Analysis of Content-Annotation-Based Methods for Triadic Citation Motivation Research: Post-print

Authors: LIU Yunmei, Zhang Shuai, Si Xiangyun, Ma Feicheng

Date: 2023-04-01T16:02:49+00:00

Abstract

[Purpose/Significance] Triangular citation is a citation structure that integrates direct citation, co-citation, and bibliographic coupling relationships, possessing significant research value and potential for application expansion. To achieve a deep-level understanding and exploration of the triangular citation phenomenon, this study attempts to uncover the underlying citation motivations through content annotation. [Method/Process] First, by reviewing existing literature, a motivation annotation framework encompassing both functional citation motivations and affective citation motivations is constructed; second, procedures for acquiring triangular citation data are established, and 9,442 triangular citation relationship records are obtained from the CNKI database; finally, professional annotators assign appropriate citation motivations to the three types of citation relationships within each triangular citation record according to the predefined motivation framework. [Results/Conclusions] The experimental results reveal that: citing behaviors within triangular citation structures tend to encompass multiple functional citation motivations, whereas affective citation motivations are predominantly unique; the distribution of citation motivations for the B→A and C→A citation relationships is essentially consistent, but differs from that of C→B; the original document A is generally an important, high-impact publication within the relevant research topic, field, or discipline, tending to provide novel, pioneering, or fundamental ideas, concepts, viewpoints, or methods.

Full Text

Preamble

Volume 65, Issue 10, May 2021

Investigation and Research on Triangular Citation Motivation Based on Content Annotation

Liu Yunmei[1],[2], Zhang Shuai[1],[2], Si Xiangyun[1],[2], Ma Feicheng[1],[2]

[1]School of Information Management, Wuhan University, Wuhan 430072
[2]Wuhan University Information Resources Research Center, Wuhan 430072

Abstract: [Purpose/Significance] Triangular citation in literature is a citation structure that integrates direct citation, co-citation, and bibliographic coupling relationships, possessing significant research value and application potential. To understand and uncover triangular citation phenomena at a deeper level, this study attempts to mine the hidden citation motivations through content annotation. [Method/Process] First, by reviewing existing literature, we constructed a motivational annotation framework comprising functional citation motivations and emotional citation motivations. Second, we established procedures for acquiring triangular citation data and obtained 9,442 triangular citation relationships from the CNKI database. Finally, professionals annotated the corresponding citation motivations for the three citation relationships in each triangular citation data instance according to the preset motivational framework. [Result/Conclusion] Experimental results revealed that citing behaviors in triangular citation structures tend to contain multiple functional citation motivations, while emotional citation motivations are mostly unique. The distribution of citation motivations for B→A and C→A relationships is basically consistent, but differs from C→B. Original literature A is generally important, high-impact literature in related research topics, fields, or disciplines, tending to provide novel, pioneering, or fundamental ideas, concepts, viewpoints, or methods.

Keywords: triangular citation phenomenon; citation motivation; citation function; citation emotion; content annotation
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2021.10.006

Direct citation, bibliographic coupling [1], and co-citation [2] are currently the three most common types of citation relationships. Direct citation represents the most basic and simple relationship between documents. Bibliographic coupling and co-citation were proposed in 1963 [1] and 1973 [2] respectively, and have been widely applied in research front detection, information retrieval, knowledge structure analysis, and other fields. Co-citation analysis exhibits time lag, making it suitable for analyzing older publication sets. Moreover, co-citation relationships and their strength change over time, facilitating analysis of disciplinary development trends and thematic evolution. In contrast, bibliographic coupling analysis offers advantages in precision and quantity of identified results, and is faster in temporal dimension than co-citation analysis. However, due to the fixed nature of references, coupling relationships remain unchanged over time and are not suitable for thematic evolution analysis [3]. Therefore, both co-citation and coupling analyses can objectively connect unrelated documents and reveal the intricate cross-relationships and connectivity among a set

of documents, though these dyadic relationships reflect different perspectives on citation patterns and structural relationships.

Triangular citation integrates direct citation, co-citation, and coupling relationships into a special citation relationship, defined as follows: If documents A and B are both cited by document C, and documents A and B also have a citation relationship between them, then a triangular citation relationship is established among A, B, and C. Triangular citation is a multivariate phenomenon containing both direct citation relationships and simultaneously co-citation (A and B cited by C) and coupling relationships (B and C both cite A). The triangular citation structure includes three documents—A, B, and C—and three citation relationships: B cites A (B→A), C cites A (C→A), and C cites B (C→B). Chronologically, B→A occurs first, followed by C citing both A and B after their formal publication. Thus, the publication order is A, B, C, while the citation order is B→A first, then C→B and C→A simultaneously. Based on the roles of the three documents as citing or cited entities and their temporal positions, we designate document A as "original literature," document B as "intermediary literature," and document C as "following literature." The citation structure is illustrated in Figure 1 [Figure 1: see original paper].

In previous research, we found that this triangular structure widely exists in real-world citation networks. From multiple perspectives—including document feature analysis and citation content feature analysis—documents A, B, and C each play distinct roles and hold different values. Typically, original literature A receives the most citations, with the strongest citation intensity and earliest citation positions. Intermediary literature B serves as a crucial link in the triangular mechanism, providing connectivity. Following literature C is the most active citer in the triangular structure, driving the formation of triangular citation relationships. Furthermore, among the three citation relationships, C→A exhibits distinct characteristics from the other two direct citations (B→A and C→B), with a prevalent "indirect citation" mechanism between A and C. When following literature C differs from original literature A in language, document type, or discipline, but shares the same linguistic environment, document type, discipline, or research topic with intermediary literature B, authors of C often exhibit lazy citation motivation. They may partially or even completely copy B's citation content referencing A, resulting in highly consistent citation content between C→A and B→A.

In Chinese scientific research, some evaluation systems have already imposed requirements on citation data [8]. For instance, the evaluation criteria for China's National Natural Science Award explicitly state: "Recognition of main academic ideas and viewpoints refers to positive citations of the academic ideas, viewpoints, or methods proposed by the awardees in formally published papers and monographs by others." Therefore, citation motivation is crucial, as it can reveal citation patterns between disciplinary fields when combined with citation content features, and provide references for citation-based academic evaluation.

2. Related Research on Citation Motivation

Current research on citation motivation can be divided into three categories: theoretical classification of citation motivation; direct acquisition of authors' real citation motivations through interviews and questionnaires; and identification of authors' citation motivations based on citation content, including manual and automatic annotation.

2.1 Citation Motivation Classification

No unified standard currently exists for classifying citation motivations, with scholars holding different perspectives. For example, E. Garfield systematically divided citation motivation into 15 classic types based on the position, content, and form of cited literature in citing documents, including paying tribute to pioneers, providing background information, making substantial claims, and negating others' work—representing one of the earliest proposals for in-depth citation motivation research [9]. M. J. Moravcsik et al. established a representative classification using a dichotomous approach across four dimensions: conceptual versus operational citations, perfunctory versus evolutionary citations, and confirmative versus negational citations [10]. P. Vinkler grouped various citation motivations into two categories: professional motivation and relational motivation. Professional motivation arises from theoretical or practical content connections, while relational motivation involves authors citing to establish social connections with the academic community [11]. Similarly, Zhu Daming also divided reference citation motivations into academic and non-academic motivations, further subdividing academic motivation into 10 subcategories [12].

2.2 Interview and Questionnaire Methods

Early citation motivation research was largely constrained by full-text analysis technology, leading scholars to primarily employ questionnaires or interviews. Two basic approaches exist: First, directly asking authors about the association between their papers and cited documents. For example, T. Rong et al. surveyed 99 biologists and psychologists about their citation motivations and asked each author to assess the importance of their citations, finding that most citations aimed to provide research background information [13]. Shi Yali et al. employed grounded theory to investigate researchers' data cognitive behaviors during citation, establishing a behavioral category system [14]. Song Weixiang conducted email interviews with authors of Sleeping Beauty and Prince literature, identifying factors such as author modesty and team instability as reasons for Sleeping Beauty formation based on their responses to three citation motivation questions [15].

Second, presetting a citation motivation framework and asking citing authors to classify citations in their papers. For instance, T. A. Brooks first divided citation motivations into seven types, then surveyed researchers at the University of Iowa, asking them to identify which category(ies) each reference in

their published papers belonged to. The experiment found that authors might have multiple motivations for citing a single document [16]. M. SingSon et al. constructed a predefined questionnaire comprising demographic data and 23 statements on citation trust and source authority, inviting 100 university faculty members to complete it. Results showed that citation behavior is complex and multifaceted, with some researchers' citation behaviors influenced by social relationships [17]. Zhang Min et al. proposed research hypotheses and models, using scenario experiments combined with questionnaires to explore factors influencing researchers' citation behaviors and their internal cognitive processes [18]. Qiu Junping et al. also proposed an influencing factor model for researchers' citation behaviors, including internal and external motivations, and analyzed interrelationships among various motivations through questionnaire results [19].

2.3 Motivation Annotation Based on Citation Content

Beyond surveying scholars directly, other studies annotate citation motivations through citation content information. Manual annotation typically involves professional annotators classifying according to predefined motivation standards. For example, Ding Wenyao et al. employed content analysis to code scientific data citation behaviors in sample papers across nine dimensions, using statistical methods to describe data citation characteristics in library and information science journals and explore associations among different dimensional features [20]. S. Teufel et al. divided citation motivations into four categories—insufficient, affirmative, contrastive, and neutral—further subdivided into 12 criteria, which were then grouped into three major categories: negative, neutral, and positive. Three annotators applied these standards to 548 citations in 26 articles, achieving high inter-annotator consistency [21].

With the maturation of natural language processing and machine learning technologies, increasing numbers of scholars have attempted computer-assisted annotation of citation motivations. S. Teufel et al. utilized content semantic structures and characteristic words for citation sentiment classification, proposing an automatic citation classification method using verbal cue words [21]. Liu Shenbo et al. automatically identified subjects of citation sentences through syntactic annotation, classifying citation content based on subject-verb collocations with Teufel's verbal cue words. They found that in the BMC Bioinformatics journal, 62.88% of citations were neutral, 33.59% were positive, and only 3.53% were negative [22]. M. Roman et al. argued that deep neural network algorithms could reveal interesting findings through in-depth citation analysis. Using a citation context dataset to annotate 10 million citation contexts, they found BERT embeddings achieved the highest accuracy [23]. M. Y. Wang et al. proposed a machine learning framework combining syntactic and contextual features of citations, distinguishing important from unimportant citations by examining syntactic and contextual information, with experimental results demonstrating good classification performance [24].

In summary, citation motivations in scientific research contain rich informa-

tion. While interviews and questionnaires can directly reflect authors' citation motivations, they suffer from small datasets and operational complexity, potentially overlooking other important citation content information. Machine annotation methods for citation motivation are suitable for identifying motivations in specific datasets but carry certain indirectness and randomness. Therefore, based on existing literature, this study constructs a citation motivation annotation framework for triangular citation structures, inviting professionals to annotate motivations for collected real triangular citation relationship data based on specific citation content information. Through statistical analysis and investigation, we attempt to uncover the real citation contexts and motivations in triangular citation phenomena, further unveiling the veil of this multivariate and complex citation structure.

3. Construction of the Citation Motivation Classification Framework

By reviewing and summarizing existing domestic and international research on citation motivation, this study constructs a citation motivation classification framework. Based on the nature of citation motivations, they are divided into two major categories: functional citation motivation and emotional citation motivation, with further subdivision of each category. The specific classification framework and reference sources are shown in Table 1 .

4. Triangular Citation Data Collection

During data collection, we first established procedures for acquiring triangular citation relationships. Starting from original literature A, we identified intermediary literature B and following literature C to determine triangular citation data with A as the original literature. The steps are as follows: First, collect all citing documents of literature A to obtain the intermediary literature set {B_0, B_1, B_2, … B …}, representing multiple "B→A" citation pairs. Second, collect the citing documents for each literature B in the intermediary set. Finally, identify common documents between A's citing documents and each B 's citing documents; these common documents constitute the following literature set C , forming "C →A" and simultaneously "C →B " relationships. Thus, literature A, B , and each document in set C compose a triangular citation relationship.

This study selected the China Academic Journals Network Publishing Database (CNKI) as the data source. Considering that journal articles in CNKI have relatively standardized citation systems and records, we chose the top 20 most frequently cited journal articles from 2015-2020 under the "Library and Information Science" category as "original literature." Following the above acquisition procedures, we collected triangular citation data samples from these 20 highly-cited articles, identifying 10,875 triangular citation relationships. Data collection occurred from January 18 to February 23, 2020, using Python web scraping tools. Table 2 presents the statistical data on triangular citation rela-

tionships regarding original literature A.

From these 20 highly-cited articles as original literature, we identified 10,875 triangular citation relationships. As shown in Table 2, among these relationships, there are 20 original literature A documents, 2,505 intermediary literature B documents, and 4,533 following literature C documents. To extract specific citation positions and content from triangular citation relationships, we used Python crawlers to obtain full-text XML data for the 2,505 B documents and 4,533 C documents. Among these, 648 documents contained only reference information without specific citation positions marked in the text, making it impossible to retrieve citation location and content data. We excluded these relationships, leaving 9,442 triangular citation relationships available for experimental analysis.

We invited two professional researchers in library and information science to annotate motivations for the remaining data according to the preset citation motivation classification framework and the three citation texts (B→A, C→A, C→B) in each triangular citation instance. The annotation principles were: (1) Each "citation-cited" relationship must simultaneously include motivations from both functional and emotional categories; (2) Each "citation-cited" relationship could correspond to one or multiple motivations within functional or emotional categories. To ensure reliability, we randomly selected 10% (944) of the triangular citation relationships as an inter-annotator test. The Cohen's kappa coefficient between the two annotators was 0.75, indicating substantial agreement [36]; other annotation disagreements were resolved through discussion. The remaining 90% of triangular citation data was then evenly distributed between the two coders for annotation.

5. Analysis of Triangular Citation Motivation Survey Results

Based on the annotation results, we statistically analyzed the distribution of functional and emotional citation motivations across the three citation relationships in triangular citation structures, calculating the coverage of 18 citation motivation types across the 9,442 triangular citation relationships. The statistical results are shown in Table 3 and Table 4 .

From the annotation and statistical results, several findings emerge:

(1) **Citing behaviors in triangular citation structures tend to contain multiple functional citation motivations, while emotional citation motivations are more singular and unique.** According to our annotation of 9,442 triangular citation instances, nearly 31% of B→A relationships contained two or more functional motivations, while only about 14% contained multiple emotional motivations. Similarly, over 28% of C→A relationships included multiple functional motivations, compared to only about 9% with multiple emotional motivations. In C→B relationships, cases with two or more functional motivations reached 35.4%,

while emotional motivations accounted for only 9.6%. Thus, in authors' citing behaviors, references serve multiple functions, but citing emotions are generally unidirectional.

(2) **Comparing the distribution of 18 citation motivations across B→A, C→A, and C→B relationships reveals that B→A and C→A distributions are basically consistent, while both differ significantly from C→B.** On one hand, since both B→A and C→A represent documents B and C citing the same document A, the citing behaviors of B and C share similarities in citation topics, contexts, and methods, leading to similar motivation distributions. On the other hand, these distribution differences indirectly indicate the existence of indirect citation behavior in C→A relationships. According to the publication and citation chronological order in triangular citation structures, after B cites A, if C indirectly cites A through B, C will largely reference and borrow B→A's citation topics, contexts, and methods, resulting in basically consistent functional and emotional distributions between B→A and C→A.

Additionally, examining the distribution of "self-citation to maintain research continuity" across the three relationships reveals a striking disparity in C→A self-citation compared to the other two relationships. This further reflects the indirect citation phenomenon described above. In triangular citation structures, documents A and B, and B and C, typically involve hereditary or similar research, making self-citation more likely between A and B or B and C. However, self-citation rarely occurs between A and C, creating a self-citation barrier and "gap"—the indirect citation phenomenon in triangular citation structures.

(3) **Detailed comparison of functional motivation distributions across the three relationships shows** that motivations such as "citing references to indicate that existing literature does not cover the topic, demonstrating novelty or frontier nature" and "citing for literature review and commentary" have much higher coverage in C→B than in B→A and C→A. Conversely, motivations like "citing relevant facts or data to demonstrate research importance/necessity" and "adopting new concepts, viewpoints, or methods from references" show much higher coverage in B→A and C→A than in C→B. This indicates that in triangular citation structures, original literature A is generally important literature in related topics, fields, or disciplines, tending to provide novel, pioneering, or fundamental ideas, concepts, viewpoints, or methods, while document B primarily serves literature review and "intermediary bridge" functions.

(4) **Detailed comparison of emotional motivation distributions** reveals that the vast majority of citation emotions are neutral, with positive citations outnumbering negative ones. Thus, in triangular citation structures, citing emotions are typically implicit, with authors usually citing content without obvious emotional 倾向. Beyond neutral citations, authors tend to express positive emotions, while negative citations such as questioning or negation are relatively rare.

Comparing the proportional distribution of seven emotional motivations across B→A, C→A, and C→B relationships shows that positive motivations like "showing respect for pioneers or founders" and "affirming referenced work" have much higher coverage in B→A and C→A than in C→B. However, "used for literature review" is far more prevalent in C→B. The high proportion of positive citations from B and C to original literature A demonstrates A's higher academic value and status compared to other documents in the triangular structure.

From the emotional motivation perspective, we reach consistent conclusions with functional motivations: document B primarily serves literature review and summarization functions in triangular citation relationships.

Conclusion

Triangular citation is a special bibliometric relationship containing important "citation-cited" connotations, mechanisms, and dynamics, with significant research meaning and application value. Building upon the concept of triangular citation, this study first constructed a citation motivation annotation framework comprising functional and emotional motivations by reviewing existing literature. Second, we established procedures for acquiring triangular citation data, obtaining 9,442 triangular citation relationships from the CNKI database. Finally, professionals annotated corresponding motivations for the three citation relationships in each instance.

Experimental results show that citing behaviors in triangular citation structures tend to contain multiple functional motivations, while emotional motivations are mostly unique. The motivation distributions for B→A and C→A relationships are basically consistent but differ from C→B. Fine-grained comparison reveals that original literature A is generally important, high-impact literature in related topics, fields, or disciplines, providing novel, pioneering, or fundamental ideas, concepts, viewpoints, or methods, while document B primarily serves literature review and summarization functions.

Theoretical contributions include: Triangular citation structure combines direct citation, co-citation, and bibliographic coupling, potentially inheriting the superior performance of existing citation relationships to form a more systematic new structure achieving 1+1+1>3 effects. It enables deeper understanding of citation relationships and behaviors, such as inter-document citation mechanisms, scholars' special citation behaviors, and indirect citation motivations reflected in triangular citations. Additionally, the constructed annotation framework applies not only to triangular citation motivation annotation but also to other citation relationship studies, with potential for further optimization.

Practical contributions include: Identifying a special triangular citation mechanism where B→A and C→A distributions are basically consistent, indicating widespread indirect citation between original literature A and following literature C. This indirect citation can artificially inflate A's citation frequency, creating Matthew effect problems. Both indirect citation and Matthew effects affect

citation data accuracy, weakening the authority and credibility of citation analysis as an academic evaluation tool, violating its essential purpose. Based on this finding, we suggest appropriately reducing the weight of citation frequency for literature A in citation-based evaluations. Furthermore, C→A indirect citation constitutes a non-standard citation issue that can be identified and excluded through technical means to enhance citation analysis reliability, providing a special perspective for improper citation research. Finally, at national and institutional levels, comprehensive full-text storage and open access platforms for research literature should be established to provide multi-language, multi-type resources; researchers should also actively address barriers in interdisciplinary and cross-language literature reading to achieve standardized and reasonable citation practices.

Limitations include: Researchers from different disciplinary backgrounds may exhibit different citation behaviors. Future research could compare citation behaviors across fields to improve model accuracy and applicability. Additionally, citation contexts, attitudes, and motivations in research activities are highly complex. Both external feature analysis and citation content mining rely on researchers' subjective judgments, potentially deviating from authors' real motivations. Therefore, future work should incorporate questionnaires or interviews with researchers to obtain more objective empirical data. We plan to conduct semi-structured interviews and surveys with productive scholars to explore the probability, motivations, and influencing factors of indirect citation in triangular citations. Based on this, we will introduce avoidance strategies from a citation ecology perspective and construct an evaluation model for indirect citation to promote healthy triangular citation development and address negative phenomena like indirect citation and Matthew effects in bibliometrics.

References

[1] KESSLER M. Bibliographic coupling between scientific papers [J]. American documentation, 1963, 14(1): 10-25.
[2] SMALL HG. Co-citation in the scientific literature: a new measure of the relationship between two documents [J]. Journal of the American Society for Information Science, 1973, 24(4): 265-269.
[3] Gao Nan, Zhao Yunhua, Peng Dingyuan. Research front identification based on citation relationships and lexical analysis: a case study of artificial intelligence [J]. Journal of Intelligence, 2020, 39(4): 44-50, 13.
[4] BO J. A comparison of two bibliometric methods for mapping of the research front [J]. Scientometrics, 2005, 65(2): 245-263.
[5] BOYACK KW, KLAVANS R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? [J]. American Society for Information Science and Technology, 2010, 61(12): 2389-2404.
[6] Liu Yu, Li Wu. Research on the legitimacy of citation evaluation: a comprehensive review based on citation function and citation motivation studies [J].

Journal of Nanjing University (Philosophy, Humanities and Social Sciences), 2013, 50(6): 137-148, 157.
[7] Liu Qian, Wang Jian, Wang Jian, et al. Research on temporal changes in citation positions and their cognitive interpretation [J]. Journal of Intelligence, 2013, 32(5): 166-169, 184.
[8] Zhu Xiaojing. Practice and reflection on team-based subject service models: a case study of Renmin University of China Library [J]. Knowledge Management Forum, 2014(4): 22-26.
[9] GARFIELD E. Citation index for science [J]. Science, 1955, 122(3159): 108-111.
[10] MORAVCSIK MJ, MURUGESAN P. Some results on the function and quality of citations [J]. Social studies of science, 1975, 5(1): 86-92.
[11] VINKLER P. A quasi-quantitative citation model [J]. Scientometrics, 1987, 12(1): 47-72.
[12] Zhu Daming. Citation motivations of references [J]. Science & Technology Review, 2013, 31(22): 19-21.
[13] RONG T, MARTIN AS. Author-rated importance of cited references in biology and psychology publications [J]. Journal of documentation, 2008, 64(2): 246-272.
[14] Shi Yali, Zhao Tong, Yang Siluo. Research on researchers' data cognitive behavior from a citation perspective: based on grounded theory [J]. Information Studies: Theory & Application, 2020, 43(6): 49-55.
[15] Song Weixiang. Motivation analysis of "Prince" citing "Sleeping Beauty literature": an empirical study based on email interviews [J]. Modern Information, 2018, 38(5): 32-36.
[16] BROOKS TA. Private acts and public objects: an investigation of citer motivations [J]. Journal of the American Society for Information Science, 2010, 36(4): 223-229.
[17] SINGSON M, SUNNYS, THIYAGARAJANS, et al. Citation behavior of Pondicherry University faculty in digital environment: a survey [J]. Global knowledge memory and communication, 2020, 69(4/5): 363-375.
[18] Zhang Min, Liu Ying, Yan Weiwei. Factors influencing researchers' citation behavior and cognitive processes: a dual-path analysis based on emotional and performance outcome expectations [J]. Library Journal, 2018, 37(6): 74-84.
[19] Qiu Junping, Chen Xiaoyu, He Wenjing. Research on researchers' paper citation motivations and their interrelationships [J]. Library and Information Service, 2015, 59(9): 36-44.
[20] Ding Wenyao, Li Jian, Han Yi. Research on scientific data citation characteristics of journal papers in China's library and information science field [J]. Library and Information Service, 2019, 63(22): 118-128.
[21] TEUFEL S, SIDHARTHAN A, DAN T. Automatic classification of citation function [J]. Conference on empirical methods in natural language processing, 2006, 14(1): 103-110.
[22] Liu Shenbo, Ding Kun, Zhang Chunbo. Research on citation evaluation based on citation content nature [J]. Information Studies: Theory & Application, 2015, 38(3): 77-81.

[23] ROMAN M, SHAHID A, KHANS, et al. Citation intent classification using word embedding [J]. IEEE Access, 2021, 9(1): 9982-9995.
[24] WANG MY, ZHANG JQ, JIAO SJ, et al. Important citation identification by exploiting the syntactic and contextual information of citations [J]. Scientometrics, 2020, 125(3): 2109-2129.
[25] GARFIELD E. Can citation indexing be automated? [J]. Essays of an information scientist, 1962(1): 84-90.
[26] Li Zhuo, Zhao Mengyuan, Liu Jiahao, et al. Research on book citation motivations based on citation content [J]. Library and Information Service, 2019(3): 96-104.
[27] Ma Feng, Wu Yishan. Questionnaire survey research on paper citation motivations: taking Chinese journal research and information science community as an example [J]. Journal of Intelligence, 2009, 28(6): 9-14, 8.
[28] CHANG YW. A comparison of citation contexts between natural sciences and social sciences and humanities [J]. Scientometrics, 2013, 96(2): 535-553.
[29] Wang Wenjuan, Ma Jianxia, Chen Chun, et al. Review of citation text classification and implementation methods [J]. Library and Information Service, 2016, 60(6): 118-127.
[30] OPPENHEIM C, RENN SP. Highly cited old papers and the reasons why they continue to be cited [J]. Journal of the American Society for Information Science, 1978, 29(5): 225-231.
[31] SEN SK. A theoretical glance at citation process [J]. International forum on information and documentation, 1990, 15(1): 1-7.
[32] BONZI S, SNYDER HW. Motivations for citation: a comparison of self citation and citation to others [J]. Scientometrics, 1991, 21(2): 245-254.
[33] Lai Fangzhong. Citation motivation analysis [J]. Journal of Sichuan Police College, 2009, 21(6): 115-119.
[34] Chen Xiaoli. Comparative analysis of citation types [J]. Library and Information, 1998(4): 3-5.
[35] Peng Qiuru, Yan Sulan, Huang Shuiqing. Research on citation indicators based on full-text analysis: taking F1000 recommended papers as examples [J]. Journal of Information Resources Management, 2019, 9(4): 82-88.
[36] VIEIRA AJ, GARRETT JM. Understanding interobserver agreement: the kappa statistic [J]. Family medicine, 2005, 37(5): 360-363.

Author Contributions

Liu Yunmei: Responsible for research design and paper writing;
Zhang Shuai: Responsible for research design and paper revision;
Si Xiangyun: Responsible for research design and paper revision;
Ma Feicheng: Responsible for research design, guidance, and paper revision.

Investigation and Research on Triangular Citation Motivation Based on Content Annotation

Liu Yunmei[1],[2], Zhang Shuai[1],[2], Si Xiangyun[1],[2], Ma Feicheng[1],[2]

[1]School of Information Management, Wuhan University, Wuhan 430072
[2]Wuhan University Information Resources Research Center, Wuhan 430072

Abstract: [Purpose/Significance] Triangular citation in literature is a citation structure that integrates direct citation, co-citation, and bibliographic coupling relationships, possessing significant research value and application potential. In order to understand and uncover triangular citation phenomena at a deeper level, this study attempts to mine the hidden citation motivations through content annotation. [Method/Process] Firstly, by reviewing existing literature, we constructed a citation motivation annotation framework comprising functional and emotional citation motivations. Secondly, we established procedures for acquiring triangular citation data and obtained 9,442 triangular citation relationships from the CNKI database. Finally, professionals annotated the corresponding citation motivations for each of the three citation relationships in the triangular citation data according to the preset motivational framework. [Result/Conclusion] The experimental results found that citing behaviors in triangular citation structures tend to contain multiple functional citation motivations, while emotional citation motivations are mostly unique. The distribution of citation motivations for B→A and C→A relationships is basically consistent, but differs from C→B. Original literature A is generally important, high-impact literature in related research topics, fields, or disciplines, tending to provide novel, pioneering, or fundamental ideas, concepts, viewpoints, or methods.

Keywords: triangular citation phenomenon; citation motivation; citation function; citation emotion; content annotation

Note: Figure translations are in progress. See original paper for figures.