When Ideas Go Viral: Measuring Scholarly Novelty and Viral Influence via Citation Network Analysis

Ching Nam Hang*§, Pei-Duo Yu†, Chee Wei Tan‡, Dah Ming Chiu§
* Yam Pak Charitable Foundation School of Computing and Information Sciences, Saint Francis University, Hong Kong
† Department of Applied Mathematics, Chung Yuan Christian University, Taiwan
‡ College of Computing and Data Science, Nanyang Technological University, Singapore
§ Data Science Research Centre for Social Policies and Services, Saint Francis University, Hong Kong
Email: cnhang@sfu.edu.hk, peiduoyu@cycu.edu.tw, cheewei.tan@ntu.edu.sg, dmchiu@sfu.edu.hk

***Abstract*—Novelty is a critical attribute in academic publishing, guiding researchers toward genuinely groundbreaking contributions and shaping influential research trajectories. Motivated by the need to better quantify novelty, this paper presents an analysis of research novelty and impact in scholarly work. We develop a quantitative metric that considers both the originality and influence of research work to measure the contributions of each publication to novelty within the constructed citation network and perform a content similarity analysis to assess how closely each work builds upon its predecessors. As a preliminary study, we apply this methodology to a set of seminal publications in the domain of artificial intelligence, tracking their current citation outcomes. Results show that the inaugural work exhibits the highest novelty and garners the most citations, whereas later, more incremental works achieve varied levels of influence. We also observe that content similarity between a citing work and the work it references tends to be inversely related to novelty, reflecting whether the new work represents a significant departure or a refinement of prior research. This combined analysis offers insights into the relationship between novelty and impact in the evolution of research publications.**

***Index Terms*—Scholarly novelty, citation network, network centrality, semantic similarity, academic publishing.**

I. Introduction

In a bustling era of scientific discovery, researchers face an unprecedented deluge of publications. The volume of scientific papers continues to grow exponentially, particularly in computer science, where fast-moving fields like artificial intelligence (AI) produce thousands of papers each year [1]. For instance, accepted papers at the Conference on Neural Information Processing Systems increased from 411 in 2014 to 4,035 in 2024. Likewise, the 40th Annual AAAI Conference on Artificial Intelligence received nearly 29,000 submissions to its main track in 2025, with about 23,000 entering review, an unprecedented scale that strains even the most established scholarly processes. Amid this torrent of submissions, researchers seeking fresh ideas face significant challenges in identifying truly novel contributions. Manual literature reviews and keyword searches have become inadequate at this scale. The emergence of methods such as Deep Research in large language models further underscores the need for quantitative tools to uncover groundbreaking work. Therefore, establishing rigorous and data-driven measures of research novelty is essential for guiding authors toward truly innovative studies.

Progress toward quantifying novelty naturally leads us to the rich web of citations that link scholarly literature. Citation count has long served as a proxy for scientific impact, providing a rough indication of influence based on how many subsequent works reference a given paper. While imperfect, a citation tally reflects the reach of an idea: a highly cited paper has evidently informed or inspired many others. Yet raw counts tell only part of the story. Beneath the numbers lies the structure of the citation network itself, a vast directed graph linking papers across time [1]–[3]. Each link encodes a relationship, showing how ideas build on prior work. In essence, the citation network captures both backward-looking originality (how a work differentiates from prior literature) and forward-looking influence (how it propagates into future research). However, citation-based measures also have limitations, as citation practices differ across fields, and truly novel ideas often take time to be recognized. Despite these caveats, citations remain the most tangible record of knowledge flow and scientific uptake. By analyzing the who-cites-whom relationships, we can extract partial but meaningful indicators of the novelty of a paper, capturing both how original its ideas are and how widely those ideas spread. This suggests that a thoughtfully designed citation network analysis could serve as a compass to navigate the sea of publications, highlighting works that are not only distinctive but also influential.

In this paper, we propose constructing a citation network to quantify the novelty of academic papers. We model the body of literature as a weighted directed graph, where nodes represent papers, directed edges connect each paper to the works it cites, and each edge carries a weight denoting the content similarity between the citing and cited papers. These content-based weights allow us to distinguish, for example, a citation that heavily builds on another paper from one that merely acknowledges peripheral background, thereby integrating semantic relatedness into the network. Within this framework, we define a novelty score of a paper as the sum of two components, *influence* and *originality*, normalized to lie between zero and one. This composite perspective makes intuitive sense, as novelty is not one-dimensional; a work is not truly trailblazing unless it is both new in its ideas and capable of stimulating further research. Influence ensures that the ideas resonate and spread, while originality ensures that the work adds something unique to the canon. Prior studies of creativity echo this duality, showing that how different a work is from prior art and how much it shapes future work are distinct axes [4]. Such an understanding sets the stage for a new approach to evaluating research contributions in an age of information overload.

To measure influence, we apply the PageRank algorithm over the citation network [5]. PageRank provides a principled method to quantify scientific influence by considering not only the number of citations a paper receives but also the influence of the papers that cite it. In other words, a paper earns a high influence score if it is cited by other important papers, which aligns with the intuition that citations from prominent works should carry more weight in assessing impact. To evaluate originality, we apply rumor centrality, a graph-theoretic measure originally introduced to identify the source of information spread within a network [6], [7]. In the context of citation analysis, rumor centrality effectively identifies papers that serve as the origin of significant citation cascades, functioning like the initial source of a novel idea. This metric serves as a maximum likelihood estimator of a source in diffusion processes, capturing the essence of originality. A highly original paper will typically emerge as a root node from which many other papers ultimately derive.

The rationale behind this method is that PageRank and rumor centrality provide complementary insights. PageRank identifies the most connected and influential nodes, while rumor centrality highlights the most seminal source nodes. By combining these two measures, we develop a composite novelty score that is grounded in network theory, scalable to large citation datasets, and interpretable through well-established methods. In an environment where researchers must sift through vast amounts of published work to find truly inspiring ideas, principled and transparent metrics for assessing novelty are essential. A reliable novelty score can empower authors to discover highly original contributions, guide them toward innovative directions, and streamline the exploration of relevant literature. Ultimately, our proposed approach seeks to provide a quantitative foundation for evaluating research novelty, helping to promote more informed and efficient assessments of scientific output.

II. Related Work

Quantifying the influence of a publication has long been a central topic in bibliometrics. Early citation analysis approaches relied on raw citation counts as a proxy for influence (e.g., the journal impact factor [8]). However, more advanced metrics based on network structures have since been developed. For instance, several studies apply the PageRank algorithm to citation data to rank the influence of publications [9], [10]. This motivates our use of PageRank to assess the influence of a paper, as it accounts not only for the number of citations received but also for the identity and relative influence of the citing papers.

A substantial body of research has focused on quantifying scientific originality. For example, the work in [11] introduces a typology-based framework that classifies papers by whether their hypotheses, methods, and results are novel or previously reported. In [12], the authors propose a network-based measure using backward and forward citations to evaluate how the originality of a paper generates new research directions. The work in [13] applies divergent semantic integration, a computational measure from creativity science, to scientific abstracts and titles to analyze patterns of originality across academic fields and their correlation with citation impact.

Our work is distinct in how it combines originality and influence into a unified novelty metric. Prior methods typically focus on one facet, either the novel use of prior knowledge by a paper or its citation influence. In contrast, we define novelty as a synthesis of both aspects, combining the originality of an idea with its impact.

III. Method

In this section, we present a comprehensive framework to quantify the scholarly novelty of scientific papers by integrating the dimensions of influence and originality. The process begins with the construction of a citation network, where each node represents a scientific paper and each directed edge captures both the citation relationship and the semantic similarity inferred from topic modeling using latent Dirichlet allocation (LDA). From this network, we derive two centrality measures: Temporal Citation Rank (TCR) to reflect influence, and a weighted rumor centrality to assess originality. These two components are then combined into a single metric to represent the overall novelty of a paper.

A. Content-Enhanced Citation Network

We define a Content-Enhanced Citation Network (CECN) as a directed weighted graph $G = (V, E, W)$, where each node $v \in V$ represents an individual paper. A directed edge $(i, j) \in E$ exists if and only if paper i cites paper j, indicating a flow of ideas from the cited paper j to the citing paper i. To incorporate both the citation relationship and the semantic similarity between papers, each directed edge (i, j) is assigned a weight $w(i, j) \in W$. In this study, we examine two complementary approaches for defining these weights based on latent topic representations derived from LDA.

The first method computes the cosine similarity between the LDA topic distributions of the abstracts of paper i and paper j. Let $P_i = \big(p_{i1}, p_{i2}, \ldots, p_{iK}\big)$ and $P_j = \big(p_{j1}, p_{j2}, \ldots, p_{jK}\big)$ denote the topic distributions derived from LDA for papers i and j, respectively. The edge weight is then defined as

$$w(i, j) = \frac{P_i \cdot P_j}{\|P_i\|_2 \, \|P_j\|_2},$$

where $\| \cdot \|_2$ denotes the L_2 norm. As an alternative, we measure directional similarity using the Kullback-Leibler (KL) divergence between topic distributions. For a directed edge (i, j), we calculate the KL divergence from P_j to P_i as

$$D_{KL}(P_j \parallel P_i) = \sum_{k=1}^{K} p_{jk} \log\left(\frac{p_{jk}}{p_{ik}}\right).$$

To convert this divergence into a similarity measure, we define the edge weight as

$$w(i,j) = \exp\Big(-D_{KL}(P_j \parallel P_i)\Big).$$

This transformation ensures that a smaller divergence, indicating a closer match between the topic distributions, yields a higher weight, suggesting a stronger likelihood that the ideas from paper j are effectively inherited by paper i.

The rationale behind this construction is twofold. First, the citation structure is represented by directed edges that encode the flow of scholarly influence, where a citation from paper i to paper j indicates that i builds upon the ideas of j. Second, each edge is weighted by semantic similarity, either through the cosine similarity between LDA topic distributions or via an exponential transformation of the KL divergence, thereby embedding rich semantic information into the network. While cosine similarity provides a direct and symmetric measure of content similarity, the asymmetric nature of KL divergence captures the directional inheritance of ideas.

B. Weighted Rumor Centrality

The concept of rumor centrality was originally introduced to measure how likely a node is to act as the source of information propagation within a network [6], [7]. In an unweighted tree-structured network, the rumor centrality for a candidate source node v is defined as

$$R(v,T) = n! \cdot \prod_{u \in G} \frac{1}{t_u^v},$$

where T is a tree with n nodes and t_u^v denotes the size of the subtree rooted at node u, assuming the propagation begins at node v. This formulation effectively counts the number of distinct ways that information can spread from node v throughout the network.

In our setting, the reverse-edged CECN (rCECN) can be interpreted as an idea propagation network. In the original CECN, a directed edge (i,j) indicates that paper i cites paper j, implying that ideas flow from paper j to paper i. By reversing the edge directions, the rCECN represents the diffusion of ideas through the scholarly landscape. Within this context, the concept of rumor centrality provides a natural foundation for quantifying originality. A paper that initiates many distinct idea propagation paths is likely to be highly original, as it acts as a central source of intellectual influence.

However, unlike the original rumor centrality, which is defined for unweighted graphs, our proposed rCECN is a weighted graph, where each edge (i,j) is assigned a weight $w(i,j)$ based on the cosine similarity between the LDA-derived topic vectors of the abstracts of paper i and paper j. These weights reflect the semantic similarity between the documents. To incorporate this semantic information into our assessment of originality, we extend the classical rumor centrality framework by introducing a weighted version, which accounts for both network structure and semantic relevance.

Let $p(i,j)$ denote the propagation probability associated with the directed edge (i,j) in rCECN. This probability is derived from the edge weight $w(i,j)$, where a higher value of $w(i,j)$ indicates stronger semantic similarity between the abstracts of papers i and j, suggesting that paper j is more likely to inherit or propagate ideas from paper i. To convert this similarity into a valid probability, we normalize over all outgoing edges from node i. Specifically, if $\mathrm{Out}(i)$ denotes the set of nodes receiving an edge from i, the propagation probability is defined as

$$p(i,j) = \frac{w(i,j)}{\sum_{k \in \mathrm{Out}(i)} w(i,k)}.$$

This normalization guarantees that the sum of the propagation probabilities from node i equals 1, that is,

$$\sum_{j \in \mathrm{Out}(i)} p(i,j) = 1.$$

By converting similarity into a probability distribution, our model effectively captures the strength of idea propagation through the network. This is a critical component for computing the weighted rumor centrality, as it allows us to assess the number of distinct propagation methods and the likelihood of each method based on the semantic similarity between papers.

For a given candidate source node v and a directed spanning tree $T \in \mathcal{T}(v)$ (the set of all propagation trees rooted at v), we assign a weight to the tree as

$$P(T) = \prod_{(i,j) \in T} \big(1 + p(i,j)\big).$$

The additive constant 1 ensures that even when a propagation probability $p(i,j)$ is small, the multiplicative term $(1+p(i,j))$ does not diminish $P(T)$. This adjustment allows $P(T)$ to act as a more robust multiplier for capturing the cumulative propagation strength across multiple edges. Then, we define the weighted rumor centrality $\mathrm{R}(v)$ for node v on a general directed weighted graph G as

$$\mathrm{R}(v) = \sum_{T \in \mathcal{T}(v)} \left[P(T) \cdot \frac{n!}{\prod_{u \in G_n} \frac{1}{t_u^v}} \right].$$

This formulation preserves the original idea of counting the number of propagation methods (i.e., the combinatorial factor $n! \prod_{u \in G_n} \frac{1}{t_u^v}$), while integrating the semantic similarity information via the product of propagation probabilities along each edge. In this way, the weighted rumor centrality not only reflects the number of distinct ways an idea can propagate from a source paper v, but also accounts for the strength of the content-based relationship between papers, providing a more nuanced measure of the originality of a paper. Given that a single rCECN may contain many directed spanning trees, computing all of them becomes intractable. To address this, we introduce a Monte Carlo-based algorithm that constructs an unbiased estimator for $\mathrm{R}(v)$, as outlined in Algorithm 1. This approach scales well to large networks. Each sample has linear complexity with respect to the graph size, and the independence of samples allows the algorithm to be parallelized with ease.

Algorithm 1 Monte Carlo Estimation of Modified Weighted Rumor Centrality

Require: rCECN; source node v; number of samples M
Ensure: Unbiased estimate $\hat{R}$ of $R(v)$

$\mathrm{Reachable}(V) \leftarrow \{u \in V \mid u$ is reachable from $v\}$ via breadth-first search
Topologically order $\mathrm{Reachable}(V)$ so that if $(i \to j) \in E$, then i precedes j
For each $u \in \mathrm{Reachable}(V)$, build $\mathrm{In}(u)$ and compute $\Sigma_{\mathrm{in}}(u) = \sum_{i \in \mathrm{In}(u)} p(i,u)$
$S \leftarrow 0$
for $m \leftarrow 1$ to M **do**
  $T \leftarrow \emptyset$, $q_{\mathrm{prob}} \leftarrow 1$
  for each $u \in \mathrm{Reachable}(V)$ in topological order **do**
    if $u = v$ **then**
      $\mathrm{parent}(u) \leftarrow \mathrm{none}$
    else
      Sample $i \in \mathrm{In}(u)$ with probability $\frac{p(i,u)}{\Sigma_{\mathrm{in}}(u)}$
      $\mathrm{parent}(u) \leftarrow i$
      $T \leftarrow T \cup \{(i,u)\}$
      $q_{\mathrm{prob}} \leftarrow q_{\mathrm{prob}} \times \frac{p(i,u)}{\Sigma_{\mathrm{in}}(u)}$
    end if
  end for
  $P(T) = \prod_{(i,j) \in T} \big(1 + p(i,j)\big)$
  Compute subtree sizes t_u^v for all $u \in \mathrm{Reachable}(V)$
  $\mathrm{inv\_subtree\_prod} \leftarrow \prod_{u \in \mathrm{Reachable}(V)} (t_u^v)^{-1}$
  $w \leftarrow n! \times P(T) \times \mathrm{inv\_subtree\_prod}$
  $\mathrm{contrib} \leftarrow w \,/\, q_{\mathrm{prob}}$
  $S \leftarrow S + \mathrm{contrib}$
end for
$\hat{R} \leftarrow S/M$
return $\hat{R}$

C. Temporal Citation Rank

Building on the insights of [14] and [15], we recognize that citation patterns evolve over time. These studies suggest that papers cited soon after publication tend to exert a stronger immediate influence, as their ideas are more rapidly assimilated into the existing knowledge base. In contrast, citations that occur much later may reflect a diluted or delayed impact. The work in [16] further argues that early citations facilitate the swift incorporation of key concepts into scholarly discourse.

Motivated by these findings, we incorporate a temporal factor into our proposed influence metric. In particular, we introduce a modified version of PageRank applied to CECN, which we call Temporal Citation Rank (TCR). In TCR, the transition probability from a citing paper i to a cited paper j, denoted by $p'(i,j)$, is adjusted by a time decay factor to reflect the recency of the citation. Specifically, we define this transition probability as

$$p'(i,j) = \frac{w(i,j) \cdot \delta(i,j)}{\sum_{k \in \mathrm{Out}(i)} w(i,k) \cdot \delta(i,k)},$$

where $\mathrm{Out}(i)$ denotes the set of papers cited by paper i and the time decay factor $\delta(i,j)$ is given by:

$$\delta(i,j) = \exp\big(-\gamma\,(t(i) - t(j))\big).$$

Here, $t(i)$ and $t(j)$ represent the publication years of papers i and j, respectively, and $\gamma > 0$ is a decay parameter that controls how strongly the time difference affects the influence. This formulation ensures that citations occurring closer in time (i.e., with smaller $t(i) - t(j)$) contribute more significantly to the influence of a paper.

The TCR for a paper v is then computed as:

$$\mathrm{TCR}(v) = \frac{1-d}{N} + d \sum_{u \in \mathrm{In}(v)} p'(u,v) \cdot \mathrm{TCR}(u),$$

where N is the total number of papers in the network, $d \in (0,1)$ is the damping factor, and $\mathrm{In}(v)$ denotes the set of papers that cite paper v. This approach extends the classic PageRank by incorporating temporal dynamics, offering a more refined measure of the influence of a paper that accounts for both the structural citation relationships and the timing of citations.

D. Unified Novelty Metric

To capture the multifaceted nature of the novelty of a paper, we propose a unified metric that integrates both its influence and originality. Specifically, the novelty of paper v is defined as a weighted sum of the influence score $\mathrm{TCR}(v)$ and the originality score $\mathrm{R}(v)$:

$$\mathrm{Novelty}(v) = \alpha \cdot \mathrm{TCR}(v) + (1-\alpha) \cdot \mathrm{R}(v),$$

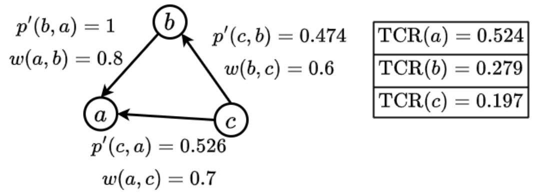
where $\alpha \in [0,1]$ is a tunable parameter that controls the balance between influence and originality. This joint model reflects the intuition that the originality of a paper can affect its long term impact, and vice versa. By combining these two dimensions, the proposed novelty metric provides a more holistic evaluation, capturing not only how quickly and widely the ideas of a paper spread (influence), but also the extent to which its content introduces new perspectives (originality). The parameter α enables flexible adjustment based on the specific context or research domain, ensuring that the measure remains both comprehensive and adaptable. Fig. 1 illustrates how our proposed framework can be applied to compute the novelty scores of three papers.

IV. Performance Evaluation

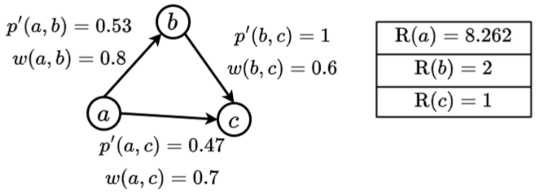
In this section, we conduct a preliminary evaluation to analyze the proposed framework for computing the novelty scores of academic papers in a citation network.

A. Setting

We construct a citation network centered on the famous "Generative Adversarial Networks" paper [17]. This seminal 2014 work is chosen as the focal point because it introduced a fundamentally new paradigm in machine learning (ML) and has since become one of the most influential ML papers of the past decade. It quickly gained widespread recognition and has been highly cited across diverse research areas, reflecting its broad and transformative impact on the field.



(a) CECN and Temporal Citation Rank



(b) rCECN and Weighted Rumor Centrality

Fig. 1. An example demonstrating the application of our framework to compute the novelty scores for three academic papers, a, b, c, where $t(a) = 2020$, $t(b) = 2021$, $t(c) = 2022$, and $\gamma = 1$. (a) The CECN, in which each directed edge $i \rightarrow j$ is labeled with its semantic similarity weight $w(i, j)$ and the time-decayed transition probability $p'(i, j)$. The resulting TCR scores reflect the influence of each paper. (b) The rCECN used to measure originality, where reversed edges represent idea propagation and retain the same weights $w(i, j)$. The computed weighted rumor centrality values $R(\cdot)$ capture the potential of each paper to serve as a source of novel ideas.

For the evaluation, we focus on a subgraph of the citation network restricted to the first year after the publication of [17] (approximately late 2014 through 2015). We include only works that cited [17] during this initial one-year period. The rationale is that early citations capture the immediate scholarly impact of the paper innovation while avoiding confounding effects from long-term trends. In other words, the first year of citations is critical for understanding how rapidly a groundbreaking idea spreads and influences subsequent research before later developments and cumulative citations set in. Among the papers that cited [17] in its first year, we further filter to include only significant early citers based on their current impact. Significance is determined by the citation counts of those papers as of 2025, effectively identifying which early-following works became highly cited themselves over the ensuing years. By concentrating on these high-impact early citations, we ensure that our evaluation examines papers that not only engaged with the idea of [17] immediately but also had substantial and lasting influence. This focus yields a set of important and representative papers that meaningfully built upon [17] in its nascent period.

From this selection process, we identify nine representative papers, [17]–[25], from the first-year citation set. For each of these nine papers, we collect key metadata, including the publication date, abstract, and list of references (i.e., the papers they cited). Using this information, we construct a directed citation network where the nine papers serve as nodes. The network contains 18 directed edges representing the citation links among these papers. Each edge is assigned a weight based on the semantic similarity between the abstracts of the citing and cited papers. These edge weights provide a quantitative measure of how closely related the content of the citation is, complementing the structural citation link with a semantic relevance score.

B. Results

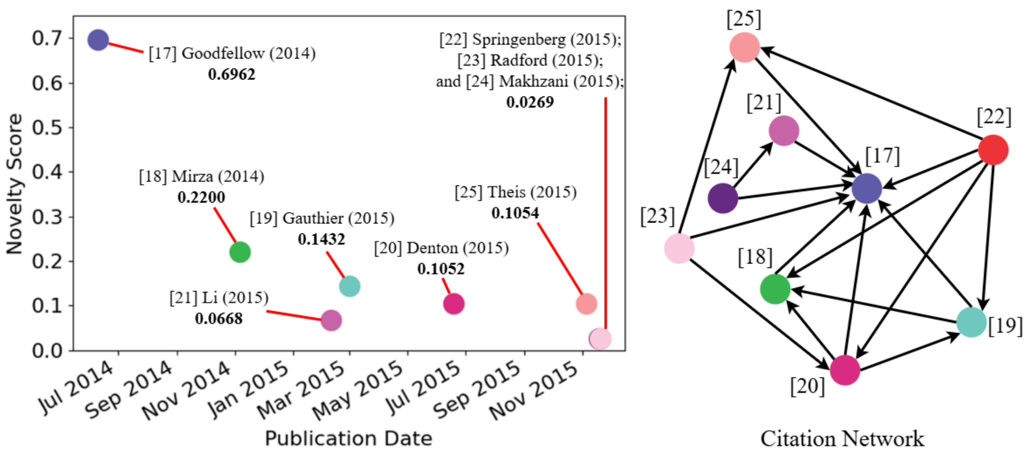
Fig. 2 summarizes the novelty scores of several influential papers and their eventual citation counts. We observe that papers with higher novelty scores tend to accumulate significantly more citations over time compared to those with lower novelty. For example, the most novel paper [17] (with a novelty score close to 0.7) eventually garnered tens of thousands of citations, vastly outperforming less novel contemporaries. In contrast, papers with minimal novelty accumulated only a few thousand citations. Overall, there is a clear positive association between the novelty of a work at publication and its long term impact, as measured by citation counts.

Fig. 3 provides a complementary perspective by illustrating the content-similarity weights of citation links among these papers. Citations directed toward the highly novel paper [17] exhibit some of the strongest content similarity scores, suggesting that subsequent research built heavily upon its core ideas. In contrast, citations involving lower-novelty papers generally show weaker content similarity, reflecting their more incremental contributions. These patterns reinforce that the novelty score effectively identifies fundamentally innovative papers that stimulate substantial follow up research. Thus, the proposed novelty score serves as an effective early indicator of the potential impact of a paper.

V. Conclusion

This preliminary study demonstrates the value of combining citation network analysis with content similarity weighting to explore the novelty of academic papers. We perform a citation network analysis of a research domain to quantitatively assess the relationship between novelty and scholarly impact. We develop a novelty metric that integrates both originality and influence to evaluate each publication, and find that the first work introduces a highly novel concept and achieves a correspondingly high impact. Subsequent works that incrementally build on this foundation exhibit lower novelty scores and show varied citation outcomes, with some attaining substantial influence through practical advances, while others remain less cited. The content similarity analysis of citing and cited pairs further illuminates how closely each follow up work aligns with prior content, explaining their relative novelty levels.

Our findings suggest that conceptual novelty and widespread impact are strongly linked in this domain, while incremental advances primarily consolidate knowledge within specific subareas. Overall, our results validate the feasibility of this integrated analytical approach and provide initial insights into the diffusion of innovation through citations. As a preliminary investigation, this work lays the groundwork for expanded studies across broader datasets and disciplines to generalize and refine these observations. In this context, our approach may also contribute to emerging AI-assisted reviewing systems by providing quantitative measures of novelty that support fairer and more transparent evaluation at scale.



Paper	Citation Count (Up to April 2025)
[17]	80,498
[23]	19,750
[18]	15,006
[24]	3,548
[20]	3,139
[25]	1,373
[21]	1,089
[22]	1,013
[19]	741

Citation Count (Sorted)

Fig. 2. Novelty scores by publication date and citation network among the selected publications.

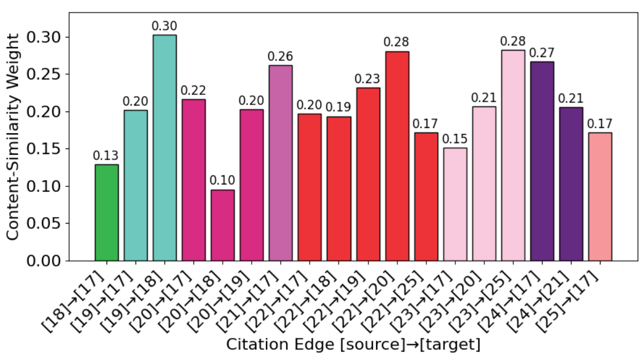


Fig. 3. Content similarity among the selected publications.

Acknowledgment

This work was supported in part by the Saint Francis University Institute Development Grant (Conference) (No. CL0129/IDGC240217), in part by the Singapore Ministry of Education Academic Research Fund RG91/22 and MOE-T2EP20224-0009, in part by the National Science and Technology Council, R.O.C, under Grant 114-2115-M-033-003-MY2, and in part by the Research Grants Council of Hong Kong under its Institutional Development Scheme Research Infrastructure Grant (No. UGC/IDS(R)11/21) and Collaborative Research Grant (No. UGC/IDS(C)11/E01/24).

References

[1] W. Zhang, J. Deakin, N. J. Higham, and S. Wang, "Etymo: A new discovery engine for AI research," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 227–230.

[2] X. Wang, L. Fu, X. Gan *et al.*, "AceMap: Knowledge discovery through academic graph," *arXiv:2403.02576*, 2024.

[3] T. Z. Fu, Q. Song, and D. M. Chiu, "The academic social network," *Scientometrics*, vol. 101, pp. 203–239, 2014.

[4] A. Elgammal and B. Saleh, "Quantifying creativity in art networks," *arXiv:1506.00711*, 2015.

[5] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web." Stanford Infolab, Tech. Rep., 1999.

[6] P.-D. Yu, C. W. Tan, and H.-L. Fu, "Epidemic source detection in contact tracing networks: Epidemic centrality in graphs and message-passing algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 2, pp. 234–249, 2022.

[7] C. W. Tan and P.-D. Yu, "Contagion source detection in epidemic and infodemic outbreaks: Mathematical analysis and network algorithms," *Foundations and Trends® in Networking*, vol. 13, no. 2-3, pp. 107–251, 2023.

[8] E. Garfield, "The impact factor," *Current Contents*, vol. 25, no. 20, pp. 3–7, 1994.

[9] P. Chen, H. Xie, S. Maslov, and S. Redner, "Finding scientific gems with Google's PageRank algorithm," *Journal of Informetrics*, vol. 1, no. 1, pp. 8–15, 2007.

[10] D. F. Gleich, "PageRank beyond the web," *SIAM Review*, vol. 57, no. 3, pp. 321–363, 2015.

[11] L. Dirk, "A measure of originality: The elements of science," *Social Studies of Science*, vol. 29, no. 5, pp. 765–776, 1999.

[12] S. Shibayama and J. Wang, "Measuring originality in science," *Scientometrics*, vol. 122, no. 1, pp. 409–427, 2020.

[13] J. H. Culbert, Y. N. Kenett, and P. Mayr, "Originality in scientific titles and abstracts can predict citation count," *arXiv:2502.01417*, 2025.

[14] W. Glänzel and U. Schoepflin, "A bibliometric study of reference literature in the sciences and social sciences." *Information Processing & Management*, vol. 35, no. 1, pp. 31–44, 1999.

[15] V. Lariviére, Y. Gingras, and E. Archambault, "The decline in the concentration of citations, 1900–2007." *J. Assoc. Inf. Sci. Technol.*, vol. 60, no. 4, pp. 858–862, 2009.

[16] J. A. Evans, "Electronic publication and the narrowing of science and scholarship," *Science*, vol. 321, no. 5887, pp. 395–399, 2008.

[17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[18] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv:1411.1784*, 2014.

[19] J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter Semester*, vol. 2014, no. 5, p. 2, 2015.

[20] E. L. Denton, S. Chintala, R. Fergus *et al.*, "Deep generative image models using a laplacian pyramid of adversarial networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[21] Y. Li, K. Swersky, and R. Zemel, "Generative moment matching networks," in *International Conference on Machine Learning*, 2015, pp. 1718–1727.

[22] J. T. Springenberg, "Unsupervised and semi-supervised learning with categorical generative adversarial networks," *arXiv:1511.06390*, 2015.

[23] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv:1511.06434*, 2015.

[24] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv:1511.05644*, 2015.

[25] L. Theis, A. v. d. Oord, and M. Bethge, "A note on the evaluation of generative models," *arXiv:1511.01844*, 2015.