International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2026, 10 (1), 64-71
DOI: 10.36287/ijmsit.10.1.7
Received: June 3, 2026; Accepted: July 23, 2026

A Graph-Based Article Recommendation System with Multilayer Citation

Umut Karadurmus[1*] and Zuleyha Akusta Dagdeviren[2]

1 Department of Computer Engineering, Graduate School of Natural and Applied Sciences, Ege University, Izmir, Turkey (umut.karadurmus@outlook.com) (ORCID: 0009-0009-2231-6376)
2 Department of Computer Engineering, Ege University, Izmir, Turkey (zuleyha.akusta.dagdeviren@ege.edu.tr) (ORCID: 0000-0001-9365-326X)

Abstract – We present a citation-centered article recommendation system that ranks works in a heterogeneous two-hop OpenAlex graph by combining structural connectivity with concept overlap and venue signals. Direct references form a primary, local-reading tier, whereas references reached through them form a secondary discovery tier. Ten target articles were evaluated from a data snapshot collected on March 8, 2026. Because concept-set Jaccard is also a scoring input, we supplemented it with an independent abstract-level TF–IDF cosine analysis and a leave-one-out hidden-reference experiment. For the full model, mean abstract cosine was 0.045 for primary and 0.029 for secondary recommendations among available abstract pairs. Across 358 held-out direct citations, 45.5% re-entered the two-hop candidate pool; Recall@10 was 0.112 overall and 0.245 when conditioned on candidate coverage. The networks were sparse and weakly connected when citation direction was ignored, although these properties are partly induced by ego-network construction. The findings show that a lightweight and interpretable hybrid can recover relevant citation-neighborhood works without model training, while also exposing candidate-generation and ground-truth limitations that require larger labelled or user-based evaluation.

Keywords – *citation networks, complex networks, recommendation systems, paper recommendation, OpenAlex*

Citation: *Karadurmus, U., Akusta Dagdeviren, Z. (2026). A Graph-Based Article Recommendation System with Multilayer Citation. International Journal of Multidisciplinary Studies and Innovative Technologies, 10(1): 64-71.*

I. INTRODUCTION

Keeping up with scientific literature has always been demanding, but the sheer volume of publications today makes it genuinely difficult even for experienced researchers to identify work that is relevant to their own. Search engines and keyword queries help, yet they often miss important connections that only become visible once you look at how papers cite one another. This work is motivated by that practical observation: we wanted a recommendation tool that could exploit the relational structure of citations, not just surface-level text similarity.

The system we built queries the OpenAlex [1] index to retrieve primary and secondary citations of a target article, along with publication venue metadata and the concept tags that OpenAlex attaches to scholarly works. These span broad areas such as computer science, environmental science, and machine learning, providing a lightweight but surprisingly expressive semantic layer on top of the raw citation graph.

Treating the resulting data as a complex network proved to be the natural choice. Highly connected articles tend to occupy central roles in their research communities, and weighting candidates by both structural prominence and topical relevance lets the system prioritize publications that are influential and on-topic. The value of network structure for identifying meaningful scholarly relations has been well documented [2], and more recent work has reinforced that multilevel citation modeling, semantic representations, and hybrid ranking all contribute to better recommendations [3, 4, 5]. Our design draws on these insights while keeping the approach lightweight enough to run efficiently over OpenAlex data.

The study's innovation is an interpretable two-tier recommendation design rather than a new representation-learning architecture. Direct references are deliberately retained as a high-confidence local-reading list, while second-hop references provide a separate discovery list; both are ranked by one transparent formula that combines structural evidence with bounded semantic and venue bonuses. In contrast to embedding-heavy, interaction-based, or supervised systems [5–7], the method requires no training data and can be reproduced from public OpenAlex metadata.

II. LITERATURE REVIEW

A. Complex Networks

A complex network is simply a graph in which the nodes and edges represent real-world entities and their dependencies, whether those entities are people, genes, routers, or academic papers. What makes such networks interesting is that they tend to develop non-trivial structural properties that are hard to predict from local rules alone. Two canonical examples are small-world networks [8], which maintain surprisingly short paths between any two nodes despite being large, and scale-

free networks [9], whose degree distributions follow a power law so that a handful of highly connected hubs coexist with many sparsely connected nodes. Both phenomena appear across natural, social, and technological systems, and both have practical implications for how information spreads through a network. Citation graphs inherit several of these properties, which is part of what makes them attractive for recommendation tasks.

B. Graph Theory

Graph theory provides the mathematical foundation for analyzing relational structures [10]. At its most basic level, a graph encodes a set of objects and the pairwise relationships between them: objects become vertices and relationships become edges. Despite this simplicity, the framework scales to arbitrarily large networks, and the metrics central to this study, namely node degree, shortest path length, density, and clustering coefficient, all derive from it.

C. Network Theory

Where graph theory focuses on combinatorial structure, network theory is more concerned with dynamics and measurement: how does information or influence propagate, what does the degree distribution reveal about how the network formed, and which nodes are most central by various criteria? These questions guide the choice of network metrics reported in the Results and Discussion section and help interpret their relevance to scholarly recommendation.

D. Citation Networks

Citation networks are a natural application of the ideas above. Each paper is a node; each reference is a directed edge. The resulting structure encodes not just who cites whom but also how knowledge has accumulated and branched over time. Highly cited papers tend to occupy central positions in the network, and clusters of mutual citation often correspond to coherent research sub-communities. Tools like CitNetExplorer and CiteSpace have made large-scale exploration of these structures practical [11, 12], and they have been used to trace the evolution of entire fields.

For recommendation specifically, the citation graph offers something that keyword search cannot easily replicate: implicit endorsement. When a paper cites another, its authors are signaling relevance in a domain-aware way. Extending this logic to two hops, from primary to secondary citations, broadens the search space without losing the topical grounding that single-hop retrieval provides. Son and Kim showed that multilevel citation structures improve recommendation quality [4], Pornprasit et al. demonstrated the value of embedding-based representations of citation graphs [6], and the systematic review by Liang and Lee confirms that hybrid and network-aware approaches have steadily gained ground over the past decade [3].

E. Recommendation Systems

At their core, recommendation systems solve an information-filtering problem: given a large catalog and some signal about what a user finds relevant, identify the items most worth their attention. Collaborative filtering, content-based matching, and hybrid combinations are the dominant paradigms in the general literature [13], with matrix factorization occupying a particularly prominent place thanks to its ability to uncover latent structure in user-item interaction data [14].

Scholarly recommendation has its own constraints because explicit ratings are rare and relevance must instead be inferred from citation links, text, topics, authorship, and publication context. Zhang et al. model semantic and relational context [5], Färber and Sampath combine context-aware signals [7], and Kanwal et al. rank multiple citation-network features [15]. More recent systems use deep link prediction [17], context-aware neural models [18], temporal graph representations, and citation networks enriched with abstract embeddings or large language models [19, 20]. These methods can capture richer semantics but commonly require learned representations, large benchmark corpora, or external embedding services. The proposed method occupies a complementary point in this design space: it prioritizes transparency, open metadata, and low implementation cost while retaining structural, semantic, and venue evidence.

TABLE I. POSITIONING AGAINST REPRESENTATIVE RECENT APPROACHES

Approach	Representation and contrast
Semantic/relational [5]	Learned semantic and citation-relation representation; richer but training-dependent.
Network embedding [6]	Citation embeddings; captures latent structure but reduces score transparency.
Deep/link models [17,18]	Neural content/context and link prediction; benchmark-oriented learned models.
LLM-enhanced hybrid [20]	Citation structure plus abstract embeddings; flexible but uses an external embedding model.
Proposed method	OpenAlex graph + concepts + venue; no training; transparent local/discovery tiers.

III.METHODOLOGY

Our approach builds a citation-centered complex network round a target article and then applies a lightweight scoring procedure to rank candidate recommendations. The network simultaneously captures publication relationships, topical concepts, and venue information (Fig. 1), making it possible to recommend papers that are both structurally connected to the target and thematically aligned with it.

Each run produces six recommendations in two intentionally separate groups. The three primary recommendations are selected only from works that occur in the target paper's own reference list; they support local reading, verification of foundational sources, and navigation within the target's immediate scholarly context. The three secondary recommendations are selected from works cited by those direct references but not by the target; they support broader discovery beyond the target's bibliography.

A. Input

The procedure takes one OpenAlex work identifier as input. The archived experiments used a March 8, 2026 data snapshot. For each target, the system retrieved up to 50 direct references and up to 30 references for each loaded direct reference, together with concept/topic, abstract, citation-count, and venue metadata. The resulting heterogeneous directed graph contains work (W), topic/concept (C), and source (S) nodes with CITES, HAS_TOPIC, and PUBLISHED_IN edges.

B. Categorization

The target's loaded direct references constitute the primary candidate pool. The union of references cited by those primary

works, excluding the target and all primary candidates, constitutes the secondary pool. Keeping the pools separate prevents highly connected second-hop works from displacing directly cited works and makes the two usage modes explicit: primary for immediate context and secondary for exploration.

Candidates receive a structural base score plus two optional bonuses. The semantic bonus is α times the Jaccard overlap between target and candidate concept sets. The venue bonus is β for an exact source match and β/2 for a publisher-only match. We fixed α = 0.3 and β = 0.2 before evaluation; they were not tuned against the ten targets. These bounded values keep the direct or two-hop citation signal dominant while allowing contextual evidence to resolve close structural scores. A sensitivity sweep over 0.0–0.5 in 0.1 increments was used to examine list membership changes rather than to select an accuracy-maximizing setting.

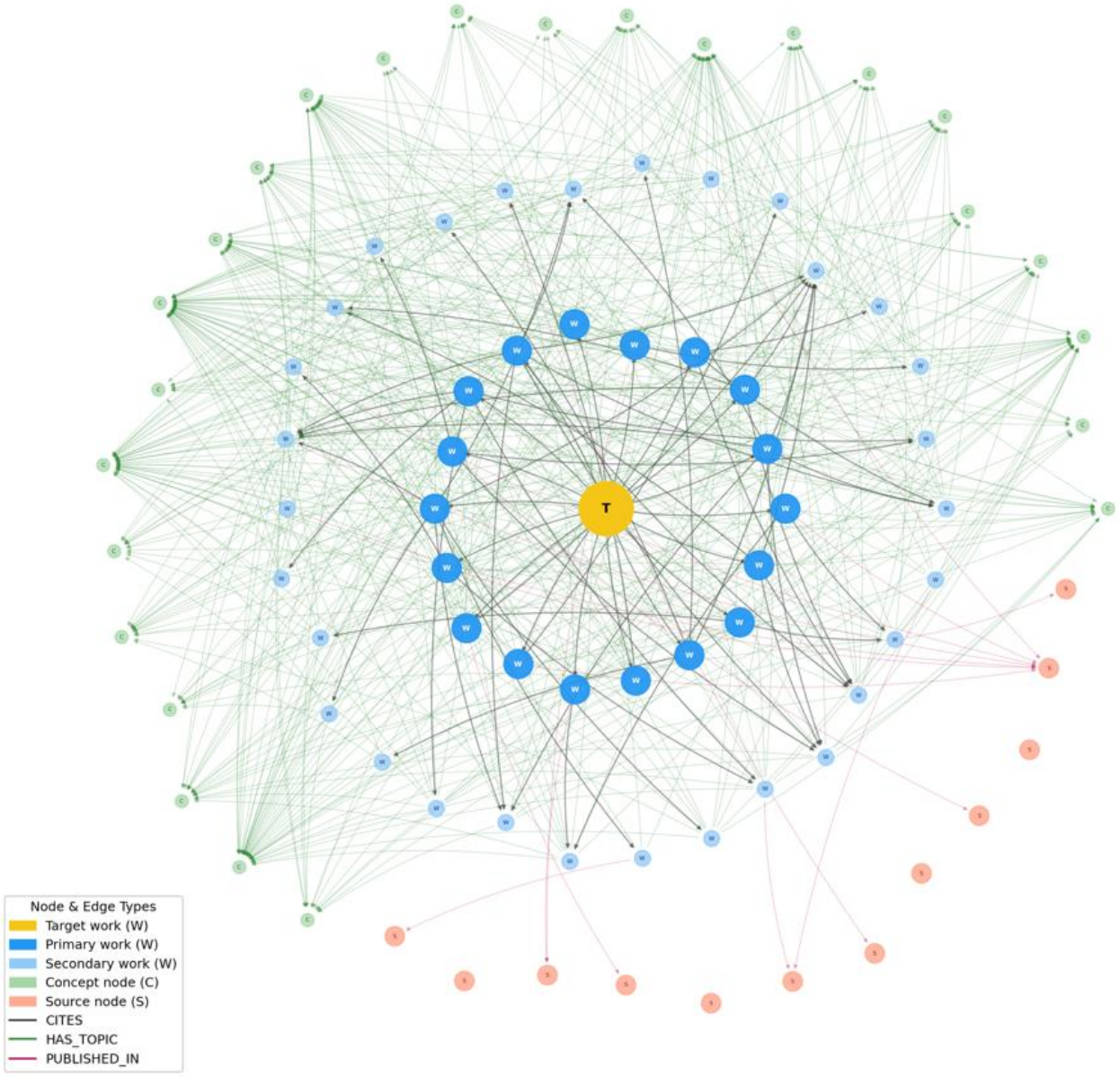

Fig. 1. Heterogeneous citation network around a single target paper, showing work nodes (W), concept nodes (C), and source nodes (S).

C. Scoring and Ranking

Both candidate pools are scored using the three-component formula in Eq. (3) and sorted in descending order. The top three entries from each pool are selected as the final recommendations, giving six results in total.

The base score differs by tier. For primary candidates directly cited by the target, the base score is:

$$BASESCORE_{\text{primary}}(c) = 1 + \frac{\text{co-citation-count}(c)}{|P|} \quad (1)$$

where $|P|$ is the total number of primary works and co-citation-count(c) is the number of primary works that also cite c. For secondary candidates cited by primary works but not by the target:

$$BASESCORE_{\text{secondary}}(\text{c}) = \frac{\text{citing-primary-count(c)}}{|\text{P}|} \qquad (2)$$

The complete scoring formula is:

$$\text{score(c)} = \text{BASESCORE (c)} + \alpha \cdot \text{J(c,w)} + \beta \cdot \text{V(c,w)} \qquad (3)$$

where $J(c,\text{w})$ is the Jaccard similarity between concept sets:

$$J(A, B) = \frac{|\text{A} \cap \text{B}|}{|\text{A} \cup \text{B}|} \qquad (4)$$

and the venue bonus $\text{V}(c,\text{w})$ is:

$$\text{V}(c,\text{w}) = \begin{cases} 1.0 \; if \; SOURCE(c) = SOURCE(w) \\ 0.5 \text{ if PUBLISHER(c)} = \text{PUBLISHER(w)} \\ 0 \text{ otherwise} \end{cases} \qquad (5)$$

Algorithm 1 Citation-Based Hybrid Recommendation
Require: Graph-derived network D, target work w,weights α, β
Ensure: Recommended work set R

source_w ← **EXTRACTSOURCE**(w)
publisher_w ← **EXTRACTPUBLISHER**(w)
C(w) ← **EXTRACTCONCEPTS**(w)
primary_w ← **EXTRACTPRIMARYCITATIONS**(w)
P ← |primary_w|
primary_cands ← ∅
secondary_cands ← ∅
for all candidate work c ∈ D such that $c \neq$ w **do**
  if c ∈ primary_w **then**
    base ← 1 + co-citation-count(c, primary_w) / P
  else
    base ← citing-primary-count(c, primary_w) / P
  end if
  J ← |CONCEPTS(c) ∩ C(w) | / |CONCEPTS(c) ∪ C(w)|
  sem_bonus ← $\alpha \cdot J$
  if SOURCE(c) = source_w **then**
    venue_bonus ← $\beta \cdot 1.0$
  else if PUBLISHER(c) = publisher_w **then**
    venue_bonus ← $\beta \cdot 0.5$
  else
    venue_bonus ← 0
  end if
  score ← base + sem_bonus + venue_bonus
  if c ∈ primary_w **then**
    primary_cands ← primary_cands ∪ {⟨c, score⟩}
  else
    secondary_cands ← secondary_cands ∪ {⟨c, score⟩}
  end if
end for
P_out ← **TOP-3**(primary_cands, key = score)
S_out ← **TOP-3**(secondary_cands, key = score)
R ← **FETCHMETADATA**(P_out ∪ S_out)
return R

D. Evaluation Protocol

Four structural ablations were evaluated: structural only ($\alpha = \beta = 0$), structural + semantic ($\alpha = 0.3$, $\beta = 0$), structural + venue ($\alpha = 0$, $\beta = 0.2$), and the full hybrid ($\alpha = 0.3$, $\beta = 0.2$). A separate topic-only baseline ranks candidates by Jaccard alone and has no structural base. Pairwise top-k Jaccard is reported as list overlap/ranking sensitivity. Independent evaluation uses TF–IDF cosine over English abstract unigrams and bigrams. For hidden-reference validation, each of 358 loaded direct citations was removed in turn and ranked as a possible secondary candidate using the remaining direct references; Recall@k and mean reciprocal rank (MRR) treat non-generated candidates as misses and are also reported conditionally on candidate coverage.

IV. RESULTS AND DISCUSSION

A. Results

We evaluated ten target articles from citation-recommendation research. Table II reports their identities and metadata, avoiding an opaque aggregate-only test description. Candidate counts refer to works successfully loaded from the archived OpenAlex snapshot after the configured caps were applied.

TABLE II. TARGET ARTICLES AND ARCHIVED TEST-SET METADATA

Target article	Field; year; citations; candidates (P/S)
W1850494429 — Context-Based Collaborative Filtering for Citation Recommendation	Recommender Systems and Techniques; 2015; 124; 30/521
W2002642763 — Citation recommendation without author supervision	Topic Modeling; 2011; 123; 32/500
W2740085866 — Paper recommendation using citation proximity in bibliographic coupling	Recommender Systems and Techniques; 2017; 23; 16/298
W2787905871 — Content-Based Citation Recommendation	Topic Modeling; 2018; 162; 33/441
W2911926823 — A Scalable Hybrid Research Paper Recommender System for Microsoft Academic	Recommender Systems and Techniques; 2019; 68; 28/450
W3201602502 — Citation recommendation using semantic representation of cited papers' relations and content	Topic Modeling; 2021; 37; 50/885
W3215214493 — Graph Neural Collaborative Topic Model for Citation Recommendation	Topic Modeling; 2021; 33; 50/795
W4206089953 — Enhancing citation recommendation using citation network embedding	Recommender Systems and Techniques; 2022; 33; 50/789
W4400698163 — Scientific paper recommender system using deep learning and link prediction in citation network	Advanced Text Analysis Techniques; 2024; 2; 23/486
W4401372619 — Research paper recommendation system based on multiple features from citation network	Recommender Systems and Techniques; 2024; 9; 46/849

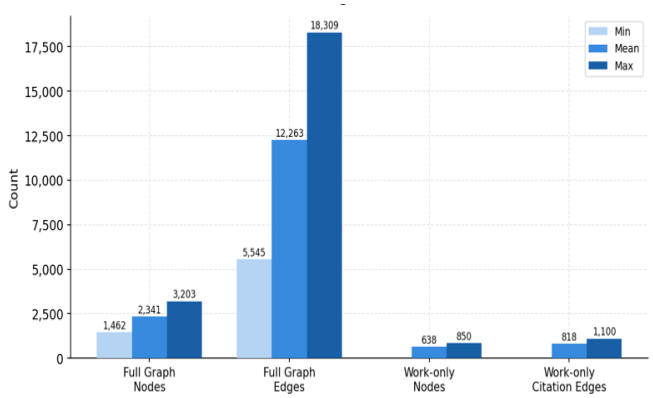

Fig. 2. Network size statistics across ten target articles: full graphs (left) and work-only subgraphs (right).

The heterogeneous networks contained 1,462–3,203 nodes (mean 2,341.7) and 5,545–18,309 edges (mean 12,263.1), with mean density 0.0022 (Fig. 2). Work-only citation subgraphs averaged 638.2 nodes and 818.7 directed edges. Connectivity and path statistics were computed after converting the work-only directed graph to an undirected projection, which is equivalent to weak-connectivity analysis. All ten projections contained one weakly connected component and had mean shortest-path length 3.77. Because every graph is a two-hop ego network grown outward from one target, this connectedness and the short paths are partly consequences of construction and should not be generalized to the global OpenAlex citation graph.

Concept overlap under the full model averaged 0.210 for primary and 0.187 for secondary recommendations, with 6.30 and 6.13 shared concepts, respectively. These values describe topical coherence but are not independent evidence of recommendation quality because the same concept Jaccard signal contributes to ranking. Independent abstract and hidden-reference results are therefore reported below.

Across the four ablation configurations, mean pairwise top-3 list overlap was 0.567 for primary and 0.443 for secondary recommendations. This statistic measures configuration-to-configuration membership overlap, not temporal or sampling stability. Lower secondary overlap indicates greater ranking sensitivity to the semantic and venue bonuses in the broader candidate pool.

The comparisons separate two different ideas that were previously conflated. The topic-only baseline ranks only by concept Jaccard and contains no structural base score. By contrast, the ablation settings named structural + semantic and structural + venue retain the structural base while activating one bonus. The full hybrid is intended to balance structural grounding and topical alignment rather than maximize the Jaccard quantity that appears in its own scoring formula.

B. Discussion

The results broadly support the viability of a lightweight hybrid approach to citation-based recommendation. The combination of edge-based structural scoring with small semantic and venue bonuses turns out to be enough to produce topically coherent suggestions without requiring expensive embedding computations or user interaction data.

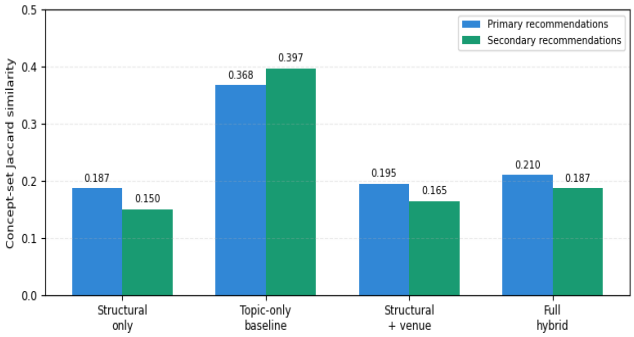

Fig. 3. Mean concept-set Jaccard across structural-only, topic-only, structural + venue, and full-hybrid configurations.

The topic-only baseline reaches concept Jaccard values of 0.368 (primary) and 0.397 (secondary), as expected because it directly optimizes that same quantity. The structural-only configuration scores 0.187 and 0.150, the structural + venue configuration 0.195 and 0.165, and the full hybrid 0.210 and 0.187 (Fig. 3). These results describe the semantic–structural trade-off; they do not by themselves establish superiority of the full method.

Precision@k, Recall@k, and NDCG ordinarily require independently judged relevant items. No human relevance labels exist for the ten target papers, so concept overlap remains a diagnostic proxy. We address this limitation with two additional tests: abstract-level TF–IDF cosine similarity, which is not used by the scorer, and leave-one-out recovery of known direct citations.

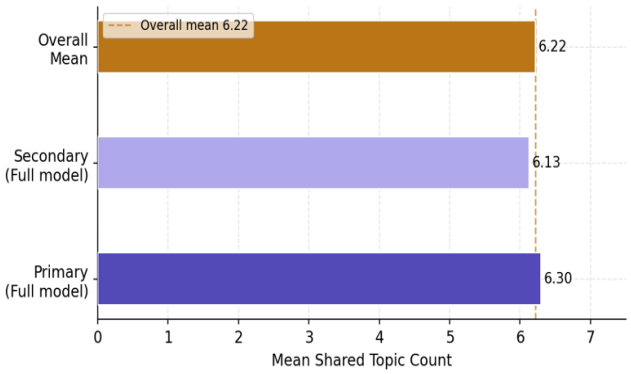

Fig. 4. Mean shared topic counts for primary, secondary, and overall recommendations under the full model.

Across the full-model recommendations, mean concept Jaccard was 0.199 and mean shared-concept count was 6.22. Primary and secondary recommendations shared 6.30 and 6.13 concepts with their targets (Fig. 4). These are internal topical-coherence indicators and are interpreted together with, rather than as substitutes for, the independent validation measures.

The topic-only baseline produces the highest concept Jaccard because it ranks directly on that measure and ignores citation structure. The full model accepts a modest reduction in concept overlap to retain structural evidence from the citation neighborhood. Whether that trade-off improves usefulness ultimately requires relevance judgements from researchers.

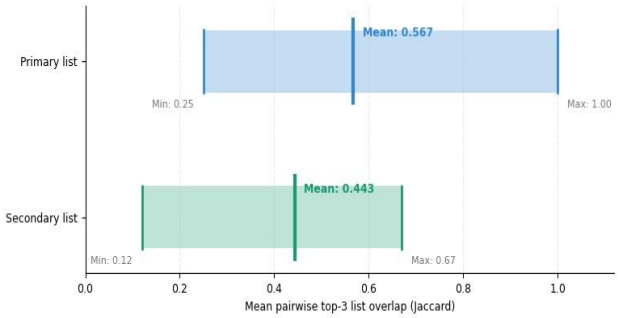

Fig. 5. Mean pairwise top-3 list overlap across the four structural ablation configurations.

Figure 5 summarizes pairwise membership overlap across the four structural ablation configurations. Primary lists overlap more because direct-reference membership and the 1.0 base term constrain their ranking; secondary lists are more sensitive because many two-hop candidates have similar structural scores. The bonuses therefore change membership most strongly in the discovery tier, but this result should be called ranking sensitivity or list overlap rather than stability.

C. Independent Validation

Abstract TF–IDF cosine provides a semantic measure that is not used in the scoring formula. For available pairs, the full hybrid averaged 0.045 (17 pairs) for primary and 0.029 (18 pairs) for secondary recommendations. Structural-only values were 0.046 (21 pairs) and 0.026 (17 pairs); structural + semantic values were 0.051 and 0.028. Thus, the semantic bonus improves primary abstract alignment in the corresponding ablation, while the full model remains comparable to the structural backbone and shows a small secondary increase. Missing abstracts prevent complete paired coverage and are reported rather than imputed.

TABLE III. INDEPENDENT ABSTRACT TF–IDF COSINE (AVAILABLE TOP-3 PAIRS)

Configuration	Primary mean (n)	Secondary mean (n)
Structural only	0.046 (21)	0.026 (17)
Structural + semantic	0.051 (21)	0.028 (20)
Structural + venue	0.039 (17)	0.023 (16)
Full hybrid	0.045 (17)	0.029 (18)

The leave-one-out test masked each loaded direct citation once. Of 358 held-out citations, 163 (45.5%) re-entered the two-hop candidate pool. Recall@3, Recall@10, and Recall@20 were 0.067, 0.112, and 0.134 overall, with MRR 0.063. Conditional on candidate coverage, the corresponding recalls were 0.147, 0.245, and 0.294, with MRR 0.139. Recovery is therefore measurable but limited mainly by candidate generation, a finding that motivates expanding beyond strict two-hop references.

TABLE IV. LEAVE-ONE-OUT HIDDEN-REFERENCE VALIDATION

Scope	Coverage	R@3	R@10	R@20	MRR
All 358	0.455	0.067	0.112	0.134	0.063
Covered 163	1.000	0.147	0.245	0.294	0.139

D. Auxiliary Signals and Future Extensions

Beyond the core recommendation model, we also examined several auxiliary ranking signals during an extended experimental analysis, including author overlap, citation count, and venue productivity. These factors were not incorporated into the main scoring formula in Eq. (3), because the proposed method was intentionally kept lightweight, interpretable, and easy to analyze. Under these extended settings, the top-3 secondary list changed in eight of the ten test cases, compared to seven for the primary list. This suggests that the exploratory half of the output is genuinely tunable, and that future versions of the system could offer configurable weighting to let users shift the balance between thematic depth and broader discovery.

The most natural extension is an author layer. Researchers often follow specific groups or individuals rather than just topics, and incorporating authorship edges would let the system surface papers along those trajectories. Citation counts and venue-level productivity scores are likewise promising candidates for future configurable extensions. Recent work on deep learning and context-aware recommendation suggests that richer semantic representations could further improve quality [17, 18], though integrating them would require moving beyond the lightweight scoring approach that currently makes the system fast and interpretable.

V. LIMITATIONS AND REPRODUCIBILITY

A. Structural Interpretation

The reported component, clustering, and shortest-path statistics were computed on the work-only graph after citation directions were ignored. They therefore describe weak connectivity in an undirected projection. Moreover, the graphs are deliberately constructed as two-hop ego networks around a single target, so connectedness and short paths are partly induced by design. These observations characterize the sampled recommendation neighborhoods, not the global scholarly graph.

Degree distributions were heavy-tailed, and simplified maximum-likelihood exponent estimates ranged from 3.55 to 5.63 (mean 4.23). An exponent estimate alone does not validate a scale-free model: no alternative-distribution comparison, goodness-of-fit test, or optimized lower cutoff was performed. We therefore treat α only as a descriptive tail indicator and avoid claiming that the networks are definitively scale-free.

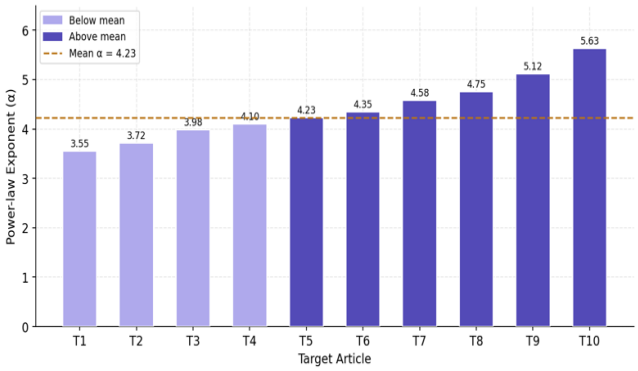

Fig. 6. Estimated power-law exponents (α) for citation subgraphs of ten target articles.

Figure 6 reports the descriptive exponent estimates for reproducibility. The coexistence of a few higher-degree nodes and many low-degree nodes is consistent with a heavy-tailed pattern, but stronger distributional claims would require formal tests against log-normal, exponential, and other alternatives on larger, independently sampled networks.

B. Evaluation Limitations

The evaluation still lacks human or curated relevance labels. Abstract text was available for eight target papers and only for subsets of their recommendations, limiting the independent TF–IDF comparison. The hidden-reference test is also bounded by candidate generation: 54.5% of held-out direct citations could not re-enter the two-hop pool at all. The reported metrics therefore support mechanism-level validation but do not replace a benchmark or user study.

C. Reproducibility

All reported core results use the March 8, 2026 OpenAlex snapshot, deterministic work-ID tie breaking, maximums of 50 primary and 30 secondary references per primary work, top-k = 3 per tier, $\alpha = 0.3$, and $\beta = 0.2$. Table II supplies target IDs and candidate counts, and the accompanying validation script exports pair-level abstract scores and all leave-one-out ranks. OpenAlex citation counts are snapshot-dependent and may change in later replications.

D. Comparative Scope

Same-pool experiments compare the full method with structural-only, topic-only, primary-neighborhood-only, structural + semantic, and structural + venue variants. Table I also positions the design against embedding, neural, temporal-graph, and LLM-assisted methods [5–7, 15, 17–20]. Direct numerical comparison with those publications would be invalid because their datasets, candidate-generation rules, and relevance labels differ; a shared benchmark evaluation remains future work.

VI. CONCLUSION

This study presented a citation-centered, two-tier recommender that separates directly cited local-reading candidates from second-hop discovery candidates and ranks both with an interpretable combination of structural, concept, and venue evidence.

Across ten targets, the method produced sparse weakly connected ego-network projections and concept-coherent lists. Independent abstract TF–IDF cosine averaged 0.045 for primary and 0.029 for secondary full-model recommendations among available pairs. In 358 leave-one-out trials, two-hop candidate coverage was 45.5%, Recall@10 was 0.112 overall, and Recall@10 was 0.245 among covered citations.

The main contribution is the transparent separation of immediate citation context from broader discovery, together with a scoring mechanism that requires no learned model or proprietary interaction history. The new validation shows that relevant known citations can sometimes be recovered after masking, while also quantifying where two-hop candidate generation fails.

The evidence should be interpreted conservatively. Abstract availability was incomplete, no human relevance judgements were collected, and the hidden-reference results reveal limited candidate coverage. The two-hop ego construction also affects connectivity and path statistics, while the power-law exponent estimates are descriptive rather than confirmatory. Future work should evaluate on a shared labelled benchmark or with researchers, expand candidate generation, test formal degree-distribution alternatives, and compare richer author, temporal, and embedding signals under the same protocol.

ACKNOWLEDGMENT

The authors thank Dr. Ahmet Anıl Müngen for providing the resources and support that made this work possible.

Authors' Contributions The authors' contributions to the paper are equal.

Statement of Conflicts of Interest
There is no conflict of interest between the authors.

Statement of Research and Publication Ethics
The authors declare that this study complies with Research and Publication Ethics.

REFERENCES

[1] J. Priem, H. Piwowar, and R. Orr, "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," arXiv preprint arXiv:2205.01833, 2022.
[2] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications, vol. 390, no. 6, pp. 1150–1170, 2011.
[3] Y. Liang and L.-K. Lee, "A Systematic Review of Citation Recommendation Over the Past Two Decades," International Journal on Semantic Web and Information Systems, vol. 19, no. 1, 2023.
[4] J. Son and S. B. Kim, "Academic paper recommender system using multilevel citation networks," Decision Support Systems, vol. 105, pp. 24–33, 2018.
[5] J. Zhang, H. Wu, Y. Lu, and X. Wang, "Citation recommendation using semantic representation of cited papers' relations and content," Expert Systems with Applications, vol. 198, 2022, Art. no. 115826.
[6] C. Pornprasit, X. Liu, and S. Tuarob, "Enhancing citation recommendation using citation network embedding," Scientometrics, vol. 127, pp. 233–264, 2022.
[7] M. Färber and A. Sampath, "HybridCite: A Hybrid Model for Context-Aware Citation Recommendation," in Proc. ACM/IEEE Joint Conf. on Digital Libraries (JCDL), 2020.
[8] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature, vol. 393, no. 6684, pp. 440–442, 1998.
[9] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 5439, pp. 509–512, 1999.
[10] D. B. West, Introduction to Graph Theory, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2001.
[11] N. J. van Eck and L. Waltman, "CitNetExplorer: A new software tool for analyzing and visualizing citation networks," Journal of Informetrics, vol. 8, no. 4, pp. 802–823, 2014.
[12] M. B. Synnestvedt, C. Chen, and J. H. Holmes, "CiteSpace II: Visualization and knowledge discovery in bibliographic databases," AMIA Annual Symposium Proceedings, pp. 724–728, 2005.

[13] X. Yang, Y. Guo, Y. Liu, and H. Steck, "A survey of collaborative filtering based social recommender systems," Computer Communications, vol. 41, pp. 1–10, 2014.
[14] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30–37, 2009.
[15] T. Kanwal, A. M. Alhanjouri, S. U. Hassan, and N. Ikram, "Research paper recommendation system based on multiple features from citation network," Scientometrics, 2024.
[16] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, "A recommendation system based on hierarchical clustering of an article-level citation network," IEEE Transactions on Big Data, vol. 2, no. 2, pp. 113–123, 2016.
[17] W. Li, H. Li, and Y. Zhang, "Scientific paper recommender system using deep learning and link prediction in citation network," Heliyon, vol. 10, no. 15, 2024, Art. no. e34685.
[18] M. A. Abbas, S. Ajayi, M. Bilal, A. Oyegoke, M. Pasha, and H. T. Ali, "A deep learning approach for context-aware citation recommendation," Journal of Ambient Intelligence and Humanized Computing, vol. 15, pp. 419–433, 2024.
[19] I. Pinedo, M. Larrañaga, and A. Arruarte, "Recent advances and trends in research paper recommender systems: A comprehensive survey," arXiv preprint arXiv:2508.08828, 2025.
[20] K. Liu, Y. Zhang, R. Pan, T. Gao, and H. Wang, "Academic literature recommendation in large-scale citation networks enhanced by large language models," arXiv preprint arXiv:2503.01189, 2025.