https://doi.org/10.51408/issi2025_082

What Type of Methodological Novelty is More Disruptive? Evidence from Citation Classics

Linlei Xie[1], Yi Zhao[2], Chengzhi Zhang[3]

[1] xielinlei@njust.edu.cn, [2] yizhao93@njust.edu.cn, [3]zhangcz@njust.edu.cn
Nanjing University of Science and Technology, No. 200, Xiaolingwei, 210094 Nanjing (China)

Abstract

The novel contributions of academic papers encompass various aspects such as methods, theories, and results, among which methodological novelty has been proven to be more disruptive compared to other types. Methodological novelty can be further subdivided into different types. However, which type of methodological novelty is more disruptive remains to be explored. Drawing on large language models (LLMs), this study first classifies methodological novelty in academic papers into three types: first-proposed, improvement, and application. Then, the study explores the relationship between the types of methodological novelty and disruption of scientific articles. Using 928 methodological novelty articles from Citation Classics as evidence, this study finds that first-proposed methods tend to be more disruptive, while improvement and application types tend to be less disruptive. Additionally, the study explores the effect of the number of authors and institutions on disruptiveness, finding that smaller and multi-institutional teams enhance the disruption of articles. This study explores a refined classification system for methodological novelty, aiming to enrich existing approaches to scientific innovation research and deepen understanding of novelty mechanisms.

Introduction

Measuring the novelty of papers is one of the hot topics in academic research. Novelty mainly emphasizes the difference between the research contributions in the paper and previous work, requiring that the contributions have not appeared in previous papers (Dirk, 1999). Currently, most research is limited to a quantitative measurement framework based on combination novelty theory (Uzzi et al., 2013; Wang et al., 2017). However, authors' new ideas do not always stem from atypical combinations of existing ideas (Tahamtan & Bornmann, 2018). Completely new ideas often have no discernible precedents, and fundamental breakthroughs often stem from the exploration of unknown knowledge spaces (Ahuja & Morris, 2001). At present, research on novelty mainly focuses on the novelty level of papers, with less research on novelty types. Exploring the types of novelty is particularly important and necessary, as it helps us decompose, evaluate, and measure novelty, thereby helping us better understand what novelty is and what drives it (Yan et al., 2020). The measurement of novelty degree can only capture a single dimension of it. In addition, existing articles on novel types often involve theoretical research and lack empirical exploration.

Moreover, the concepts of novelty and influence have long dominated theoretical research on scientific change, attempting to explain how new ideas change the course of knowledge (Leahey et al., 2023). Researchers have long observed that papers containing more novel ideas are more likely to be in the top 1% of citation distributions (Lee et al., 2015). Furthermore, when these novel elements are combined with an appropriate amount of conventional content, these papers are more

likely to become highly cited "hot papers" (Uzzi et al., 2013). Kuhn (1962) mentioned in his book *The Structure of Scientific Revolutions* that new ideas promote paradigm shifts in science, where a new way gradually replaces an old one. So, how do these novel ideas interact with previous work to influence future knowledge flows? Leahey et al. (2023) took a new step in this field by abandoning traditional quantitative measurement methods and dividing the novel contribution of papers into three types: new theory, new method, and new result, and deeply exploring the relationship between these types and the nature of scientific impact (measured by the CD index (Consolidating/Disruptive index, CD index) (Funk & Owen-Smith, 2017)). Leahey et al. (2023) argue that the citation count of an article can only capture the quantity of scientific impact, while the level of disruption (measured by the CD index) can better capture the nature of scientific impact, that is, the changes the article makes to the subsequent knowledge flow. Their research found that new methods tend to be more disruptive, whereas new theories tend to be less disruptive, and new results do not have a robust effect on disruptiveness; (Leahey et al., 2023). In addition, among the 2540 articles in its novelty classification dataset, there are 1459 papers on methodological novelty, accounting for over 57%.

Despite this, Leahey et al.'s (2023) typology mainly focuses on the structural level of novelty in papers and cannot distinguish specific novel ways. Evaluating how new papers can change the subsequent knowledge flow is undoubtedly a topic worth exploring in depth (Leahey et al., 2023), which can provide new insights for scientific innovation. In addition, the strong disruptiveness and dominant proportion in the classification results demonstrated by the methodologically novelty papers have also aroused our interest in further exploration. Methodological novelty can not only change the direction of knowledge flow (Leahey et al., 2023) but also the direction of scientific practice, and is often an independent foundation for future scientific discoveries (Leahey, 2008; Shi et al., 2015). Furthermore, according to the connection between methods and existing methods, methodological novelty papers can be further divided into different subtypes. Papers that propose completely new methods may be more disruptive, while papers that innovatively improve or apply existing methods may have relatively lower disruptiveness. However, the relationship between methodological novelty types and disruptiveness remains unverified in existing research. Is the high disruptiveness of methodological novelty papers caused by original methods? And what is the disruptiveness in method improvement and application-oriented articles? These questions remain to be further explored. To this end, this study will further classify methodologically novelty articles and explore the relationship between their novelty types and the essence of their scientific impact. This study aims to combine Large Language Models (LLMs) for this novel classification task.

This paper mainly studies the following two questions：

RQ1*:* How effective are LLMs in the task of classifying methodological novelty in papers？

RQ2*:* What is the relationship between different types of methodological novelty papers and disruptiveness？

Related Work

This study mainly focuses on the measurement of novelty (especially novelty classification) and its relationship with disruptiveness. So, we will review previous work from three aspects: the measurement of novelty in papers, classification, and its relationship with scientific impact.

Measurement of Novelty in Papers

For measuring the novelty of papers, researchers often develop indicators based on the logic of element novelty and recombination novelty to measure whether a paper is novel or to what extent it is novel (Kaplan & Vakili, 2015). These are mainly divided into two approaches: external indicators and internal indicators.

The measurement of novelty based on external indicators basically adopts the idea of recombinant novelty. Recombination is widely considered a source of novelty in the literature. Literature related to creativity suggests that connecting distant elements is a pathway to creativity (Uzzi et al., 2013). Management-related literature shows that a new invention stems from the synthesis of multiple ideas (Fleming, 2001; File, 2001). For academic papers, if they contain new or rare combinations of knowledge elements, they are considered novel. The main source of combinatorial novelty is the combination of previously unconsolidated elements or the combination of established elements with new concepts (Mukherjee et al., 2016). The most widely used method is to treat cited journals as a knowledge element (Uzzi et al., 2013; Tahamtan & Bornmann, 2018; Shibayama et al., 2021). If a paper cites literature from two journals that are rarely cited together, it is considered novel. Citations imply that the knowledge in the cited literature is utilized by the citing literature (Matsumoto et al., 2020). Therefore, a paper that cites a rare journal pair implies the integration of rare knowledge. Uzzi et al. (2013) proposed a method to capture the combinatorial process of research papers by calculating the relative commonality of journal pairs cited by the paper. The lowest tenth percentile commonality score in a series of commonality scores of the paper is used to measure the novelty of the paper, and the median commonality score is used to measure its conventionality. This strategy has been applied and adapted in a series of subsequent related work due to its completeness and originality. Lee et al. (2015), based on Uzzi et al.'s previous work, treated the novelty of academic papers as the scarcity of pairwise combinations of previous work (i.e., references), and measured the novelty of academic papers based on the scores of cited reference pairs. Wang et al. (2017) treated scientific research as a combinatorial process, where novelty is the exploration process of combining new knowledge with existing knowledge, and measured the novelty of science based on whether the paper is the first to combine reference journals. However, journals as a knowledge element are highly aggregated units, and citation indicators designed accordingly, although to some extent reflecting the novelty of papers (Shibayama et al., 2021), their effectiveness is still controversial (Matsumoto et al., 2020).

On the other hand, content-based novelty mainly follows two logics: one is based on element novelty, and the other is based on recombinant novelty. The main limitation of recombinant novelty is that it ignores the novelty of the knowledge elements

themselves. Completely new ideas often have no discernible precedents, and fundamental breakthroughs often stem from the exploration of unknown knowledge spaces (Ahuja & Morris Lampert, 2001). Such isolated novelty events may not be captured by recombinant novelty measures. Based on the logic of element novelty, Azoulay et al. (2001) calculated the average age of MeSH keywords to assess the novelty of articles. Some studies also believe that novelty includes both the creative development of knowledge and the inheritance and reconstruction of existing or conventional knowledge. Mishra and Torvik (2016), while exploring the relationship between MeSH terms and paper novelty, proposed that in biomedicine, a single subject term is difficult to express novelty, while combined subject terms can better reflect the novelty of papers. The most influential papers often introduce some novel combinations (atypical combinations) on the basis of traditional combinations (typical combinations). Foster et al. (2015) used entity combinations to construct a chemical knowledge network, defining the combination of knowledge entities in different clusters in the knowledge network as novel. These measurement methods are relatively intuitive, but inevitably suffer from the problem of ambiguity in textual information (such as synonyms). Although they can be solved through controlled vocabulary dictionaries, building a dictionary requires a lot of expert effort, and existing dictionaries are often domain specific.

Classification of Novelty in Papers

Regarding the classification of novelty, current classifications of novelty in papers are mainly based on two ideas. One is based on the structure and content of the article, dividing novelty types according to novel content, and the other is based on the level of novelty, dividing novelty types according to the degree of it.

Classifying articles according to novel content can be seen as a multidimensional conceptualization of novelty (Rosenkopf & McGrath, 2011), allowing us to process it more richly. Early researchers mostly based their classification standards on expert experience or questionnaire and interview results, directly classifying articles into novelty types. Dirk (1999), starting from the structure of papers, believed that if the three elements of scientific work (hypothesis, method, and result) have not been reported in previous work, scientific originality can be divided into eight types (P-P-P, P-P-N, P-N-P, N-P-P, N-N-P, N-P-N, N-N-N), and asked authors to classify their papers through questionnaires. Guetzkow et al. (2004), through interviews with panel members of scholarship competitions in social sciences and humanities, divided originality into seven types: original strategy, under-researched field, original topic, original theory, original method, original data, and original result, and found that on different dimensions of originality, both social sciences and humanities generally value the originality of methods. In addition, humanists also emphasize the originality of the data used, while social scientists appreciate more types of originality (Guetzkow et al., 2004). Heinze et al. (2009) divided originality into five types: proposing new ideas, discovering new phenomena, developing new methods, inventing new tools, and integrating existing theories from new perspectives, and invited more than 400 authoritative researchers in human genetics and nanotechnology to judge the types of 20 highly creative research results in the field.

Recently, Leahey et al. (2023) divided the novel contribution of articles into new theory, new method, and new result through rule matching, and explored which type of novelty is most disruptive to knowledge flow. The research results show that new methods are often disruptive, new theories are less disruptive, and new results have no significant relationship with scientific impact.

In addition, some researchers classify articles according to the degree of novelty. Arnqvist et al. (2013) divided the novelty of articles into high incremental, low incremental, and completely novel according to the degree of connection with existing research. Sánchez et al. (2019) divided the novelty of articles into four levels according to the degree of knowledge increment: fundamental, high incremental, incremental, and low incremental. However, whether from the perspective of novel content or degree of novelty, current classifications of novelty remain at a coarse-grained level and cannot reveal the specific ways and reasons for the novelty of articles. On the basis of Leahey et al.'s (2023) classification, we further classify methodologically novelty articles to clarify how these articles, which have strong destructive effects on subsequent knowledge flows, change the knowledge process.

Relationship Between Novelty Types and Scientific Impact

The scientific impact in this study mainly comes from peer evaluations of research and academic publications (Van, 2000). Currently, the evaluation of the scientific impact of papers is mainly through external indicators, i.e., the citation situation of the articles. However, relying solely on citation counts can only capture the quantity of scientific impact, not its nature. Therefore, some studies have begun to use citation patterns to better evaluate scientific impact (Leahey et al., 2023). It is not surprising that novel contributions often have a disruptive impact on the scientific literature. Lin et al. (2022), in a large-scale study of more than 87 million scientific papers, found that novel articles are more disruptive, with the probability of disrupting science being almost twice that of traditional papers, but this is a slow process that takes ten years or more to achieve. Ruan et al. (2023), using nearly 900,000 PubMed articles published between 1970 and 2009, measured the relationship between topic combination novelty and scientific impact, and found that topic combination novelty has an inverted U-shaped relationship with citation counts, but is positively correlated with disruptiveness. So, do different types of novel articles differ in disruptiveness? Leahey et al. (2023) have conducted related research and found that there is indeed an interesting relationship between the novelty types of articles and disruptiveness.

According to the research results of Leahey et al. (2023), methodological novelty articles are more disruptive. The portability (Porter, 1996) and wide applicability of some quantitative techniques (Abbott, 2004) promote their dissemination. New methods are often introduced from other disciplines or sub-disciplines (Abbott, 2004) and applied to problems and data related to the problem at hand. The interdisciplinary nature of most methods makes potential users unfamiliar with their foundations, and scholars who introduce and adapt methods from other fields are less constrained by existing usage conventions, so they can apply them in qualitatively new ways, resulting in more disruptive research (Leahey et al., 2023). Methods can be easily

transferred to a new environment and applied to new problems without being changed in the process (Leahey, 2005).

New theory articles are more consolidating. The new theory "requires... significant changes in conventional scientific problems and technology" (Kuhn, 1962), Therefore, they should only be constructed when existing theories can no longer explain unexpected (and cumulative) observations. In addition, "any scientific theory must be evaluated together with its auxiliary hypotheses, initial conditions, etc., especially with its predecessor, so that we can see what kind of changes it has produced" (Lakatos & Musgrave, 1970). For those who apply new theories, it is also difficult to completely separate them from their foundations (Leahey et al., 2023). A new result, even if it is truly unexpected and contradicts previous research on the topic, is unlikely to be cited alone by subsequent authors (Leahey et al., 2023). In addition, new results (usually generated at the active research "frontier") must be linked to existing theories (residing in the consensus and paradigm "core") to be recognized and understood (Cole, 1983).

We already know that articles that are novel in different elements have different disruptiveness, but in fact, even within a single type of novel article, there are still different novel patterns and strategies. For example, for methodological novelty articles, some propose an unprecedented method, some improve existing methods, and some articles innovatively apply existing methods. So, for different novel patterns within the same type of novel article, will there be some differences in disruptiveness? Subjectively, papers proposing completely novel methods should be more disruptive than improvement and application-type papers, but this speculation has not been verified. At present, no researchers have explored the potential relationship, so this paper intends to further distinguish different novel patterns in methodological novelty articles, to deeply explore how such articles change subsequent knowledge flows.

Data and Methodology

Since this study is based on Leahey et al.'s (2023) research to further subdivide the novelty types of methodological novelty papers and explores the relationship between their different subtypes and disruptiveness, we adopt the same method as theirs to first divide papers into three types: new theory, new method, and new result, and obtain the methodological novelty articles required for this study. Specifically, we use Citation Classics essays as the data source, use the synonym dictionary developed by Leahey et al. (2023), and adopt a rule-based method to obtain three types of novel sentences, thereby performing article-level novelty classification. For disruptiveness, we use the CD index (Consolidating/Disruptive index, CD index) developed by Funk and Owen-Smith (2017) and employed by Leahey et al. (2023) to measure. Leahey et al. (2023) also mentioned in the article that subsequent scholars have used this measure and re-labeled it the “disruption index” (Bornmann et al. 2020; Wu et al. 2019), but it is equivalent to Funk and Owen-Smith’s (2017) CD index. Moreover, the CD index has demonstrated robust performance across multiple validation tests conducted by Funk and Owen-Smith (2017), as well as subsequent studies adopting the metric, such as research by Wu et al. (2019) and

Azoulay et al. (2020). There are actually other indicators for measuring disruptive behavior. For example, Chen (2006) proposed Freeman's Betweenness centrality, but this indicator is suitable for identifying key nodes of cross domain connections and revealing the mediating role of knowledge flow. There are also FV index (Prabhakaran et al., 2015) and FV gradient (Lathabai et al., 2015; Prabhakaran et al. 2018), which rely on complex network path analysis and have high computational costs. In addition, there are also the multidimensional evaluation framework proposed by Bu et al. (2021) and the semantic based evaluation method proposed by Yan and Fan (2024), which lack the simplicity and practicality of the CD index and cannot effectively measure changes in knowledge flow. The CD index uses citation patterns to quantify the degree to which a focal paper increases or decreases its dependence on its predecessor papers (i.e., its cited references). The logic of its calculation is that papers with a stronger consolidating impact should increase their citations to predecessor papers, while papers with a greater disruptive impact should do the opposite. Since this indicator quantifies whether and how a paper changes the knowledge flow on which it is based, it can be conceptualized as a scientific impact indicator (Leahey et al., 2023). In this way, this study will deeply explore how methodological novelty papers with high disruptiveness and a large proportion of articles change subsequent knowledge flows. The research framework is shown in Figure 1:

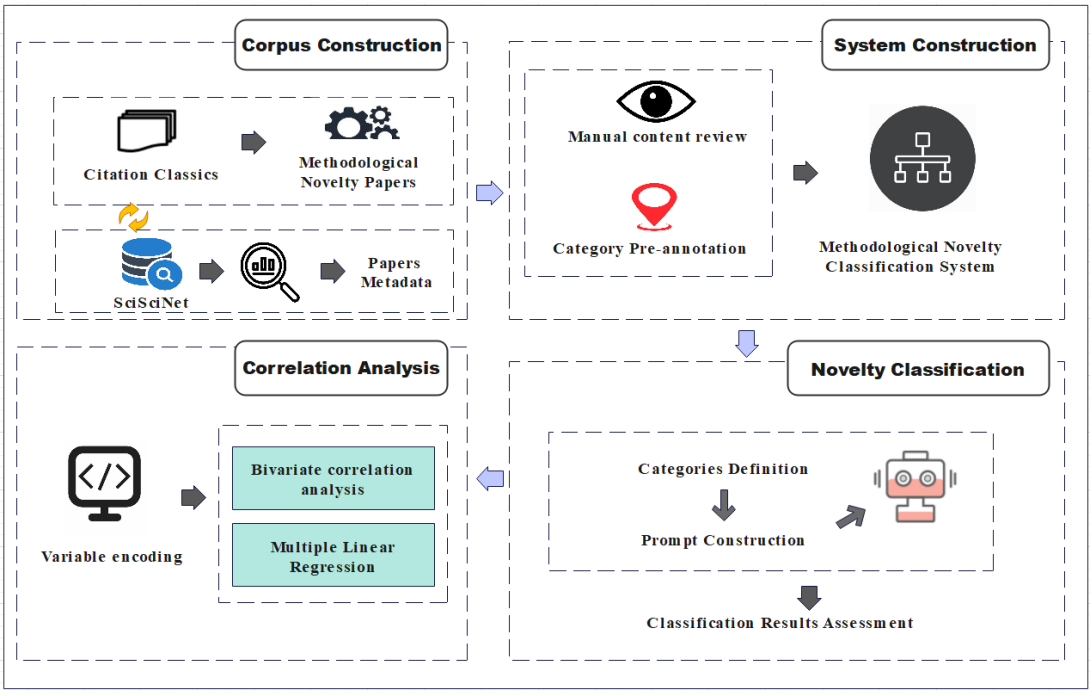

Figure 1. Framework of this study.

Data

This study uses Citation Classics as the data source. Citation Classics refer to journal articles published between 1936 and 1987 that have been cited more than a specified

number of times in Web of Science[1]. Citation Classics essays are written by Citation Classics authors and are solicited by Eugene Garfield, the developer of Web of Science. Many years after the original papers were published, these authors were invited to write a short (about one page) essay reviewing the origins of their projects, the challenges they encountered, and the reasons they believe their work has had a profound impact. These Citation Classics essays encourage scientists to "construct their own contributions," promoting the production of "intellectual self-narratives" (Gross, 2008), so they contain rich sociological information. These essays provide us with a rare humanized perspective on science, which is rarely seen in traditional journal articles or bibliometric metadata. These one-page essays span 17 years (1977 to 1993) and cover all major scientific fields. Leahey et al. (2023) OCR scanned the Citation Classics essays to form text files. By constructing a synonym table, they used a rule-based method to identify different types of novel sentences in Citation Classics essays and aggregated them at the article level to obtain the novelty type of each article. We mainly conducted further novelty classification of methodological novelty articles, we use the same method as Leahey et al. (2023) to obtain the novelty classification dataset, and separated 1459 methodologically novelty articles from it for further novelty classification in this study.

It is worth mentioning that to test whether the views of Citation Classics authors are consistent with those of the scientific community, Leahey et al. (2023) obtained the "citation context" of each Citation Classics article studied from the Microsoft Academic Graph (MAG)[2] to understand how other papers expressed themselves when citing these classic articles. They used regression models and confusion matrices to compare the Citation Classics author perspective (collected from Citation Classics essays) and the scientific community perspective (collected from MAG citation contexts), and the results confirmed that the two are consistent in their views on the novelty types of articles. This to some extent also ensures the reliability of the methodological novelty data used for further classification in this study.

Due to the limited access to Web of Science resources, we were unable to obtain the metadata of some articles, so we decided to link our data with the SciSciNet database and matched a total of 1226 articles. SciSciNet is a large-scale open data lake for the science of science research, covering over 134M scientific publications and millions of external linkages to funding and public uses (Lin et al., 2023). In addition, we further obtained the corresponding metadata of the articles through the DOI, including the title, journal name, publication year, author names and affiliations, number of co-authors and institutions, and the CD index used to measure disruptiveness.

Classification System for Methodological Novelty in Papers

This study mainly further distinguishes different novel patterns in methodological novelty articles to obtain subtypes of methodological novelty, and on this basis, explores their relationship with disruptiveness. Therefore, on the basis of Leahey et

[1] https://www.webofscience.com/wos

[2] https://www.microsoft.com/en-us/research/project/microsoft-academic-graph

al.'s (2023) data, we extract methodological novelty articles to construct the corpus to be classified in this study.

There is relatively little research on the further classification of methodological novelty articles. German scientist Mensch divided novel contribution into three types according to importance: Basic novelty, Improving novelty, and Fake novelty (Mensch, 1979). Among them, Basic novelty marks the beginning of previously unknown new products or new processes based on new scientific principles; Improving novelty refers to minor but important improvements to products, processes, and services; and Fake novelty refers to external modifications to products or processes that do not lead to changes in their consumer characteristics. In addition, the National Science Board (US) divided novelty into incremental and transformative according to the way science develops (2011). Arnqvist et al. (2013) also divided the novelty of articles into high incremental, low incremental, and completely novel according to the degree of connection with existing research. This paper synthesizes previous classifications of novelty and combines the characteristics of academic paper research methods to divide the types of methodological novelty in papers into three subtypes: First-proposed, Improvement, and Application. The category definitions are shown in Table 1.

After formulating the classification system for methodological novelty papers in this study, we manually annotated 100 articles for the construction of the subsequent evaluation dataset and as examples to be added to the Prompt to help LLMs better understand the classification task.

Table 1. Definitions of Methodological Novelty Types in Papers.

Novelty Type	*Definition*	*Example*
First-proposed	This method is first proposed in this paper and has never appeared in other scientific works before. This method is not an improvement or application of other methods, nor is it a combination of several other methods.	This paper described the first completely automatic method for colorimetric analysis.
Improvement	This method is an improvement or modification of methods that have appeared in previous scientific works, or a combination of previously proposed methods.	Our first attempts to improve the method used the incredibly laborious ion chamber technique.
Application	This method is the introduction or application of methods already proposed in previous scientific works.	It was the first attempt to apply one of the numerical hydrodynamic codes to the problem of the collapse and explosion of a star.

Classification Method for Methodological Novelty in Papers

In recent years, LLMs have shown significant progress in various challenging tasks, including solving mathematical problems (Romera-Paredes et al., 2024), proving mathematical theories (Wang et al., 2023a), and generating code to solve analytical or computational tasks (Huang et al., 2024). These advances have opened up new possibilities for using LLMs to accelerate research (Wang et al., 2023b), including research on novelty classification (Huang et al., 2025). For the novelty classification task in this study, we adopt the method of using LLMs with Few-Shot Prompting. This study uses deepseek_v3[1], llama-3[2], qianwen-2[3], and gpt-3.5-turbo[4] models to further classify methodological novelty papers. By comparing the classification results of multiple models, the model with the best performance was selected to participate in the subsequent regression analysis with the CD index.

Specifically, referring to Huang et al.'s (2025) research on LLMs, we design a prompt to elucidate the criteria and methodology for novelty classification of methodological novelty articles, and added some examples to the Prompt to assist the LLMs in understanding the methodological novelty classification task. Before using LLMs for formal classification, we randomly selected some articles and conducted LLMs classification and manual category labeling, and manually reviewed and compared the results. By analyzing the erroneous data identified by LLMs, we iteratively improve the Prompt to make it clearer in terms of task and novelty category definition, thereby improving the performance of LLMs in this task. Table 2 shows the specific content of the final Prompt, which mainly includes three parts: "###Instruction", "###Input", and "###Output". LLMs classify novelty based on the relevant text of the input article and provide classification criteria by understanding the definition of methodological novelty types and learning from a small number of annotated examples.

[1] https://www.deepseek.com
[2] https://www.llama.com
[3] https://tongyi.aliyun.com
[4] https://openai.com

Table 2. Prompt for Methodological Novelty Classification.

###Instruction	As a proficient scholar, your task is to evaluate the methodological novelty of a given paper based on the definitions of methodological novelty and its types, as well as analyzing the provided artificial classification examples. **Definition of Methodological Novelty** Methodological novelty refers to the extent to which a scientific output contributes to the knowledge of a particular research field in terms of methods. Methodological novelty exists in anything that adds new things to the knowledge of the method in the field. Methodological novelty can be first proposed, improved on existing methods, application-oriented, or even a mixture of them. **Definition of Methodological Novelty Type:** **1.First-proposed:** this method was first proposed in this paper and has never appeared in other scientific works before. This method is not an improvement or application of other methods, nor is it a combination of several other methods. **2.Improvement:** this method is an improvement or modification of methods that have appeared in previous scientific works, or a combination of previously proposed methods. **3.Application:** this method is the introduction or application of methods already proposed in previous scientific works. Please provide Methodological Novelty Type and Methodological Novelty Description. **Methodological Novelty Type:** choose Methodological Novelty Type of the given paper from [First-proposed, Improvement, Application]. If there are no suitable options, output 'None'. **Methodological Novelty Description:** write a concise paragraph (no >500 words) to explain the reasons for choosing Methodological Novelty Type. **Examples of classification of Methodological Novelty :** **Example 1:** **Sentence:** This paper described the first completely automatic method for colorimetric analysis. **Methodological Novelty Type:** First-proposed **Example 2:** **Sentence:** Our first attempts to improve the method used the incredibly laborious ion chamber technique. **Methodological Novelty Type:** Improvement **Example 3:** **Sentence:** It was the first attempt to apply one of the numerical hydrodynamic codes to the problem of the collapse and explosion of a star. **Methodological Novelty Type:** Application
###Input	**the relevant text of a given paper:** …
###Output	**Methodological Novelty Type (MNT):** … **Methodological Novelty Description (MND):** …

"###Instruction" helps LLMs understand the conceptual basis, analysis methods, and goals of the novelty evaluation of methodological novelty articles. In terms of conceptual basis, a concise definition of methodological novelty is proposed: "Methodological novelty refers to the extent to which a scientific output contributes

to the knowledge of a particular research field in terms of methods. Methodological novelty exists in anything that adds new things to the knowledge of the method in the field. Methodological novelty can be first proposed, improved on existing methods, application-oriented, or even a mixture of them." According to this definition, Methodological Novelty Types (MNT) include three categories: First-proposed, Improvement, and Application, and detailed definitions of these three categories are given. In this part, three manually classified examples are added to help LLMs understand the classification task. To achieve the evaluation goal, the LLM needs to provide the Methodological Novelty Type (MNT) and Methodological Novelty Description (MND).

"###Input" includes the relevant text of the given paper. This text is the corresponding Citation Classics essay, i.e., the one-page author's self-narrative content, which is regarded as the author's self-construction of the article's contribution many years later.

"###Output" includes MNT and MND, which are generated by the LLMs. MNT represents the novelty type of the methodological novelty paper, which can be first-proposed, improved, or application-oriented; MND is a concise paragraph to clarify the reasons for assigning the MNT.

Correlation Analysis Techniques Between Methodological Novelty Types and Disruptiveness

First, we set all three novelty types as binary variables. If an article is classified into that type, the variable is coded as 1, otherwise as 0. To deeply explore the potential relationship between novelty types in methodological novelty papers and disruptiveness, we first conducted an independent samples t-test on the relationship between method novelty types and CD index to determine whether there is a significant difference in the mean CD index between two sample populations belonging to and not belonging to a certain methodological novelty type.

To control for the influence of other variables such as the size of the paper team and the mutual influence between our three independent variables, we further conducted a multiple linear regression with the three methodological novelty types as independent variables and the CD index as the dependent variable. We built a multi-level linear regression model by successively adding control variables.

Since some articles are quite old, some articles in the SciSciNet database have publication years beyond the original time range of Citation Classics (1931-1987), and there may also be some errors in the publication year of articles in the database. To ensure the accuracy of our results, we decided not to include the publication year of articles as a control variable in our regression analysis model. Specifically, this study included key control variables to ensure the accuracy and reliability of the analysis results. The control variables include:

- Article type: whether the article is a journal paper, conference paper, or other.
- Number of authors: referring to the research of Singh & Fleming (2010) and Wu et al. (2019), the number of authors is used as an indicator of team

collaboration, which may affect the diversity and depth of novelty achievements.

- Number of institutions: institutions, as support for resources and technical elements, may affect the advancement and research depth of methodological novelty achievements.

Results

Due to the fact that there are 1226 articles in our data that can be found in the SciSciNet database for corresponding article records when obtaining other metadata of the article, of which 928 articles have a CD index. Therefore, we further extracted these articles to participate in the subsequent regression analysis.

Evaluation of the Classification Results of Methodological Novelty in Papers

We answer RQ1 in this section. We first reviewed the fine-grained classification results of methodological novelty of the four models: deepseek_v3, llama-3, qianwen-2, and gpt-3.5-turbo. We randomly selected 100 articles and manually annotated the methodological novelty categories to construct an evaluation dataset to evaluate the classification results. The evaluation mainly includes two parts: one is the accuracy of the Methodological Novelty Type (MNT) classification, and the other is the completeness and logic of the Methodological Novelty Description (MND) content. This study selected Precision(P), Recall(R), and F_1 score as evaluation indicators for MNT. These indicators are calculated based on the "confusion matrix", which is 2 * 2 for the most common binary classification. It includes True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN), as shown in the table 3.

Table 3. confusion matrix.

Real / *Predict*	*1*	*0*
1	TP	FP
0	FN	TN

The detailed definitions of the above indicators are as follows:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$\mathrm{F1} = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (3)$$

The evaluation comparison of the MNT classification results of the four models is shown in Table 4.

Table 4. Evaluation of Methodological Novelty Classification Results.

Model	*P*	*R*	F_1
deepseek_v3	0. 718	0. 710	0. 705
llama-3	**0. 815**	**0. 810**	**0. 804**
qianwen-2	0. 723	0. 670	0. 662
gpt-3.5-turbo	0. 554	0. 540	0. 513

Through comparison, we found that llama-3 outperforms the other three models in the accuracy of MNT. We further evaluated MND, taking the classification result of one article as an example, as shown in Table 5.

Table 5. Example of LMS Model Classification Results.

ocr_article	*deepseek-v3*	*llama-3*	*qianwen-2*	*gpt-3.5-turbo*
Number 51 Citation Classics December 19, Singer J M & Plotz C M. The latex fixation test. I. Application to the serologic diagnosis of rheumatoid arthritis. American Journal of Medicine 21:888-92, 1956. The authors describe a latex fixation test for the serologic diagnosis of rheu- matoid arthritis which improves on standard tests depending on erythrocyte agglutination by using biologically in - ert polyvinyl toluene and polysterene latex particles of uniform size.… (Due to space limitations, the remaining parts of this article have been omitted. For details, please refer to: https://garfield.library.upenn.edu/classics1977/A1977EC98100002.pdf)	###MNT: **Improvement** ###MND: The paper describes the development of the latex fixation test, which improves upon existing methods for the serologic diagnosis of rheumatoid arthritis. The authors replaced the previously used erythrocyte agglutination method with biologically inert polyvinyl toluene and polystyrene latex particles of uniform size. This modification addressed the limitations of earlier methods, such as the nonspecific reactions observed with collodion particles. The novelty lies in the use of latex particles, which were first developed in 1954, and their application to create a more reliable and standardized diagnostic test. While the latex particles themselves were not novel, their specific application in this context, combined with the identification of the antigen as human IgG, represents a significant improvement over prior techniques. Thus, the method novelty type is classified as ""Improvement"" because it builds upon and refines existing methodologies rather than introducing an entirely new concept."	###MNT: **Improvement** ###MND: The paper describes the development of a latex fixation test for the serologic diagnosis of rheumatoid arthritis, which improves on standard tests that depend on erythrocyte agglutination. The authors utilized biologically inert polyvinyl toluene and polystyrene latex particles of uniform size, which was a novel approach at the time. This improvement is built upon previous work, such as the use of collodion particles and the identification of the antigen as human IgG. The introduction of latex particles of uniform size, developed by Backus and Vanderhoff, was a key factor in the development of this test. The combination of this new latex with previous findings led to the creation of the latex fixation test, which has remained a standard in the field. The improvement lies in the use of a more reliable and efficient method for serologic diagnosis, making it a significant contribution to the field of immunology. Note: Although the paper builds upon previous work and introduces a new method, it is not considered ""First-propose	###MNT: **Improvement** ###MND: The method described in this paper represents an improvement on existing methods rather than a first-proposed or purely application-oriented approach. The authors, Morgan and Lazarow, developed a two-step procedure for radioimmunoassay of insulin by building upon prior research. Specifically, they utilized the foundational work of Moloney & Coval (1955), who demonstrated that guinea pigs could be reliably immunized with mammalian insulins, and Arquilla & Stavitsky (1956), who showed the feasibility of using immunological procedures for insulin assays. Additionally, Skom & Talmage (1958) had reported on the use of anti-human gamma globulin to precipitate non-precipitating insulin antibodies in insulin-resistant human serum. By integrating these previous findings, Morgan and Lazarow introduced a novel two-antibody system: in the first step, insulin forms a soluble complex with its specific antibody from immunized guinea pigs; in the second step, this complex is precipitated by an antibody to guine	###MNT: **Improvement** ###MND: The method described in the paper is an improvement on standard tests for the serologic diagnosis of rheumatoid arthritis. By utilizing biologically inert polyvinyl toluene and polystyrene latex particles of uniform size, the authors improved on the standard methods that depended on erythrocyte agglutination. This method represents an enhancement or modification of existing methods, making it a clear example of an Improvement in terms of Method Novelty."

	d"" because it improves upon existing methods r ather than proposing a c ompletely new concept. It is also not solely an "" Application"" because it involves significant novel ty and improvement in t he method itself, rather t han just applying an exis ting method to a new pr oblem. Therefore, the m ost appropriate classifica tion is "Improvement".	a pig serum obtained fro m immunized rabbits. Th is method significantly i mproved the precision a nd reproducibility of insu lin measurement, especi ally for samples containi ng less than a nanogram of insulin. The novelty li es in the combination an d refinement of existing techniques to create a m ore efficient and reliable immunoassay method. Therefore, the method n ovelty type is classified a s "Improvement" becaus e it builds on and enhan ces previously establishe d methodologies rather than proposing a compl etely new or unexplored method.

The article "describes a latex fixation test for the serologic diagnosis of rheumatoid arthritis, which improves on standard tests depending on erythrocyte agglutination by using biologically inert polyvinyl toluene and polystyrene latex particles of uniform size." Our several models all correctly classified this article as the 'Improvement' type.

For MND, the llama-3 model gave the result: "This paper describes the development of a latex fixation test for the serologic diagnosis of rheumatoid arthritis, which improves on standard tests that depend on erythrocyte agglutination. The authors utilized biologically inert polyvinyl toluene and polystyrene latex particles of uniform size, which was a new approach at the time. This improvement is built upon previous work, such as the use of collodion particles and the identification of the antigen as human IgG. The introduction of latex particles of uniform size, developed by Backus and Vanderhoff, was a key factor in the development of this test. The combination of latex with previous findings led to the creation of the latex fixation test, which has remained a standard in the field. The improvements are in the use of a more reliable and efficient method for serologic diagnosis, making it a significant contribution to the field of immunology. Note: Although the paper builds upon previous work and introduces a new method, it is not considered 'First-proposed' because it improves upon existing methods rather than proposing a completely new concept. It is also not solely an 'Application' because it involves significant novelty and improvement in the method itself, rather than just applying an existing method to a new problem. Therefore, the most appropriate classification is 'Improvement'."

It can be seen that llama-3 clearly and completely gave the reasons for classifying the article as 'Improvement', as well as the reasons for not classifying it as the other two novelty types, and the narrative is also more logical. Llama-3 not only gave the specific improvements and novel contributions of the article but also provided relevant information about the original method. In contrast, we found that other models only gave reasons for not classifying it as 'First-proposed' and did not give

reasons for not classifying it as 'Application', and their narratives were not as logical as llama-3.

In addition, although we only limited the Prompt to no more than 500 words, by comparing the output results of the models, we found that the gpt-3.5-turbo model's classification results were much shorter than other models, with an average of only about 80 words; deepseek_v3 had an average of about 150 words; llama-3 model and qianwen-2 model had an average of about 250 words.

Therefore, considering the evaluation results of MNT and MND, we finally selected the classification results of the llama-3 model as the final methodological novelty classification results of this study.

Descriptive Statistical Results of Methodological Novelty Types

According to the classification results of llama3, among the 928 articles, 191 were classified as First-proposed, 572 were classified as Improvement, and 146 were classified as Application; 19 articles were judged by the model as MNT being None, meaning they were not in our three categories. Table 6 shows the descriptive statistical results of our data, including the control variables involved in this study.

Table 6. Descriptive Statistical Results of Methodological Novelty Classification.

Variable \ *Metric*	*Mean*	*SD*	*Min.*	*Max.*
CD Index	0.15	0.23	-0.16	0.99
MNT				
First-proposed	0.21	0.41	0	1
Improvement	0.62	0.49	0	1
Application	0.16	0.36	0	1
Controls				
Team_Size	2.41	1.69	1	16
Institution_Count	1.17	0.59	1	9
Doc_Type(1- Journal;0-other)	0.97	0.17	0	1

We found that there are more disruptive papers in our data, with less consolidating papers. This may be because the Citation Classics we selected are often highly cited and influential. However, as this article aims to explore how innovative methods can change the way subsequent research cites focused papers, that is, to investigate the impact of three types of methodological novelty on disruptiveness. So this is not a problem in our research. The overall disruptiveness in the articles is not very high, with an average of only 0.15. However, there are significant differences in disruptiveness between articles, with the CD index of the article with the highest disruptiveness reaching 0.99.

In our classification results, the proportion of 'Improvement' articles is the highest, reaching 62%; Next is 'First-proposed', accounting for 21%, with the lowest proportion being 'Application' type. In addition, the number of collaborating scholars in different studies varies greatly, with the largest team size reaching 16 people, and

the standard deviation reaching 1.69; but the mean is 2.41, which means that the collaboration team usually consists of 2-3 people. The number of collaborating institutions corresponding to different articles also varies, but the number is generally small, usually consisting of 1-2 institutions.

Correlation Analysis Results Between Methodological Novelty Types and Disruptiveness

We answer RQ2 in this section. We first conducted an independent samples t-test, and the results are shown in Table 7.

Table 7. Independent Samples t-test Results Between Methodological Novelty Types and CD Index.

	Yes	*No*
First-proposed	0.199*** (n = 191)	0.135 (n = 737)
Improvement	0.147 (n = 572)	0.150 (n = 356)
Application	0.092 (n = 146)	0.159*** (n = 782)

Note: *p < .05;**p < .01;***p < .001.

According to the t-test results, articles belonging to the 'First premise' category (meanCD =0.199) are more disruptive than articles not belonging to this category (meanCD =0.135), and are significant at the p=.001 level; There is no significant difference in disruptiveness between articles that belong to and do not belong to the category of 'Improvement'; On the contrary to 'First premise', articles that do not belong to 'Application' (meanCD =0.159) are more disruptive than articles that belong to this type (meanCD =0.092), and are significant at the p=.001 level.

Due to the previous t-test not considering control variables, more accurate results need to be further estimated using multiple linear regression with the addition of control variables. In Model 1, we only studied the correlation between the three methodological novelty types and the CD index; Model 2 added "Number of Collaborating authors" (Team_Size); Model 3 added the control variable "Number of Collaborating Institutions" (Institution_Count); Model 4 added all control variables. Table 8 shows the results of the multiple linear regression analysis.

Table 8. Multiple Linear Regression Results of Methodological Novelty Types and CD Index.

	Model 1	Model 2	Model 3	Model 4
MNT				
First-proposed	0.130*	0.128*	0.276**	0.276**
	(0.055)	(0.054)	(0.099)	(0.099)
Improvement	0.077	0.078	0.239*	0.241*
	(0.053)	(0.053)	(0.097)	(0.097)
Application	0.026	0.028	0.162	0.165
	(0.055)	(0.055)	(0.098)	(0.098)
Controls				
Team_Size		-0.012**	-0.022***	-0.023***
		(0.005)	(0.006)	(0.006)
Institution_Count			0.065***	0.064***
			(0.019)	(0.019)
Doc_Type (1-Journal;0-other)				0.109*
				(0.055)
R^2	0.020	0.027	0.054	0.061
N	928	928	928	928

Note: *p < .05;**p < .01;***p < .001. Standard errors in parentheses.

According to the results of Model 1, when only considering the relationship between the three methodological novelty types and the CD index, we found that all three methodological novelty types are positively correlated with the CD index, but only the relationship between 'First-proposed' and the CD index is significant (b = 0.130*); the coefficient of 'Improvement' is slightly smaller than that of 'First-proposed' (b = 0.077), and the coefficient of 'Application' is even smaller (b = 0.026). The newly proposed method has no basis in the original method, so when subsequent scholars cite this article, many will not choose to cite the references of this article as supplementary discussions of the method, which leads to the generally higher CD index of this type of article. For 'Improvement' and 'Application' type methodological novelty articles, authors often cite related articles of the original method when introducing them, and subsequent scholars will also cite the references of this article when citing it to better introduce the principle of the method or to clarify the founder of the method, which leads to their CD index being relatively smaller than that of 'First-proposed' type articles. Therefore, although the first-proposed method has risks due to its uncertainty, once successful, its return is often very high, which can change subsequent knowledge flows and significantly promote the development of the field.

When we added the control variable 'Team_Size' in Model 2, the results still hold, 'First-proposed' is still positively correlated with the CD index and significant (b = 0.128*); the positive correlation between the other two novelty types and the CD index is still not significant, and the coefficients are smaller than that of 'First-proposed'; moreover, we found that the number of co-authors is negatively correlated with the CD index and statistically significant (b = -.012**). This result is also consistent with the conclusion of Wu et al. (2019), whose research is based on large-scale papers, patents, and software products data with various levels of influence. We have also obtained consistent conclusions in high impact Citation Classics datasets. To test whether this negative correlation is due to an inverted U-shaped correlation between the number of co-authors and the CD index, we squared the value of 'Team_Size' and participated in the regression analysis with the CD index, but the results showed that there is no inverted U-shaped correlation between the squared number of co-authors and the CD index.

We further added a control variable 'Institution_Cunt' in Model 3. We found that the positive correlation between 'First-proposed' and the CD index still holds, and the significance level has increased (b = 0.276**); in addition, we surprisingly found that the positive correlation between 'Improvement' and the CD index becomes significant (b = 0.239*), and the number of co-authors is still negatively correlated with the CD index and significant (b = -0.022**). Moreover, the number of collaborating institutions is positively correlated with the CD index and particularly significant (b = 0.065***). The more resources and broader research networks brought by multi-institutional collaboration may be the reason for this relationship.

We ultimately added the control variable 'Doc_Type' in Model 4. It can be seen that the positive correlation between the 'First proposed' type and the CD index still holds and is relatively significant (b=0.276**); the positive correlation between 'Improvement' and CD index still holds and is significant (b=0.241*); The relationship between 'Application' and CD index is still not significant. The negative correlation between 'Team_Size' and the CD index, as well as the positive correlation between 'Institution_Cunt' and the CD index, still hold and are significant. In addition, we found that journal articles are more disruptive than non journal articles. Similar to the results of Leahey et al. (2023), the impact of our methodological novelty type on disruptiveness is statistically significant, but also small. However, in reality, it is difficult to explain highly complex results such as the CD index, which rely on citation behavior not only by the authors of the article, but also by the broader scientific community (Leahey et al., 2023). Leahey et al. (2023) also converted the CD index into percentile of disruptiveness as Wu et al. (2019) did and found that the impact was comparable in scale to the team size coefficient they proposed. Moreover, our main focus is on the comparison between the three types of methodological novelty. Overall, the most compelling conclusion we have drawn is that there exists a significant positive correlation between the "First-proposed" type and the CD index, and disruptiveness of this type of articles is significantly higher than the other two types. Furthermore, disruptiveness of "Improvement" type is also higher compared to "Application" type. This result is consistent with our cognition and hypothesis.

Discussion

In this study, the high disruptiveness shown by papers proposing new methods is consistent with the cognition of the scientific community. Research proposing methods for the first time has no original method as a basis and may completely disrupt existing research paradigms or introduce completely new concepts. These methods often break existing knowledge frameworks and have greater potential to promote changes in scientific knowledge flows. On the other hand, improvement and application-type research is more about optimizing and expanding on existing knowledge, with less impact on subsequent knowledge flows.

In addition, smaller teams have advantages in both communication costs and decision-making processes, allowing them to adjust research directions more quickly, thereby helping to produce more disruptive research results. Large teams may be more inclined to adopt more conservative research methods to reduce risks and ensure the stability and reproducibility of research. The complexity of multi-scholar collaboration may hinder the implementation of innovative ideas in the research process, thereby reducing the disruptiveness of research.

Multi-institutional collaboration can integrate more resources, such as experimental equipment, data sets, and funding, and the integration of these resources helps to carry out more complex and innovative research. In addition, multi-institutional collaboration usually involves a broader research network, and different institutions may focus on different research fields. This makes multi-institutional collaboration more likely to come into contact with more research frontiers and emerging fields, and combine the latest advances in different fields to produce disruptive research results.

Theoretical Implications

Overall, this study has the following theoretical implications.

Firstly, this study enriches the research content of novelty evaluation in articles. This study is the first to propose dividing methodological novelty articles into three subtypes: first-proposed, improvement, and application. This classification method not only enriches the research content of articles' novelty evaluation but also provides a more detailed analysis framework for subsequent research. Through this classification, researchers can more deeply understand the unique characteristics and impacts of different types of novel articles. By deeply exploring the underlying mechanisms of articles' novelty, this study improves the interpretability of methodological novelty. This mechanism analysis helps to reveal the internal logic of novelty generation and provides a theoretical basis for future research.

Secondly, it expands the research perspective of scientometrics. This study analyzes the relationship between the novelty types of methodological novelty articles and their disruptiveness (CD index), thereby exploring which type of methodological novelty can better change subsequent knowledge flows. This research not only expands the research horizon of scientometrics but also provides a new perspective for understanding the dissemination and evolution of scientific knowledge. By exploring the impact mechanisms of different types of novel articles on knowledge flows, this study provides a new research direction for the field of scientometrics and

helps to further understand the role of scientific novelty in promoting the knowledge system.

Furthermore, this study validates the feasibility of artificial intelligence technology in classifying articles' novel contribution. This study applied advanced artificial intelligence technology, especially LLMs, to the task of article novelty classification, verifying its feasibility in handling complex text classification tasks. This application not only expands the methods of articles' novelty classification but also provides references for other text analysis tasks. The generalization ability and complex feature capture ability of LLMs make up for the shortcomings of manual classification caused by personal disciplinary background and subjective factors. This technical application improves the accuracy and scientificity of classification results and provides reliable tools for future research.

Practical Implications

The practical implications of this study can be summarized in the following three aspects.

Firstly, optimize scientific research management and policy-making. By understanding the impact and disruptiveness index of different types of novel articles (first-proposed, improvement, application), scientific research managers and policy makers can more scientifically allocate resources, prioritize support for research with high disruptiveness and potential impact, thereby maximizing the return on scientific research investment. In addition, the results of this study can provide a basis for the formulation of scientific research policies and educational training programs, encourage cross institutional and interdisciplinary cooperation, and support high-risk and high return research projects. Moreover, in talent cultivation, special attention should be paid to original thinking and abilities, heuristic teaching should be encouraged, and innovative practical activities should be carried out in a timely manner.

Secondly, improve the academic evaluation system. Traditional academic evaluation systems usually rely on quantitative indicators such as citation counts. Although they can reflect the dissemination scope and influence of research, they are difficult to accurately measure the novelty of research. By introducing the novelty classification of methodological novelty articles, the academic evaluation system can be improved, and the novelty, influence, and long-term value of research results can be more comprehensively evaluated. The positive correlation between the novelty type and disruptiveness of methodological novelty articles derived from research can also motivate researchers to engage in more innovative and disruptive research, encourage exploration of unknown fields, and promote scientific progress.

Thirdly, promote scientific research cooperation and achievement transformation. This study found that cross-institutional collaboration helps to produce more disruptive research results. Therefore, scientific research managers and policy makers can promote more cross-institutional and cross-disciplinary collaborative projects, promote knowledge sharing and resource integration. By identifying scientific research achievements with high disruptiveness potential, scientific research institutions can accelerate their transformation and application, promote the

combination of scientific and technological novelty and economic development, and deepen industry-university-research collaboration.

Limitations

The focus of this study is to explore the further classification of methodological novelty articles. To ensure the quality of methodological novelty articles, our research is based on the first-level classification of novelty by Leahey et al. (2023), so the data scale is relatively small. In addition, since methodological novelty articles have strong disruptiveness, we have only further divided the novelty types of such articles at present, and the further classification of theoretical novelty and result novelty remains to be explored. Furthermore, our methodological novelty classification only utilizes the currently popular and widely recognized four LLMs models with good performance, and adopts a Few-Shot Prompting approach. The performance of other classification methods and models in this classification task still needs further exploration. In addition, the novelty classification method of this study largely depends on the clear statements made by the authors of Citation Classics when reviewing the paper. But authors may implicitly describe their methodological contributions without explicitly labeling them as "first-proposed" "improvement" or "application, or may use outdated or different terminology. This situation may not be well captured and correctly classified by large models. Finally, this article used retrospective essays from Citation Classics (often decades-old papers). Since these texts are reflections written many years after original publication, the original authors' descriptions and terminology choices may no longer align clearly with present-day understandings. Terms and concepts once considered novel or groundbreaking can become standard practice or even obsolete over time. When contemporary LLMs interpret these historical reflections, they probably do so with the knowledge patterns learned from more recent textual corpora, potentially misclassifying or overlooking nuances related to past methodological innovations. This is a potential limitation of our research.

Conclusion and Future Work

This study is the first to divide methodological novelty into three types: first-proposed, improvement, and application, and introduces LLMs to classify methodological novelty. Through independent samples t-test and multiple linear regression analysis, the impact of different types of methodological novelty on disruptiveness is revealed. The study found that articles proposing new methods for the first time have higher disruptiveness, while improvement and application-type articles have relatively lower disruptiveness. In addition, we found that the number of co-authors has a significant negative correlation with disruptiveness, while the number of collaborating institutions has a significant positive correlation with disruptiveness.

At present, this study has only further classified methodological novelty papers, and will subsequently explore other novelty categories, such as the novelty classification of theoretical novelty. In the future, we will also conduct our novelty classification experiments on a larger scale of data to verify the universality of the results of this

study. And combined with more complex machine learning models to improve the accuracy and efficiency of articles' novelty classification. In the future, we will continue to study how to better explore the potential "novelty descriptions" in papers using LLMs, thereby improving the performance of LLMs in novelty classification. In addition, we noticed that methodological novelty articles are often more disruptive, which may be closely related to their portability (Porter, 1996) and interdisciplinary nature (Abbott, 2004). Therefore, articles in different disciplines may have significant differences in the way they change subsequent knowledge flows. Therefore, in the future, we will also combine disciplinary differences for more in-depth exploration.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant No.72074113) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX24_0792). Special thanks to Dr. Jina Lee for sharing the relevant data and code.

References

Abbott, A. D. (2004). Methods of discovery: Heuristics for the social sciences.

Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal*, 22(6-7), 521-543.

Arnqvist, G. (2013). Editorial rejects? Novelty, schnovelty! *Trends in ecology & evolution*, 28(8), 448-449.

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3), 527-554.

Azoulay, P., Jones, B. F., Kim, J. D., & Miranda, J. (2020). Age and high-growth entrepreneurship. American Economic Review: Insights, 2(1), 65-82.

Bornmann, L., Devarakonda, S., Tekles, A., & Chacko, G. (2020). Disruptive papers published in Scientometrics: Meaningful results by using an improved variant of the disruption index originally proposed by Wu, Wang, and Evans (2019). Scientometrics, 123(2), 1149-1155.

Bu, Y., Waltman, L., & Huang, Y. (2021). A multidimensional framework for characterizing the citation impact of scientific publications. Quantitative science studies, 2(1), 155-183.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for information Science and Technology, 57(3), 359-377.

Cole, S. (1983). The hierarchy of the sciences? *American Journal of sociology*, 89(1), 111-139.

Dirk, L. (1999). A measure of originality: The elements of science. *Social Studies of Science*, 29(5), 765-776.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, 47(1), 117-132.

File, D. (2001). The nber patent citation data file: lessons, insights and methodological tools.. *NBER Working Paper*, *8498*, 40.
Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American sociological review*, 80(5), 875-908.
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management science*, 63(3), 791-817.
Gross, N. (2008). *Richard Rorty: The making of an American philosopher*. University of Chicago Press.
Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is Originality in the Humanities and the Social Sciences? *American Sociological Review*, 69(2), 190-212.
Heinze, T., Shapira, P., Rogers, J. D., & Senker, J. M. (2009). Organizational and institutional influences on creativity in scientific research. *Research Policy*, 38(4), 610-623.
Huang, Q., Vora, J., Liang, P., & Leskovec, J. (2024). MLAgentBench: Evaluating Language Agents on Machine Learning Experimentation. *International Conference on Machine Learning*.
Huang, S., Huang, Y., Liu, Y., Luo, Z., & Lu, W. (2025). Are large language models qualified reviewers in originality evaluation? *Information Processing & Management*, 62(3), 103973.
Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10), 1435-1457.
Kogabayev, T., & Maziliauskas, A. (2017). The definition and classification of innovation. *HOLISTICA–Journal of Business and Public Administration*, 8(1), 59-72.
Kuhn, T. S. (1962). *The structure of scientific revolutions* (Vol. 962). Chicago: University of Chicago press.
Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Cambridge university press.
Lathabai, H. H., Prabhakaran, T., & Changat, M. (2015). Centrality and flow vergence gradient based path analysis of scientific literature: A case study of biotechnology for engineering. Physica A: Statistical Mechanics and its Applications, 429, 157-168.
Leahey, E. (2005). Alphas and asterisks: The development of statistical significance testing standards in sociology. *Social Forces*, 84(1), 1-24.
Leahey, E. (2008). Methodological memes and mores: Toward a sociology of social research. *Annu. Rev. Sociol*, 34(1), 33-53.
Leahey, E., Lee, J., & Funk, R. J. (2023). What types of novelty are most disruptive? *American Sociological Review*, 88(3), 562-597.
Lee, Y. N., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research policy*, 44(3), 684-697.
Lin, Y., Evans, J. A., & Wu, L. (2022). New directions in science emerge from disconnection and discord. *Journal of Informetrics*, 16(1), 101234.

Lin, Z., Yin, Y., Liu, L., & Wang, D. (2023). SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data*, 10(1), 315.

Matsumoto, K., Shibayama, S., Kang, B., & Igami, M. (2020). A validation study of knowledge combinatorial novelty.

Mensch, G. (1979). *Stalemate in Technology*. Cambridge Massachusetts: Ballinger Publishing Company.

Mishra, S., & Torvik, V. I. (2016). Quantifying Conceptual Novelty in the Biomedical Literature. *D-Lib magazine: the magazine of the Digital Library Forum*, 22(9-10), 10.1045/september2016-mishra. https://doi.org/10.1045/september2016-mishra.

Mukherjee, S., Uzzi, B., Jones, B., & Stringer, M. (2016). A new method for identifying recombinations of existing knowledge associated with high-impact innovation. *Journal of Product Innovation Management*, 33(2), 224-236.

National Science Board (US). (2011). *National Science Foundation's merit review criteria: review and revisions*. National Science Foundation.

Prabhakaran, T., Lathabai, H. H., & Changat, M. (2015). Detection of paradigm shifts and emerging fields using scientific network: A case study of Information Technology for Engineering. Technological Forecasting and Social Change, 91, 124-145.

Prabhakaran, T., Lathabai, H. H., George, S., & Changat, M. (2018). Towards prediction of paradigm shifts from scientific literature. Scientometrics, 117, 1611-1644.

Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., ... & Fawzi, A. (2024). Mathematical discoveries from program search with large language models. *Nature*, 625(7995), 468-475.

Rosenkopf, L., & McGrath, P. (2011). Advancing the conceptualization and operationalization of novelty in organizational research. *Organization Science*, 22(5), 1297-1311.

Ruan, X., Ao, W., Lyu, D., Cheng, Y., & Li, J. (2023). Effect of the topic-combination novelty on the disruption and impact of scientific articles: Evidence from PubMed. *Journal of Information Science*, 01655515231161133.

Sánchez, I. R., Makkonen, T., & Williams, A. M. (2019). Peer review assessment of originality in tourism journals: critical perspective of key gatekeepers. *Annals of Tourism Research*, 77, 1-11.

Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*, 43, 73-85.

Shibayama, S., Yin, D., & Matsumoto, K. (2021). Measuring novelty in science with word embedding. *PloS one*, 16(7), e0254034.

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? Management science, 56(1), 41-56.

Tahamtan, I., & Bornmann, L. (2018). Creativity in science and the link to cited references: Is the creative potential of papers reflected in their cited references? *Journal of informetrics*, 12(3), 906-930.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157), 468-472.

Van Houten, B. A., Phelps, J., Barnes, M., & Suk, W. A. (2000). Evaluating scientific impact. *Environmental health perspectives*, 108(9), A392-A393.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416-1436.

Wang, H., Xin, H., Zheng, C., Li, L., Liu, Z., Cao, Q., ... & Liang, X. (2023a). Lego-prover: Neural theorem proving with growing libraries. *arXiv preprint arXiv:2310.00656*.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., ... & Zitnik, M. (2023b). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972), 47-60.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. Nature, 566(7744), 378-382.

Yan, Y., Tian, S., & Zhang, J. (2020). The impact of a paper's new combinations and new components on its citation. *Scientometrics*, 122, 895-913.

Yan, Z., & Fan, K. (2024). An integrated indicator for evaluating scientific papers: considering academic impact and novelty. Scientometrics, 129(11), 6909-6929.