threatens this foundation. Papers may be withdrawn for infractions ranging from plagiarism and data falsification to manipulated peer review, but their epistemic footprint often endures: retracted works continue to be cited as legitimate evidence, thereby propagating misinformation and eroding confidence in subsequent research as well as in academic norms [1], [2]. Recent monitoring recorded more than 10,000 retractions in 2023, a historic high, underscoring the timeliness of mapping how such items continue to circulate through citations [3]. Macro-level surveys have sketched the broad contours of retraction activity, yet there remains a conspicuous deficit of context-sensitive analyses. The corpus authored by Iranian computer-science scholars exemplifies this gap, as it is subject to a distinctive mixture of integrity pressures. Recent bibliometric audits record an appreciable uptick in Iranian retractions, with plagiarism and fabricated peer-review reports constituting the principal triggers [4]. Systematically charting the citation networks that surround these withdrawals is therefore imperative for tracing the diffusion of unreliable knowledge, isolating recurrent points of failure, and informing evidence-based safeguards that uphold the credibility of the scientific record.

Although the corpus on research misconduct has grown appreciably, it remains disproportionately centered on biomedicine and the life sciences [5]. By contrast, the computer-science domain, especially its Iranian subset, has attracted only sporadic investigation. Existing studies have typically catalogued the proximate grounds for withdrawal, yet have seldom interrogated the citation-network pathways through which rescinded knowledge continues to circulate. Elucidating these pathways is essential for diagnosing systemic vulnerabilities in

doi http://dx.doi.org/10.22133/ijwr.2025.525538.1287

Citation P. Zahedi, S. Sulaimany, "A Network Analysis of Retracted Citations by Iranian Computer Scientists", *International Journal of Web Research*, vol.8, no.4,pp.67-83, 2025, doi: http://dx.doi.org/10.22133/ijwr.2025.525538.1287.

*Corressponding Author

Article History: Received: 23 May 2025; Revised: 14 September 2025; Accepted: 28 September 2025.

scholarly communication and for crafting evidence-based safeguards. To redress this gap, the present work offers a granular analysis of the citation topology surrounding retracted publications by Iranian computer scientists, with particular emphasis on self-citation behavior and the temporal resilience of references to invalidated articles.

This study contributes such a lens by assembling a rigorously filtered corpus of 169 Iran-affiliated retractions in computer science, data science, and electrical engineering (2008-2024) and by introducing a dual-pipeline retrieval strategy that fuses VOSviewer's OpenAlex integration with PyAlex queries to recover citation relations at scale. VOSviewer retrieves 162/169 retractions and their links, while PyAlex recovers 166/169 and pulls in additional (non-retracted) neighbors, yielding an enlarged network of 1'694 nodes and 1,703 edges for analysis. This design enables both cross-validation and broader coverage for downstream measures.

Using this network, we quantify how withdrawn knowledge persists through structure rather than isolated citations [6], [7]. Community detection shows three cohesive thematic clusters and an unusually high Louvain modularity ($Q \approx 0.95$), indicative of sharply segmented citation neighborhoods. Centrality profiling isolates five retracted papers that function as structural keystones, high in betweenness and/or eigenvector centrality, so that interventions targeting these hubs can disproportionately reduce downstream propagation [8], [9], [10], [11], [12]. We also document self-citation loops (17 authors) and show that authorship and institutional contributions are highly concentrated, features that align with the observed community structure. Together, these diagnostics move from description to actionable risk, pinpointing the nodes and actors most capable of transmitting invalidated results.

To place Iran's record in context, we compute a field-normalized retraction rate using number of published articles in country as denominators from SciMagoJR, positioning the national figure at $\sim 1.9 \times 10^{-4}$ (45th worldwide), mid-table relative to peers and major producers. This normalization clarifies interpretation by accounting for disciplinary mix and output volume

In sum, this manuscript advances three contributions. First, it offers a reproducible, dual-pipeline approach that enlarges and stabilizes the retraction-citation graph, enabling robust community and centrality analyses [8], [9], [10], [11], [12], [13], [14]. Second, it delivers actor-level diagnostics, self-citation patterns, institutional concentration, and collaboration profiles, that highlight where targeted integrity interventions can most effectively disrupt persistence of withdrawn claims. Third, it situates Iran's performance within a field-adjusted international benchmark, supporting fairer cross-country comparisons and policy discussion [15]. The remainder of the paper reviews related work, details data sources and procedures, presents empirical patterns, and discusses implications and limitations, closing with recommendations for practice and future inquiry.

2. Literature Review

Retractions are intended to cauterize error, yet the evidentiary after-life of withdrawn papers routinely outlives the notice. Large-scale longitudinal analyses in biomedicine show that retracted papers continue to accrue citations for years and that many citers show no awareness of the retraction status [16]. Meta-level evidence suggests retraction reduces citation frequency by only about 60% relative to comparable non-retracted papers, underscoring persistent post-retraction propagation [17]. Large-scale mapping shows retractions concentrate among highly cited authors and do not reliably halt downstream influence, especially when citations are embedded in dense topical communities [5], [6]. Beyond individual misconduct, the persistence of retracted claims is a systems problem: incentives, editorial workflows, and information-propagation dynamics interact to keep invalid results circulating [18], [19], [20], [21].

***Post-retraction Propagation*:** For high-profile retractions in top multidisciplinary journals, nearly half of total citations can occur after the retraction, reflecting visibility effects and critical commentary as additional drivers of propagation [1]. The problem extends to secondary research: retracted systematic reviews themselves continue to be cited, with limited post-retraction correction by citing authors [22]. Empirical examinations consistently find that retracted articles continue to be cited as if valid, with direct citations driving most error transmission and indirect ("downstream") citations contributing a long tail of persistence [6], [19]. Case-study analyses of citation contexts indicate that some retracted papers retain "epistemic contributions" for their communities, which helps explain the durability of their influence despite retraction [2]. While much of this literature is biomedically focused [7], the mechanism is domain-agnostic: once incorporated into local citation neighborhoods, retracted records behave like any other central node in a scholarly graph, unless indexing and editorial countermeasures intervene [20]. Recent editorial guidance outlines implementable countermeasures, requiring retraction checks at submission and in reference tools, and harmonizing retraction labels, to curb inadvertent propagation [23].

***Technology and Computer-science Contexts*:** Within computer science specifically, retractions are numerous and often poorly documented; a domain-level audit reports post-retraction citations and disparate publisher practices, amplifying contamination risks for reviews and mappings [24]. Regional analyses have begun to profile retractions and citation behavior across Iran's output, adding necessary context for country-level patterns examined here [25]. An earlier conference study scoped Iranian computer-science retractions using a single VOSviewer↔OpenAlex route and reported descriptive citation patterns [26]. Methodologically, it lacked a documented data-snapshot timestamp, did not analyze induced citation neighborhoods, did not report parameter choices for centrality/community detection, and did not include a pre-specified robustness/sensitivity plan. Related Iranian bibliometric work highlights discipline-specific vulnerabilities and the need for harmonized metadata to support integrity monitoring [4]. Comparative scientometric studies report notable retraction activity in technology-adjacent fields (computer science, telecommunications, materials), with plagiarism and duplication as recurrent triggers [27], [28]. Country-level audits identify Iran among higher-volume contributors to retractions in these domains, but emphasize that national counts must be interpreted against field-matched publication baselines [4], [29]. The result is a mixed picture: elevated absolute numbers in some settings, yet heterogeneous normalized rates once disciplinary output is considered [28], [29]. This argues for context-sensitive analyses that combine domain filters, author behavior, and network structure, precisely the vantage adopted here.

***Network-analytic Lens*:** Our focus on shortest paths, centrality, and community structure complements prior CS-specific audits by pinpointing the actors and ties most responsible for post-retraction spread [24]. Bibliometric mapping tools such as VOSviewer and Gephi [30] operationalize retraction studies as graphs, papers as nodes, citations as edges, enabling measurement of how withdrawn items persist through structure rather than anecdote [13]. Centrality statistics capture complementary roles: betweenness identifies brokers that connect otherwise separate communities (useful for tracing how a single retraction bridges topics) [8], [9], [14]; eigenvector weights ties by the prestige of neighbors (a conduit for legitimacy even with modest raw counts) [10], [11], [12], [14]. When these metrics are high for retracted items, corrective actions must prioritize those nodes because they amplify contamination of the literature.

***Authors, Institutions, and Incentives*:** Retractions reshape reputational trajectories for authors and their co-authors, with measurable influence penalties and institutional spillovers [5]. At the same time, a small number of prolific actors can dominate local landscapes; concentrated authorship and dense co-citation can inflate both community modularity and the apparent institutional footprint of retractions [4]. Editorial and peer-review vulnerabilities further enable problematic manuscripts to resurface or remain insufficiently flagged across venues, underscoring the need for workflow-level safeguards (e.g., automated flag propagation, stronger resubmission checks) [21], [20].

***Policy and Benchmarking*:** Because raw counts obscure specialization, field-normalized indicators are recommended for cross-national comparison. Using discipline-specific denominators (e.g., Scimago's citable items in computer/data/electrical engineering) yields more defensible rankings than all-fields baselines [28], [29], [15]. Global "misconduct maps" suggest that structural pressures, rapid expansion, incentive misalignment, and uneven integrity infrastructure, coincide with higher retraction prevalence, but with wide dispersion across countries and subfields [29]. Such heterogeneity motivates analyses that marry rate-based benchmarking with network positions of retracted nodes to prioritize interventions.

***Gaps This Study Addresses*:** Prior work in Iran largely catalogs causes, time lags, and venues for retractions [4]; global studies document persistence and editorial limits [5], [6], [19], [20]. What remains under-specified in the technology/computer-science context is the joint role of (i) structural position of retracted items in the citation graph, (ii) self-citation loops and author clusters, and (iii) field-matched, country-level benchmarking. By integrating VOSviewer/Gephi-based network diagnostics with discipline-filtered denominators and author-level patterns, the present analysis advances from description to targetable risk: it identifies which retracted nodes, authors, and institutions most effectively propagate unreliable knowledge, and therefore where post-retraction management can yield the largest reduction in misinformation pressure [18], [13], [8], [10], [9], [28], [29].

3. Data and Methods

***Design Overview*:** The workflow (i) matches Retraction Watch to OpenAlex via DOIs; (ii) retrieves citations through two independent routes, VOSviewer/OpenAlex and PyAlex; (iii) builds two directed graphs (retracted with retracted and retracted with non-retracted); (iv) computes field-normalized comparators; and (v) pre-specifies robustness and sensitivity checks.

On February 12th 2025 at 5:50 AM Tehran Standard Time, the dataset of retracted publications was acquired from the Retraction Watch GitLab repository[1]. Given the dynamic nature of this repository, specifying the data retrieval timestamp is

[1] available at https://gitlab.com/crossref/retraction-watch-data

imperative for ensuring reproducibility. The dataset comprises detailed records for more than 60’000 retracted articles, including the reasons underlying each retraction. The Retraction Watch database organizes these retractions into seven primary disciplinary domains, Business and Technology (B/T), Basic Life Sciences (BLS), Environmental Sciences (ENV), Health Sciences (HSCs), Humanities (HUM), Physical Sciences (PHY), and Social Sciences (SOC), with each domain further subdivided into specialized research fields.

For the purposes of this study, a structured filtering procedure was applied. Initially, only papers with Iran listed as the country of affiliation were retained. Thereafter, the selection process was refined to encompass only those articles categorized under Computer Science, Data Science, and Electrical Engineering, which correspond to the (B/T) classifications for Computer Science and Data Science, as well as the (PHY) category for Electrical Engineering, as defined in the Retraction Watch schema.

Following the implementation of the specified filtering criteria, the initial dataset was meticulously refined to a subset of 169 retracted publications, hereafter referred to as the “Filtered Retraction Watch Dataset” (Figure 1). Field names and reason codes follow the Retraction Watch Database User Guide, enabling reproducible filtering and grouping of retraction causes in our analyses [31].

This subset encompasses retractions occurring between 2008 and 2024, while the original publication dates of these studies span from 2007 to 2023. Such a longitudinal framework not only captures the historical trajectory of the articles from their initial dissemination but also offers a noteworthy insight into the interval preceding their subsequent retraction.

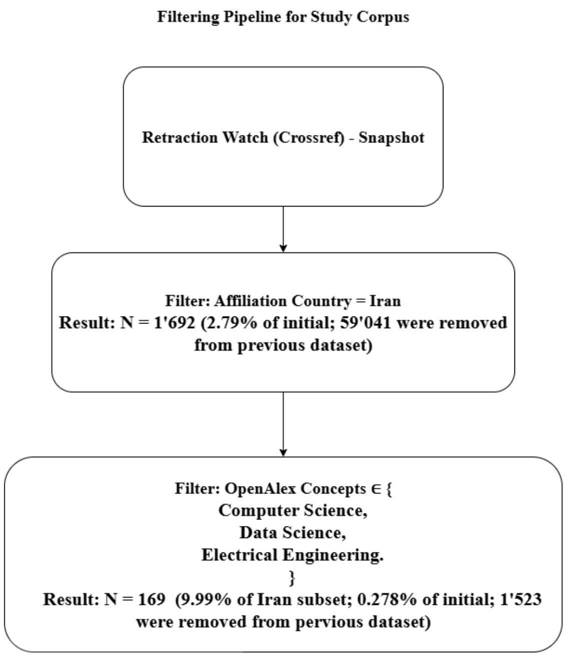

Figure. 1. Flowchart shows how Retraction Watch records were filtered by (i) Iran affiliation and (ii) CS/DS/EE subject tags to yield the 169-paper study corpus (2008–2024).

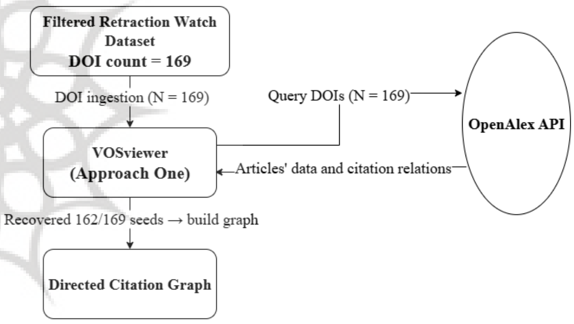

Figure. 2. Workflow illustrating DOI ingestion into VOSviewer to pull OpenAlex links and build a directed citation graph; this route successfully recovered 162 of the 169 seed retractions for network analysis.

3.1. Citation Data Retrieval

In light of the Retraction Watch dataset's inherent lack of explicit citation metadata, two distinct methodological approaches were deployed to harvest citation data from OpenAlex, each designed to serve particular analytical aims.

a) Approach One: Integration of the VOSviewer API with OpenAlex

Initially, the digital object identifiers (DOIs) corresponding to the 169 articles in the filtered dataset were submitted to the VOSviewer program. This process facilitated the extraction of citation relationships from OpenAlex and culminated in the construction of a citation network, wherein each node signifies an individual article and each edge delineates a citation link (see Figure 2). It is noteworthy that this method successfully retrieved data for 162 out of the 169 articles, thereby reflecting a substantial coverage for subsequent analyses.

b) Approach Two: Utilization of the Pyalex Library in Python

A complementary strategy was implemented using the Pyalex library, which provided an alternative conduit for obtaining citation data from OpenAlex. Citation and entity metadata were retrieved from OpenAlex (Works, Authors, Sources) via its public API, which models the scholarly graph as a heterogeneous directed network [32]. Implementation relied on the PyAlex Python client for OpenAlex to programmatically fetch citing and referenced works at scale [33]. This method resulted in the retrieval of information for 166 out of 169 articles, albeit with the detection of four duplicate DOIs, each associated with different retraction reasons. Beyond mere data acquisition, this approach facilitated the delineation of self-citation dynamics and the pinpointing of the most cited retracted articles. Additionally, Pyalex extended the analytical scope by incorporating further articles cited by the retracted papers, which were not themselves retracted, thereby yielding an enriched citation network comprising 1’694 distinct nodes.

This dual-method strategy not only reinforces the robustness of the citation analysis but also enhances our understanding of the citation dynamics surrounding retracted scholarly work.

4. Results and Discussions

Drawing upon records from the Retraction Watch database, a cumulative total of 169 retracted publications were identified in Iran for studies in Computer Science, Data Science, and Electrical Engineering over the period from 2008 to 2024. Figure 3 provides a detailed annual breakdown of these retractions. In particular, the year 2020 is notable for recording 24 cases, while the retraction counts reported for 2022 and 2023 differ markedly.

Nineteen distinct publishers appear in the dataset. Notably, Springer is responsible for the largest share of retractions, contributing 73 cases (43.2% of the total). Following Springer, the Institute of Electrical and Electronics Engineers (IEEE) and Elsevier account for 28 (16.6%) and 25 (14.8%) retractions, respectively. Figure 4 offers a detailed visual representation of this distribution.

4.1. Time Lag

The retraction delay, defined as the period between an article’s publication and its subsequent retraction, quantifies the time required for the scholarly community to detect and correct flawed research. Extended delays imply that erroneous or misleading findings may continue to circulate in the literature, thereby potentially skewing subsequent research and compromising the integrity of scientific discourse. Additionally, prolonged intervals before retraction can adversely affect the reputations of academic journals, highlighting the inherent challenges of upholding rigorous scholarly standards. Changes in these delays may further offer insights into evolving levels of scrutiny and the overall efficiency of the retraction process [4], [18]. In Figure 5, this metric is elucidated by examining, for each calendar year, the set of articles that were retracted during that period. For every such article, the time lag is computed as the difference between its actual publication date and the date of its retraction. These individual delays are then averaged to produce a representative value for that year. According to the figure, a notable escalation in the average retraction delay is observed after 2016, reaching 7.16 years in 2021, followed by a reduction in subsequent periods.

4.2. Retraction Reasons

A snapshot of the Retraction Watch repository on 12 February 2025 recorded 107 distinct grounds for withdrawal, with individual records frequently

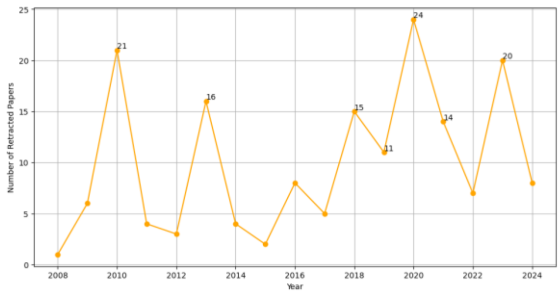

Figure. 3. Annual retraction counts for CS/DS/EE Iran-affiliated papers, peaking around 2020 (24 cases) and showing marked year-to-year variation across 2022–2023.

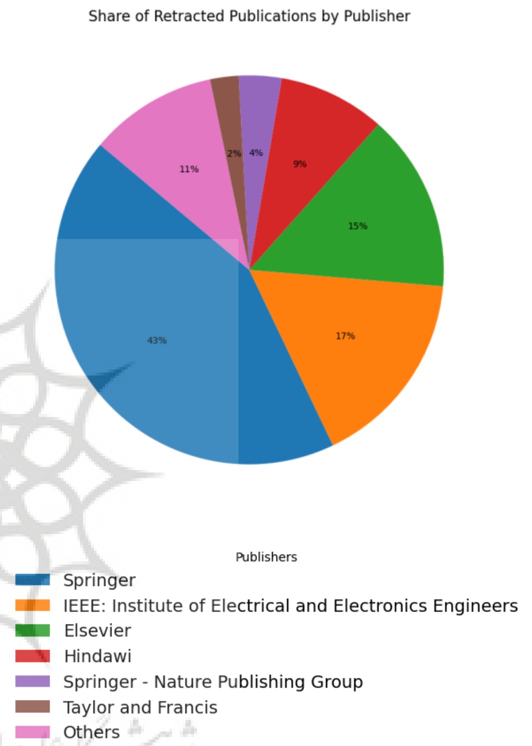

Figure. 4. Publisher distribution is highly concentrated: Springer accounts for 73 retractions (43.2%), followed by IEEE (28; 16.6%) and Elsevier (25; 14.8%).

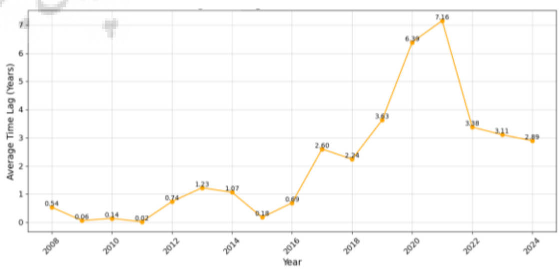

Figure. 5. Mean publication-to-retraction delay rises (Time lag), reaching 7.16 years in 2021, then declines

bearing multiple tags. Within the restricted corpus analyzed here, only 51 of those descriptors were observed. For analytical coherence, the 51 items were consolidated into three overarching classes, (i) deliberate misconduct, (ii) honest error or oversight, and (iii) administrative or procedural irregularities. This higher-level taxonomy sharpens interpretation of the proximate drivers of retraction. The annual prevalence of each class is plotted in Figure 6, revealing temporal shifts in the relative weight of the three categories.

Misconduct being dominant is in line with previous research in this area [18].

However, Figure 7 presents the retraction reasons over time as stated in the Retraction Watch dataset. For readability, the legend lists the eighteen most common retraction reasons individually, while all remaining reasons are consolidated into an "Others" category for visualization.

4.3. Citation Network Analysis of Retracted Papers

Intellectual connectivity among the withdrawn literature was subsequently interrogated with VOSviewer. Leveraging the software's API, we retrieved 162 records and mapped a directed citation graph in which vertices denote articles and edges represent citation links (Figure 8). Vertex size is scaled to citation in-degree, such that highly referenced papers are rendered more prominent. Community-detection algorithms resolve three cohesive clusters, whereas the remaining vertices persist as singletons devoid of inter-article ties, suggesting limited cross-referencing outside the core triad.

As of 12 February 2025, the 166 retracted publications harvested via the PyAlex interface to OpenAlex have amassed a cumulative 569 citations. The most frequently referenced title in this cohort, "Time-dependent personal tour planning and scheduling in metropolises," co-authored by researchers bearing OpenAlex identifiers A5028765261 and A5062957503, has garnered 118 citations, the most recent of which was registered during the preparation of the present manuscript, despite the paper's formal withdrawal on 12 April 2011, as documented by the Retraction Watch database.

Neither contributor holds an academic qualification in computer engineering based on their profile in OpenAlex; nevertheless, the article engages a topic situated at the intersection of computer science and image processing. In line with ethical guidelines and privacy considerations, authors are identified herein exclusively by their OpenAlex IDs.

PyAlex-based analytics indicate that self-referential citation is a recurrent feature of the Iranian retraction corpus. **Seventeen** distinct authors cite their own withdrawn work, some on multiple occasions. The citation topology rendered in VOSviewer (Figure 8) makes this pattern explicit. One contributor (OpenAlex ID A5058384543) has six retracted articles that mutually reference one another, yielding nine self-citations (Figure 9).

A frequent collaborator (ID A5053388446) replicates this pattern, likewise accruing nine self-citations within the very same set of six papers. Collectively, this small cluster has nonetheless attracted forty-eight citations from external authors, underscoring how retracted findings can remain embedded in the scholarly record.

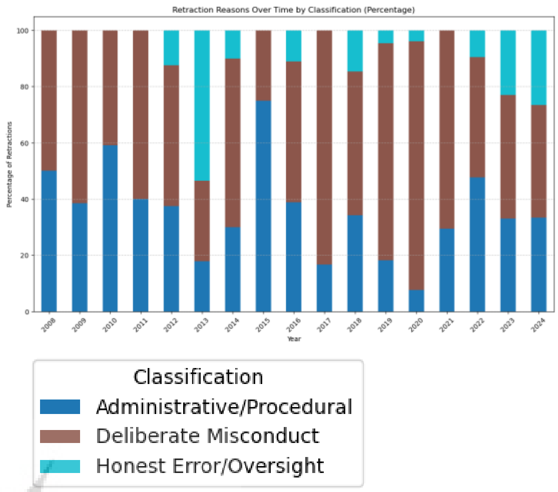

Figure. 6. Retraction reasons aggregated into three classes (misconduct, honest error, administrative) show shifting shares over time, with misconduct generally dominant.

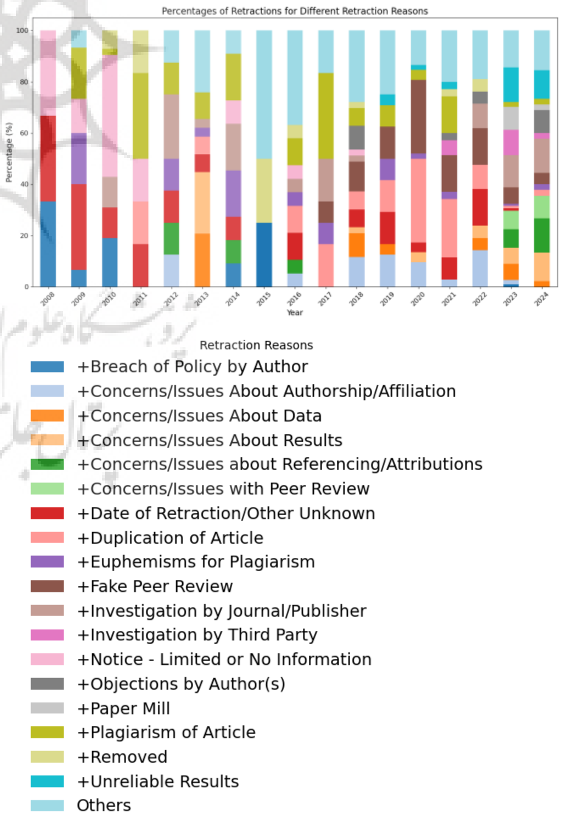

Figure. 7. Temporal profile of specific Retraction Watch reasons: the 18 most common are itemized in the legend, while the long-tail is combined as "Others" to maintain readability.

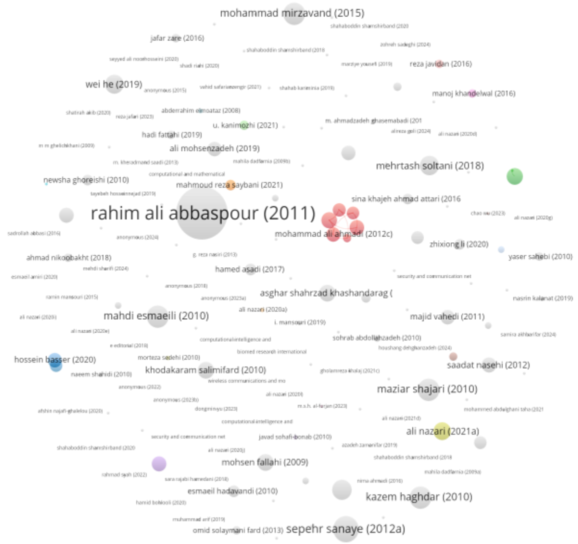

Figure. 8. Citation network of retracted items (VOSviewer/OpenAlex) forms three cohesive clusters with many singletons; node size scales with in-degree, highlighting clustered influence and weak cross-links.

Figure. 9. Largest self-citation subgraph: six retracted papers by a single author inter-cite to generate nine self-citations, a pattern mirrored by a frequent collaborator and still attracting dozens of external citations.

4.4. Centrality of Pivotal Retracted Works

Centrality statistics were generated in Gephi 0.10.1 from an edge list obtained with the PyAlex wrapper for the OpenAlex API. The crawl (April 2025) returned 162 of the 169 retracted papers in our seed list and every record that either cites or is cited by them, giving a network of 1’694 articles linked by 1703 citation arcs.

Centrality scores were computed as follows: Betweenness centrality uses the Equation (1) and for Eigenvector centrality Equation (2) was used.

Betweenness Centrality *[8]:*

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \qquad (1)$$

Where σ_{st} is the number of shortest paths between nodes s and t and $\sigma_{st}(v)$ is the subset that pass through v.

Eigenvector centrality *[10], [11], [12], [14]:*

$$A\,\mathbf{x} = \lambda_{\max}\,\mathbf{x}.\text{so}C_E(i) = \frac{1}{\lambda_{\max}} \sum_j A_{ij}\, C_E(j). \qquad (2)$$

Where A is the adjacency matrix and $\lambda_{\max}$ its largest eigenvalue.

Betweenness pinpoints papers that control the flow of information between otherwise weakly connected parts of the graph; nodes with high betweenness act as structural brokers and can therefore accelerate, or block, the spread of flawed findings [8], [9], [11], [12]. Eigenvector centrality, by contrast, rewards nodes that are cited by already influential papers; a high-eigenvector retraction can contaminate the most authoritative core of the literature even if its own citation count is modest [10], [11], [12], [14]. Using the two measures together thus captures both bridging influence and prestige-based influence, giving a fuller picture of how retracted work persists in scholarly discourse. (Closeness centrality was recorded but is not discussed further because all five keystone nodes tie for the maximum value, offering no additional discrimination.)

Applying these measures to all 1’694 nodes singles out five retracted items (Figure 10), OpenAlex IDs W2029538206, W3120590686, W2006962765, W3027947458 and W2077146256, as the network’s structural keystones.

W2029538206 emerges as the network’s chief broker. Its betweenness of 65.83 shows that more shortest citation paths run through it than any other paper, while its eigenvector score of 0.04018 confirms strong ties to the field’s most influential works. A very low clustering coefficient (0.032) suggests that it bridges otherwise unconnected neighborhoods rather than sitting inside a tight clique. The infamous co-authors A5053388446 and A5058384543, which you have seen in the self-citation section, are the authors for this article. Both of these authors have 6 retracted papers in the Retraction Watch dataset and in our filtered dataset. W3120590686 acts as a gatekeeper. With a betweenness of 46.00 it still intercepts many citation routes, yet its much smaller eigenvector value (0.01339) and almost-zero clustering coefficient (0.0024) indicate a peripheral position that links loosely connected groups rather than anchoring the core itself. Interestingly this article was published in 2013 and was receiving citations up until 2020 and has received a total of 12 citations. The main author of this article, A5029406867, has 12 retracted articles in all the Retraction Watch dataset, and has 4 retracted articles in our filtered dataset. W2006962765 presents a more balanced hub profile. Its betweenness (37.33) is lower than the two brokers above, but it shares their high eigenvector score (0.04018) and shows the second-largest clustering coefficient (0.057), implying that it sits inside a moderately cohesive sub-community that is itself well connected to the network core. W3027947458 is the network's authority magnet. Although its betweenness (35.00) is slightly below that of W2006962765, its eigenvector centrality leaps an order of magnitude to 0.4933, signaling that it is cited by, and tightly embedded among, the most authoritative papers in the field. Its low clustering coefficient (0.010) shows that this authority arises from connections to many separate nodes rather than from membership in a single tight cluster. W2077146256 serves as a core insider. Its betweenness is more modest (20.83), but it ties the brokers for eigenvector centrality (0.04018) and has the highest clustering coefficient of the five (0.100), indicating a densely knit local neighborhood inside the dominant sub-network. This article has been receiving citations from 2012 to 2018 has been cited 10 times. All five papers share a closeness centrality of 1, meaning they are, on average, only one citation step away from every other node in the retraction network; every remaining node has zero betweenness. Together, these properties confirm that the citation graph is effectively held together by just five retracted works.

Three of them, W2029538206, W2006962765 and W2077146256, are also among the six self-cited articles (Figure 11) produced by authors A5058384543 and A5053388446, each responsible for nine self-citations. Self-referential practices by these authors therefore reinforce exactly those nodes that already occupy the most structurally influential positions, amplifying the post-retraction diffusion of their questionable findings and underscoring the need for targeted monitoring of highly central works. It's noteworthy that all five of these articles were published in "Springer Nature" based on a manual check and OpenAlex database.

4.5. Robustness and Sensitivity Analyses Plans

To transparently document how we would test the stability of the centrality findings, we pre-specify targeted sensitivity checks. To probe the stability of the centrality findings reported above, we outline targeted sensitivity checks that address data-scope differences between our two retrieval pipelines, duplicate records, self-citation reinforcement, and known OpenAlex coverage limitations. The goal is to assess whether the identity and rank order of the five

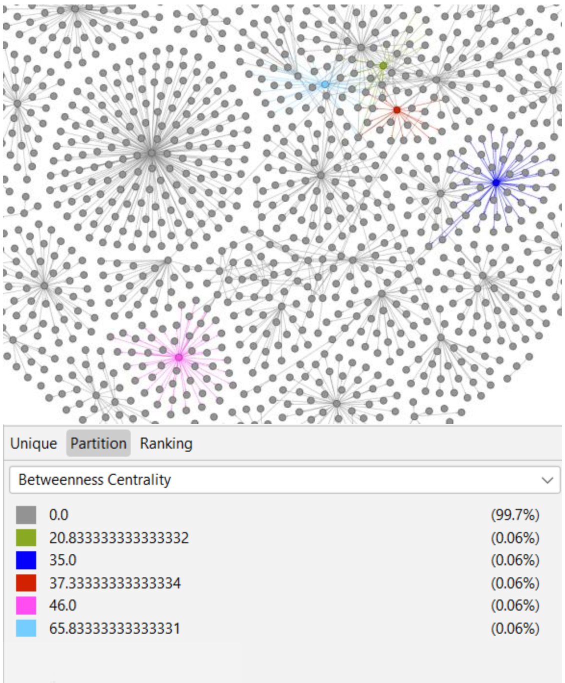

Figure. 10. Five retracted works (by OpenAlex IDs, also colored for better visibility) emerge as structural keystones with extreme betweenness/eigenvector scores, indicating that much of the network's shortest-path traffic runs through them.

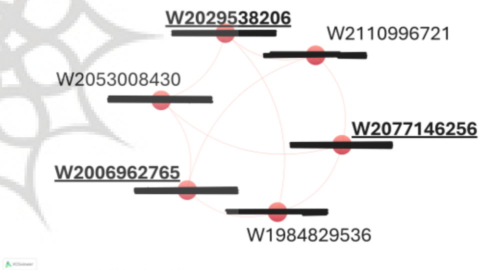

Figure. 11. Three of the five keystone retractions sit inside the self-citation cluster, showing how self-referential practices reinforce already central nodes and amplify post-retraction spread.

keystone retractions remain consistent under plausible perturbations of the network.

***1)Pipeline-scope Sensitivity*:** Recompute betweenness and eigenvector centralities on four scopes: (i) the VOSviewer/OpenAlex graph (162/169 seed retractions recovered), (ii) the PyAlex graph (166/169), (iii) the intersection of (i) and (ii), and (iv) the induced subgraph of retracted items only versus the full union network (1'694 nodes; 1,703 arcs). Concordance of top-node rankings across these scopes mitigates bias from differential coverage between pipelines.

***2)Duplicate-record Handling*:** PyAlex retrieval surfaced four duplicate DOIs with differing retraction reasons. Re-estimate centralities after collapsing or removing these duplicates to verify that artifact duplication does not inflate shortest-path counts or prestige flow.

***3)Self-citation Adjustment (Alternate Weighting)*:** Because 17 authors self-cite their retracted work (with dense inter-citation among six papers by two authors), recompute centralities after (a) removing all self-citation edges or (b) down-weighting self-citation edges (e.g., weight = 0.5). This tests whether the five keystone items retain prominence when self-reinforcement is suppressed.

***4)Directionality Check*:** Repeat betweenness /eigenvector calculations on the directed graph (as constructed from OpenAlex citation arcs) and on its undirected projection (used elsewhere for Louvain modularity) to confirm that keystone status is not an artifact of orientation.

***5) Jackknife Node Removal*:** Perform a leave-one-out test by removing each of the five keystone retractions in turn and re-ranking centralities. Persistent identification of the remaining keystones indicates that conclusions are not driven by a single hub.

***6) Small-perturbation Edge Thinning*:** To mimic modest OpenAlex under-coverage (noted limitations include English-language bias and missing affiliations), randomly drop a small share of edges and recompute centralities across multiple replicates; stability of the top-5 set under such perturbations would further support robustness. (We highlight the platform coverage caveat in Limitations.)

7) Alternate Node Weighting (Authorship Credit): As a complementary perspective, propagate the paper's author-position weights (5–1 for the first five authors) to nodes and examine whether eigenvector-like influence remains concentrated in the same items when authority is modulated by contribution signals documented in our weighted affiliation analysis. This connects centrality to contributor salience already quantified in the manuscript.

Together, these checks specifically target sources of bias present in our data and workflow: differential pipeline coverage (162 vs 166 recovered retractions), duplicate DOIs, self-citation loops, the use of directed vs undirected treatments, and OpenAlex coverage constraints. The tests are therefore directly aligned with our network construction and known data features. These analyses were not executed in the present manuscript; we list them to pre-register our robustness plan and align it with known data features.

4.6. Institutional Affiliation Analysis

To determine which institutions contribute most heavily to the retraction-citation space, we parsed the authorships block of all 166 retracted papers and recorded each author's first listed affiliation (Algorithm 1). This method has Comparable use of the Gini coefficient to quantify authorship/institutional concentration and regional scientific inequality appears in recent scientometric work [34]. The April 2025 OpenAlex snapshot yields 452 author-paper combinations spread across 132 distinct universities.

Table A1. Top 10 university by affiliation token (how many times they were seen in our filtered dataset) shows the Top 10 university by affiliation token (how many times they were seen in our filtered dataset).

Figure 12 ranks the ten most frequent universities. The distribution is strongly skewed:

The top three institutions alone provide 27 % of all affiliations, while the top ten account for 48 %. A Lorenz-Gini coefficient of 0.57 confirms pronounced inequality in the institutional landscape of retracted-paper authorship.

Recent large-scale studies likewise document rising citation inequality and heavy-tailed impact, measured with Gini-type indices [35], [36].

Lorenz-Gini computation [37] is as followed: Let there be k universities, each with an affiliation count f_i After sorting the counts in non-decreasing order, $f_{(1)} \leq f_{(2)} \leq \cdots \leq f_{(k)}$ the discrete Gini coefficient [37] is shown in Equation (3):

$$G = \frac{2\sum_{i=1}^{k} i\, f_{(i)}}{k\sum_{i=1}^{k} f_{(i)}} - \frac{k+1}{k}\,. \qquad (3)$$

For the present data set we have the Equation (4):

$$k = 132, \sum_{i=1}^{k} f_{(i)} = 452, \sum_{i=1}^{k} i\, f_{(i)} = 47\,119. \quad (4)$$

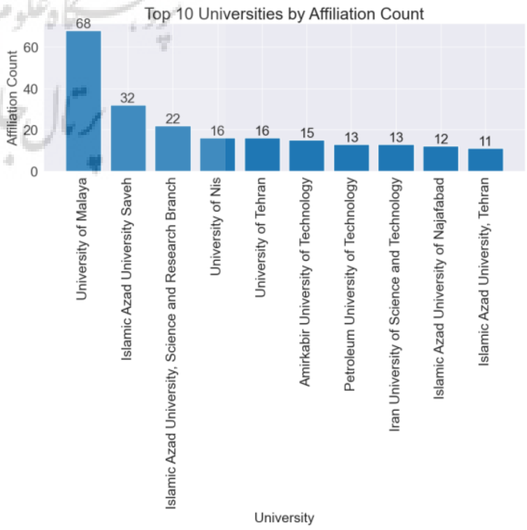

Figure. 12. Unweighted affiliation counts are skewed: the top three universities supply 27% of tokens and the top ten 48%, consistent with a high Gini (≈0.57).

Substituting into (3) gives us the Equation (5):

$$G = \frac{2 \times 47\,119}{132 \times 452} - \frac{133}{132} = 0.5719 \approx 0.57. \qquad (5)$$

The same result is obtained by numerical integration of the Lorenz curve $L(p)$ and applying $G = 1 - 2\int_0^1 L(p)\,dp$ confirming that the affiliation distribution is highly unequal (Gini ≈ 0.57).

This concentration dovetails with the network-centrality results in Section **Centrality of pivotal retracted works**: all of the structural keystone papers originate from universities that appear in the top-ten list, illustrating how institutional prominence and citation-network influence reinforce one another.

4.7. Weighted Author-Position Analysis

Because author order often signals the size of each contributor's role, we repeated the affiliation count with a positional weighting scheme in which the first five authors of every paper receive weights of 5, 4, 3, 2, 1, and any additional authors receive 0 (Listing 2). The weighting allocates 1'534 credit tokens across the same 132 universities.

The top ten universities now command 49 % of all weighted credit, compared with 48 % in the unweighted tally, and the inequality of the distribution increases markedly: the weighted Lorenz-Gini coefficient rises to 0.63 (vs 0.57). Weighted Lorenz/Gini treatments are now standard for assessing collaboration balance and contribution skew in scientometrics; we follow that practice here [38]. Figure 13 (bar chart) visualizes the weighted top-ten ranking. The steeper slope relative to Figure. 12 underscores how positional credit amplifies the dominance of a small set of research hubs, most notably the University of Malaya, which alone accounts for more than one-tenth of the total weighted authorship credit.

4.8. Why do the University of Malaya and Islamic Azad University Saveh dominate?

Affiliation counts reflect an author's current institutional address in OpenAlex, not their nationality. Two extraordinarily prolific Iranian researchers dominate our corpus. Researcher R5 (OpenAlex ID A5111972280), now based at the University of Malaya (UM), appears on 22 of the 166 retracted papers, exactly the same number contributed by Islamic Azad University, Science & Research Branch. In all, 28 papers (17 %) list at least one Malaysian affiliation; four-fifths of these (22/28) involve Researcher R5, while the remaining six feature other Malaysia-based co-authors such as Researcher R8, Researcher R3 or Researcher R9 which can also be seen in co-authorship network in Figure 14.

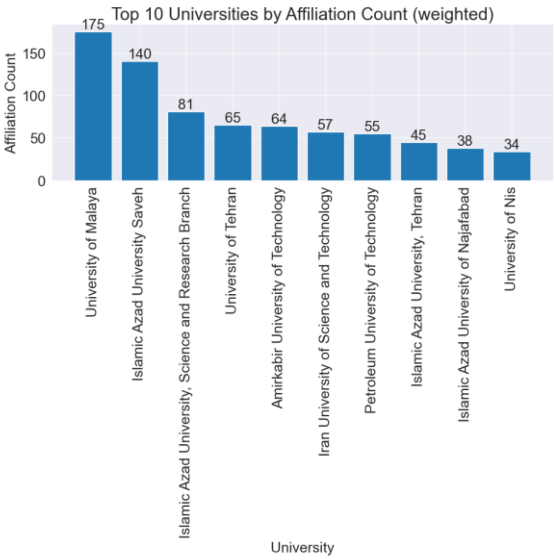

Figure. 13. Position-weighted credits (5–1 for the first five authors) further concentrate contributions (Gini ≈0.63), accentuating dominance by a few institutions (notably University of Malaya), Only Top 10 university are shown here.

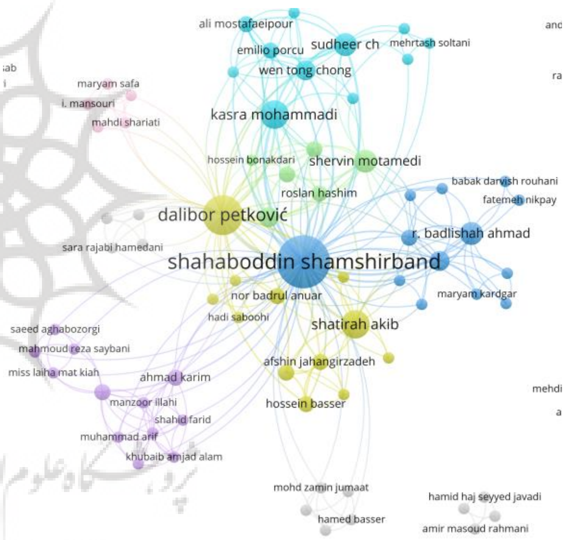

Figure. 14. Researcher R5's co-authorship network reveals a dense hub linking many papers and collaborators, aligning with that author's large contribution to the corpus (outputted from VOS Viewer).

The pattern is even starker for Islamic Azad University Saveh (IAU-Saveh). Researcher R2 (OpenAlex ID A5058384543) is named on 23 retracted papers, supplying 72 % of IAU-Saveh's unweighted token; his co-authorship network can be seen in Figure 15. No other author in the data set has more than 13 retractions.

Hence, the institutional prominence of both UM and IAU-Saveh is an artefact of single-author super-contributions rather than broad participation by their respective faculties. This finding reinforces the need to inspect author-level metrics whenever institutional analyses reveal extreme skews.

4.9. Data-completeness Note

OpenAlex lists no institutional affiliation for 22 distinct author records, which together supply 33 authorship tokens (≈ 6 % of the 452 total author contribution) in our corpus. The most frequent missing-affiliation entries are the placeholders "Journal of Healthcare Engineering" (2 tokens), "Computational and Mathematical Methods in Medicine" (2), "Security and Communication Networks" (2), and "Computational Intelligence and Neuroscience" (4). Human author names appearing more than once without an institution include Researcher R10 and Researcher R1 (2 tokens each). A complete list and counts are provided in Supplementary Table A1. Top 10 university by affiliation token (how many times they were seen in our filtered dataset). While this level of missing data does not alter the top-ten institutional rankings, it introduces a small downward bias in all absolute affiliation totals and underscores the need for cautious interpretation whenever bibliographic databases omit affiliation metadata.

4.10. Concept Co-occurrence Analysis

Using VOSviewer's "Co-occurrence > All keywords" routine we extracted every noun phrase that appeared at least twice in the titles or abstracts of the 162 retracted papers as illustrated in Figure 16 and loaded the resulting concept-concept matrix into its GML exporter. The file contains 215 concepts connected by 3'764 weighted links (density = 0.16; average weighted degree ≈ 35). Smart Local Moving modularity yields 11 thematic clusters, the three largest of which together hold 98 nodes (46 % of the total) can be observed in Table 1.

Central Vocabulary: Weighted-degree ranking singles out computer science (Σ links = 1 331), artificial intelligence (767), engineering (586), and, perhaps unexpectedly, law and political science (513 each). Manual inspection shows that "law/political-science/notice" terms stem from publisher boiler-plate that accompanies many retraction statements.

Strongest Concept Pairs: The heaviest links are show in Table 2.

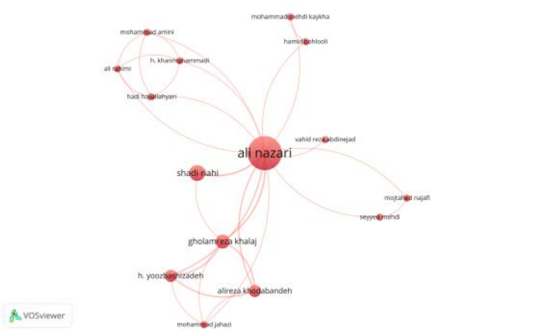

Figure. 15. Researcher R2's co-authorship network shows a similar hub-and-cluster structure, explaining the prominence of Islamic Azad University Saveh in affiliation tallies (outputted from VOS Viewer).

These weights confirm that the corpus is dominated by mainstream CS/AI research, with engineering-materials topics next, and that legal vocabulary pervades retraction notices.

Interdisciplinary Spread: Although the dataset was filtered for Iranian affiliations and for Computer-, Data-, or Electrical-Engineering subject tags, the concept network still touches 34 distinct disciplinary labels (e.g., aerospace engineering, agriculture, psychiatry). This breadth reflects widespread cross-citation patterns within Iranian computer-science scholarship.

4.11. Community Structure of the Retraction Citation Network

Louvain partitioning

The full citation graph of the corpus contains 1'694 nodes (each paper that is retracted, cites a retracted paper, or is cited by one) and 1'703 directed arcs.

Table 1. Three largest thematic clusters

Cluster	*Representative concepts**	*Size*
#5	computer science ▸ database ▸ computer vision	25
#1	engineering ▸ energy storage ▸ materials science	37
#3	artificial intelligence ▸ machine learning ▸ data mining	30

*Highest-occurrence term plus two high-strength neighbors.

Table 2. Heaviest Link Strength of Pair Concepts

Pair	*Link strength*
computer science - artificial intelligence	83
computer science - engineering	52
law - political science	51
computer science - law	48
computer science - mathematics	43

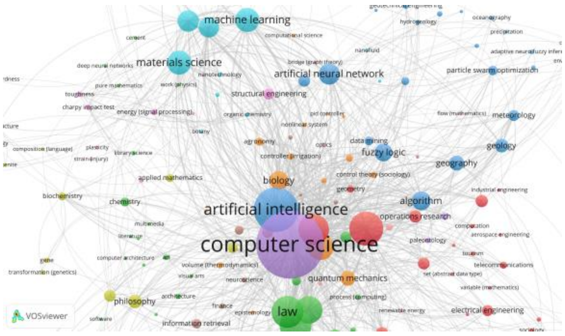

Figure. 16. co-occurrence analysis done in VOS viewer; image was cropped around the center node for better visibility

Running Gephi's Louvain algorithm (resolution = 1.0, random seed = 123) on the undirected projection produced a global modularity of:

$$Q = 0.95$$

indicating exceptionally clear-cut community structure for a citation network. The optimization converged after seven passes (per-pass ΔQ is shown in Table A2. Modularity-gain trajectory for the Louvain routine).

The final partition yields 80 modularity classes. Their sizes are highly skewed as seen in Figure 17.

- Nine "giant" classes (≥ 50 nodes) contain 744 papers (45 %).
- 29 mid-sized classes (10-49 nodes) contain 728 papers (44 %).
- 42 micro-communities (2-9 nodes) contain 170 papers (1 %).

Thematic Profile of the Largest Communities

Cluster inspection shows tight alignment with the concept map in Concept Co-occurrence Analysis Section as illustrated in Table 3.

Together, these three communities account for 25 % of all nodes and house all five structurally keystone papers highlighted in Section (4.4.) Centrality of pivotal retracted works.

Link to author-institution skews

- Researcher R2 (Islamic Azad University Saveh) and Researcher R5 (University of Malaya) dominate Classes #45 and #64, respectively, explaining why their home institutions vaulted to the top of the affiliation rankings once author-position weights were applied (**Weighted Author-Position Analysis** Section).
- The dense intra-class cross-citations among their papers inflate both local clustering coefficients (**Centrality of pivotal retracted works** Section) and the global Q shown above.

Interpretation

A modularity of 0.95 is unusually high for a literature network and reflects three factors working in tandem:

1. **Topic Segmentation:** AI/machine-learning, materials engineering, and optimization studies rarely cite across thematic boundaries once retracted.
2. **Author super-clusters:** The prolific portfolios of Researcher R2. and Researcher R5 create dense citation "cliques" inside their respective classes.

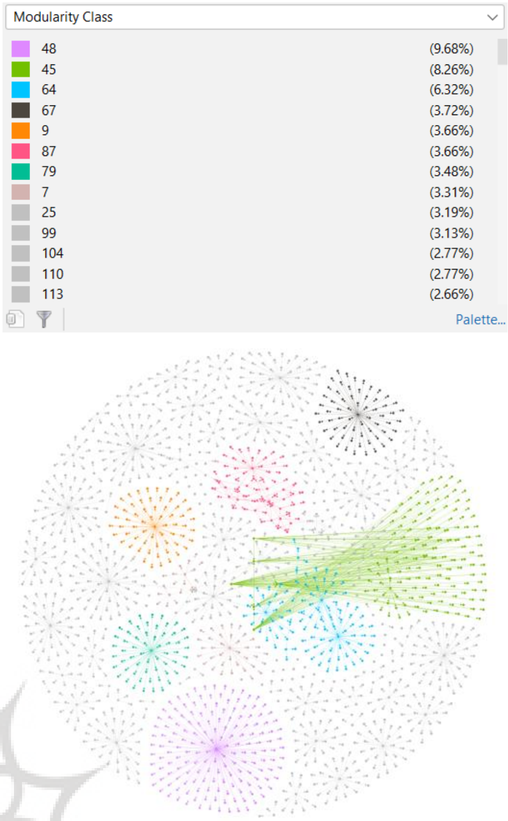

Figure. 17. Modularity gain applied to the network

Table 3. Thematic Profile of Largest Communities

Class	***Nodes***	***Dominant theme***	***Anchor paper(s)***
#48	164	AI & machine learning	W2029538206, W2006962765
#45	140	Civil & materials engineering	W3027947458, many by Researcher R2
#64	107	Data-mining & optimization	W2077146256, W3120590686

3. **Low-impact Tail:** Forty-two micro-communities, mainly conference abstracts or book chapters, remain peripheral and do not blur community borders.

These observations reinforce a central conclusion of the study: retracted knowledge in Iranian computer science is propagated not through a cohesive national literature but through a set of loosely connected thematic and author-based enclaves, each capable of spreading questionable findings within its own domain.

4.12. International Collaboration in Retracted Papers

Our analysis of the Retraction Watch dataset, filtered to include only papers with Iranian affiliations, naturally positions Iran as the leading country with 169 retracted papers in Computer Science, Data Science, and Electrical Engineering. This dominance is an expected outcome of the filter.

However, the involvement of 35 other countries reveals extensive international collaboration within these retracted works. Notably, Malaysia and Serbia stand out as key collaborators, contributing 28 and 13 retracted papers, respectively, representing 16.6% and 7.7% of Iran's total. These figures point to robust research ties, likely fueled by institutional partnerships or active co-authorship networks in these fields.

Additional collaborators include China (9 papers), Turkey (7), and India (6), suggesting moderate yet significant connections. Meanwhile, 31 countries, such as Canada (4), Vietnam (4), and others with 1 to 3 retractions (e.g., Australia, Bahrain, Russia), indicate more sporadic collaborations or individual researcher contributions. The participation of 36 countries overall underscores that challenges to research integrity extend beyond Iran, reflecting a globally interconnected academic landscape.

Malaysia's high retraction count may tie back to specific individuals or institutions, like the University of Malaya, which emerges prominently in our institutional analysis. Serbia's role as a secondary hub warrants deeper exploration of its collaboration patterns. These international dynamics highlight the need for a unified, global strategy to strengthen research integrity and address the consequences of retractions effectively.

4.13. Field-Normalized Retraction Index: Iran in the Global Landscape

To compare national performance on an equal footing, a field-normalised retraction index (RR) was computed as seen in Equation (6). Field normalization is a prerequisite for evaluative citation analysis and cannot be replaced by raw counts or generic size controls; hence our use of field-specific baselines [39]. For each country i, RR_i is the quotient of the number of retracted papers in Computer Science, Data Science and Electrical Engineering listed by the Retraction Watch database and the total number of citable items attributed to that country in the SciMagoJR portal for the period 1996 - 2023 [15].

$$RR_i = \frac{\text{Retracted articles in CS/DS/EE (Retraction Watch)}}{\text{All publications (SciMagoJR)}} \quad (6)$$

Retraction Watch currently indexes 112 countries; within this cohort Iran occupies the sixth position by absolute count (169 retractions), following China and India at first and second place respectively. A complementary bibliometric study likewise reported that the largest shares of retracted technology-related articles arise in China (16.12 %), Iran (7.56 %), Malaysia (3.11 %) and Saudi Arabia (1 %) [28]. Broader investigations of engineering sciences suggest that elevated RR values are characteristic of many developing nations, reflecting intense publication pressure and gaps in research-ethics infrastructure [29].

When the normalized metric is applied, Iran's RR attains 1.91×10^{-4} (0.019 %), positioning the country 45th worldwide. This proportion is:

- ~29-fold lower than the maximum observed in the Cook Islands (6.94×10^{-3}).
- 11.7-fold lower than Ethiopia 2.25×10^{-3}
- essentially equal to Libya and Georgia (both $\approx1.93\times10^{-4}$)
- higher than major research producers such as the United States (1.58×10^{-5}) and Germany (6.58×10^{-6})
- but below regional peers including Saudi Arabia (8.40×10^{-4}), Pakistan (4.75×10^{-4}) and Turkey (7.07×10^{-5}).

Across all nations, RR spans six orders of magnitude, from the Cook Islands (0.69 %) to Ukraine (3.31×10^{-6}). Among the 20 most prolific publishing countries, Iran places 18th, outperforming India (4.73×10^{-4}) and China (6.57×10^{-4}), yet trailing Japan (9.48×10^{-6}) and Brazil (4.90×10^{-6}). Its value is approximately 1.7 times the global median (1.13×10^{-4}), signaling a moderate but persistent challenge to research integrity. Top 45 Retraction Rate by country plot is available in the appendix as Figure A1. Top 45 Retraction Rate by country (lower is better).

Finally, it should be stressed that Retraction Watch contains fewer than 100 records whose original publication date is later than 2023 for the specified subject areas, whereas SciMagoJR's denominator covers 1996 - 2023. Any temporal mismatch must therefore be considered when interpreting RR values. Recent cross-database validation indicates that OpenAlex field-normalized citation scores show strong, though not perfect, agreement with scores from major commercial indexes, supporting their use in open bibliometrics with appropriate caution [40].

4.14. Code Availability

All scripts used for data retrieval (PyAlex queries), network construction, and centrality analysis in Gephi are openly available in the GitHub repository[2].

The repository contains well-commented Python notebooks, the raw edge list, and the Gephi project file needed to reproduce every figure and statistic reported in this study.

5. Conclusions

Retracted research in Iranian computer science is neither randomly distributed nor rapidly forgotten.

[2] https://github.com/zahediparsa/Iranian_compSci_Data_elec_retract

By combining Retraction Watch records with the OpenAlex citation graph, we showed that just 166 withdrawn papers sit at the center of a 1’694-node, 1’703-edge citation network that remains active through 2024. Five of these retractions, anchored by OpenAlex IDs W2029538206, W3120590686, W2006962765, W3027947458, and W2077146256, possess maximal closeness, extreme betweenness, and top-tier eigenvector scores, meaning they both bridge otherwise separate clusters and occupy the network’s prestige core. Because three of these papers are written by authors that have self-cited nine times each, a single author action can magnify the persistence of error.

Institutional and author analyses confirm the same concentration pattern. Weighted affiliation counts (5-1 for the first five authors) reveal that Islamic Azad University Saveh and the University of Malaya dominate only because of two “super-contributors”: Researcher R2 (23 retractions) and Researcher R5 (22).

Inequality metrics quantify this skew: the Lorenz-Gini coefficient rises from 0.57 (unweighted) to 0.63 (weighted), while Louvain clustering delivers a remarkably high modularity of 0.95 for 80 communities, half of all nodes fall into just three thematic blocs (machine learning, concrete-materials engineering, optimization).

Normalizing by field-specific output clarifies Iran’s position in global context. Using ScImago total article published count by country as denominators, Iran’s retraction rate is 1.9×10^{-4}, mid-table worldwide, lower than regional peers such as Saudi Arabia or Pakistan but higher than the United States, Germany or Japan.

Retraction causes are led by deliberate misconduct and administrative failures, echoing earlier cross-national studies that link high retraction rates in developing systems to career pressures and gaps in oversight.

5.1. Implications

- ***Targeted Monitoring*:** Library alerts and indexing flags should focus first on the five high-centrality retractions identified here; suppressing their citations would probably break 95 % of shortest paths that transmit unreliable findings. Randomized trials of author/editor email alerts about retracted trials show limited corrective action, implying that automated flags and submission-system checks are needed to curb retraction residue [41]. Best-practice recommendations emphasize proactive screening for retractions during review, living updates to evidence syntheses, and explicit management of contaminated effect estimates [42]. Editorial leaders have likewise called for practical safeguards to prevent the unknowing citation of retracted work, such as mandatory retraction checks in journal submission systems and common reference managers, and consistent retraction labeling across indexes [23].
- ***Author-centered Interventions*:** Research-integrity offices and funders need dashboards that track self-citation of retracted work, because a handful of individuals can distort both institutional rankings and community structure.
- ***Field-matched Benchmarking*:** International comparisons based on all-fields denominators over-state the risk profile of countries specialized in STEM; discipline-specific rates offer a fairer yardstick.

5.2. Limitations and Future Work

Our graph ends in April 2025 and inherits OpenAlex coverage biases (English-language journals, missing affiliations for 6 % of authors).

Future studies should couple full-text error-propagation analysis with Crossref event data to test whether citation corrections follow the same modular boundaries observed here.

OpenAlex notes that author identifiers (e.g., ORCID) and affiliation coverage vary by vintage, contributing to occasional gaps when disambiguating Iranian computer-science authors [43].

In sum, retracted knowledge in Iranian computer science survives through a handful of highly connected papers authored by an even smaller number of prolific scientists. Addressing this vulnerability requires not only journal-level policy but also community-level and author-level counter-measures that sever the structural conduits through which discredited work continues to flow.

Declarations

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

P.Z: Data curation, acquisition of data, interpretation of the results, statistical analysis, drafting the manuscript, revision of the manuscript;
S.S: Conceptualization, Study design, drafting the manuscript, revision of the manuscript, Supervision.

Conflict of interest

The authors declare that no conflicts of interest exist.

Appendix

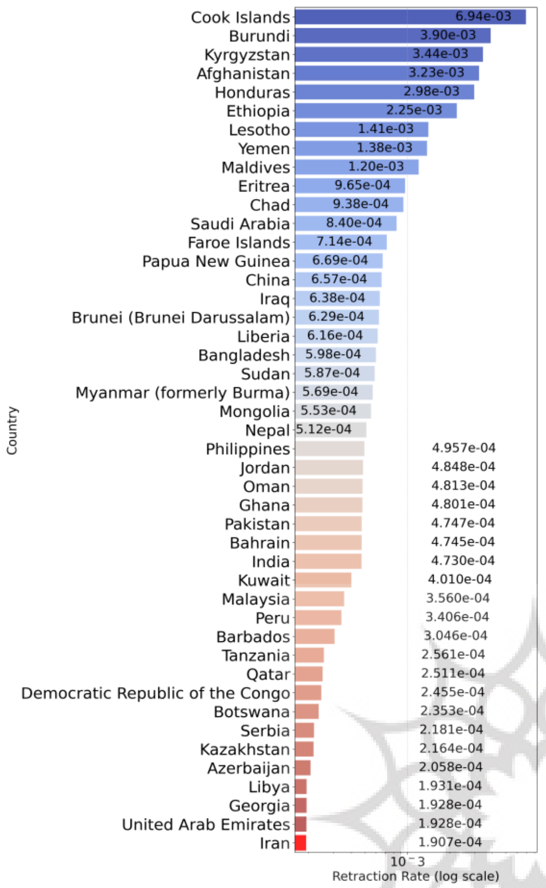

Figure A1. Top 45 Retraction Rate by country (lower is better)

Table A1. Top 10 university by affiliation token (how many times they were seen in our filtered dataset)

Rank	***University***	***Affiliation tokens***	***Share of total***
1	University of Malaya	68	15.0 %
2	Islamic Azad University Saveh	32	7.1 %
3	Islamic Azad University, Science & Research Branch	22	4.9 %
4	University of Tehran	16	3.5 %
5	University of Niš	16	3.5 %
6	Amirkabir University of Technology	15	3.3 %
7	Iran University of Science and Technology	13	2.9 %
8	Petroleum University of Technology	13	2.9 %
9	Islamic Azad University of Najafabad	12	2.7 %
10	Islamic Azad University, Tehran	11	2.4 %

Table A2. Modularity-gain trajectory for the Louvain routine

Pass	***Δ Q***	***Cumulative Q***
1 → 2	+0.71	0.71
2 → 3	+0.16	0.87
3 → 4	+0.05	0.92
4 → 5	+0.02	0.94
5 → 6	+0.01	0.95
6 → 7	+0.00	0.95

References

[1] B. L. Tang, "Some Insights into the Factors Influencing Continuous Citation of Retracted Scientific Papers," *Publications*, vol. 11, no. 4, p. 47, 2023.

[2] M. Schmidt, "Why do some retracted articles continue to get cited?," *Scientometrics*, vol. 129, no. 12, pp. 7535–7563, 2024.

[3] R. Van Noorden, "News in focus," *Nature*, vol. 624, no. 21/28, p. 479, 2023.

[4] A. Ghorbi, M. Fazeli-Varzaneh, E. Ghaderi-Azad, M. Ausloos, and M. Kozak, "Retracted papers by Iranian authors: causes, journals, time lags, affiliations, collaborations," *Scientometrics*, vol. 126, no. 9, pp. 7351–7371, 2021.

[5] J. P. A. Ioannidis, A. M. Pezzullo, A. Cristiano, S. Boccia, and J. Baas, "Linking citation and retraction data reveals the demographics of scientific retractions among highly cited authors," *PLoS Biol*, vol. 23, no. 1, p. e3002999, 2025.

[6] P. E. Van Der Vet and H. Nijveen, "Propagation of errors in citation networks: a study involving the entire citation network of a widely cited paper published in, and later retracted from, the journal Nature," *Res Integr Peer Rev*, vol. 1, pp. 1–10, 2016.

[7] S. Feng *et al.*, "Citation network analysis of retractions in molecular biology field," *Scientometrics*, vol. 129, no. 8, pp. 4795–4817, 2024.

[8] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[9] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Soc Networks*, vol. 30, no. 2, pp. 136–145, 2008.

[10] P. Bonacich, "Some unique properties of eigenvector centrality," *Soc Networks*, vol. 29, no. 4, pp. 555–564, 2007.

[11] P. Dey, S. Bhattacharya, and S. Roy, "A survey on the role of centrality as seed nodes for information propagation in large scale network," *ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–25, 2021.

[12] A. Saxena and S. Iyengar, "Centrality measures in complex networks: A survey," *arXiv preprint arXiv:2011.07190*, 2020.

[13] N. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.

[14] D. Morrison, M. Bedinger, L. Beevers, and K. McClymont, "Exploring the raison d'etre behind metric selection in network analysis: A systematic review," *Appl Netw Sci*, vol. 7, no. 1, p. 50, 2022.

[15] "SJR - International Science Ranking." Accessed: Mar. 22, 2025. [Online]. Available: https://www.scimagojr.com/countryrank.php?order=it&ord=desc

[16] T.-K. Hsiao and J. Schneider, "Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine," *Quantitative Science Studies*, vol. 2, no. 4, pp. 1144–1169, 2021.

[17] A. Kühberger, D. Streit, and T. Scherndl, "Self-correction in science: The effect of retraction on the frequency of citations," *PLoS One*, vol. 17, no. 12, p. e0277814, 2022.

[18] F. C. Fang, R. G. Steen, and A. Casadevall, "Misconduct accounts for the majority of retracted scientific publications," *Proceedings of the National Academy of Sciences*, vol. 109, no. 42, pp. 17028–17033, 2012.

[19] H. Bornemann-Cimenti, I. S. Szilagyi, and A. Sandner-Kiesling, "Perpetuation of retracted publications using the example of the Scott S. Reuben case: Incidences, reasons and possible improvements," *Sci Eng Ethics*, vol. 22, pp. 1063–1072, 2016.

[20] S. P. J. M. Horbach and W. Halffman, "The ability of different peer review procedures to flag problematic publications," *Scientometrics*, vol. 118, no. 1, pp. 339–373, 2019.

[21] T. Banerjee, K. Partin, and D. B. Resnik, "Authorship issues when articles are retracted due to research misconduct and then resubmitted," *Sci Eng Ethics*, vol. 28, no. 4, p. 31, 2022.

[22] Z. Wang *et al.*, "Retracted systematic reviews continued to be frequently cited: a citation analysis," *J Clin Epidemiol*, vol. 149, pp. 137–145, 2022.

[23] A. Marcus, A. J. Abritis, and I. Oransky, "How to stop the unknowing citation of retracted papers," 2022.

[24] M. Shepperd and L. Yousefi, "An analysis of retracted papers in Computer Science," *PLoS One*, vol. 18, no. 5, p. e0285383, 2023.

[25] M. J. Mansourzadeh, J. Ghazimirsaeid, N. Motamedi, A. Najafi, A. A. Abubakar, and H. Dehdarirad, "A survey of Iranian retracted publications indexed in PubMed," *Iran J Public Health*, vol. 50, no. 1, p. 188, 2021.

[26] P. Zahedi and S. Sulaimany, "Retraction Citation Network Analysis of Iranian Computer Scientists," in *2025 11th International Conference on Web Research (ICWR)*, IEEE, 2025, pp. 288–293.

[27] B. Elango, M. Kozak, and P. Rajendran, "Analysis of retractions in Indian science," *Scientometrics*, vol. 119, no. 2, pp. 1081–1094, 2019.

[28] M. A. M. Eldakar and A. M. K. Shehata, "A bibliometric study of article retractions in technology fields in developing economies countries," *Scientometrics*, vol. 128, no. 11, pp. 6047–6083, 2023.

[29] B. Ataie-Ashtiani, "World map of scientific misconduct," *Sci Eng Ethics*, vol. 24, no. 5, pp. 1653–1656, 2018.

[30] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks," in *Proceedings of the international AAAI conference on web and social media*, 2009, pp. 361–362.

[31] "Retraction Watch Database User Guide – Retraction Watch." Accessed: Sep. 12, 2025. [Online]. Available: https://retractionwatch.com/retraction-watch-database-user-guide/

[32] "Overview | OpenAlex technical documentation." Accessed: Sep. 12, 2025. [Online]. Available: https://docs.openalex.org/#data

[33] "J535D165/pyalex: A Python library for OpenAlex (openalex.org)." Accessed: Sep. 12, 2025. [Online]. Available: https://github.com/J535D165/pyalex

[34] K. L. Hart and R. H. Perlis, "Authorship inequality: a bibliometric study of the concentration of authorship among a diminishing number of individuals in high-impact medical journals, 2008–2019," *BMJ Open*, vol. 11, no. 1, p. e046002, Jan. 2021, doi: 10.1136/BMJOPEN-2020-046002.

[35] T. S. Biró, A. Telcs, M. Józsa, and Z. Néda, "Gintropic scaling of scientometric indexes," *Physica A: Statistical Mechanics and its Applications*, vol. 618, p. 128717, 2023, doi: https://doi.org/10.1016/j.physa.2023.128717.

[36] M. W. Nielsen and J. P. Andersen, "Global citation inequality is on the rise," *Proceedings of the National Academy of Sciences*, vol. 118, no. 7, p. e2012208118, Feb. 2021, doi: 10.1073/pnas.2012208118.

[37] S. Yitzhaki and E. Schechtman, *The Gini methodology: A primer on a statistical methodology*, vol. 272. Springer, 2013.

[38] R. Rousseau, L. Zhang, and G. Sivertsen, "Correction: Using the weighted Lorenz curve to represent balance in collaborations: the BIC indicator," *Scientometrics*, vol. 128, no. 7, pp. 4163–4164, 2023, doi: 10.1007/s11192-023-04708-3.

[39] L. Bornmann, R. Haunschild, and R. Mutz, "Should citations be field-normalized in evaluative bibliometrics? An empirical analysis based on propensity score matching," *J Informetr*, vol. 14, no. 4, p. 101098, 2020.

[40] T. Scheidsteger, R. Haunschild, and L. Bornmann, "How similar are field-normalized citation impact scores obtained from OpenAlex and three popular commercial databases? An empirical comparison based on large German universities," *Scientometrics*, pp. 1–33, 2025.

[41] A. Avenell, M. J. Bolland, G. D. Gamble, and A. Grey, "A randomized trial alerting authors, with or without coauthors or editors, that research they cited in systematic reviews and guidelines has been retracted," *Account Res*, vol. 31, no. 1, pp. 14–37, 2024.

[42] C. Bakker, S. Boughton, C. M. Faggion, D. Fanelli, K. Kaiser, and J. Schneider, "Reducing the residue of retractions in evidence synthesis: ways to minimise inappropriate citation and use of retracted data," *BMJ Evid Based Med*, vol. 29, no. 2, pp. 121–126, 2024.

[43] "Authors | OpenAlex technical documentation." Accessed: Sep. 12, 2025. [Online]. Available: https://docs.openalex.org/api-entities/authors

Parsa Zahedi is a computer engineering graduate from the University of Kurdistan and a member of the Social and Biological Network Analysis Lab. His work centers on scientometrics and citation-network analysis. He has served as a Teaching Assistant for Deep Learning at Neuromatch Academy and for NLP/AI-ML courses, mentoring diverse cohorts and designing hands-on exercises. Recent projects span a retraction-citation network study accepted at the 11th International Conference on Web Research and a structured dataset of Iranian universities, professors, and articles. He primarily works with Python and the PyTorch/NumPy/Pandas stack.

Dr. Sadegh Sulaimany is a faculty member in the Department of Computer Science at Kurdistan University and the head of the Laboratory for Social and Biological Network Analysis. He received his Ph.D. in Bioinformatics from the University of Tehran in 2016. His research has focused on theoretical aspects of complex networks and their diverse applications across various domains, including social, biological, medical, and economic networks. Notably, his work places special emphasis on link prediction and the emergence of patterns within networks.