Postprint: Research on Methods for Analyzing Interdisciplinary Trends in Thematic Research Fields Based on Web of Science Subject Categories

Authors: Ye Chunlei

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] To better reveal the disciplinary objects and research content of interdisciplinary development in thematic research fields, a comprehensive interdisciplinary trend analysis method is proposed. [Method/Process] First, all disciplinary categories involved in the thematic field are identified, a disciplinary influence index is defined, and a disciplinary influence network is constructed. Then, centrality, structural hole, and visual analysis are conducted on the disciplinary influence network to identify core disciplinary categories. Finally, a keyword-discipline category co-occurrence network is constructed, network centrality analysis is used to obtain thematic content represented by keywords, and interdisciplinary thematic content is obtained by combining core disciplinary categories and domain expert opinions. [Results/Conclusions] Empirical results show that the proposed analysis method can to some extent reveal the interdisciplinary development trends in thematic research fields, and its effectiveness has been verified to a certain extent.

Full Text

Preamble

Research on Interdisciplinary Trend Analysis Methods for Subject Research Fields Based on Web of Science Subject Categories

Ye Chunlei
Library of Beijing University of Agriculture, Beijing 102206

Abstract

[**Purpose/Significance**] To better reveal the disciplinary objects and research content of interdisciplinary development in subject research fields, this paper proposes a comprehensive interdisciplinary trend analysis method. [**Method/Process**] First, all subject categories involved in the subject field are identified, and a Discipline Influence Index is defined to construct a discipline influence network. Then, centrality, structural holes, and visualization analyses are performed on this network to identify core subject categories. Finally, a keyword-subject category co-occurrence network is constructed, and network centrality analysis is used to obtain subject content represented by keywords, which is then combined with core subject categories and domain expert opinions to identify interdisciplinary subject content. [**Results/Conclusion**] Empirical results demonstrate that the proposed method can reveal interdisciplinary development trends in subject research fields to a certain extent, and its effectiveness has been partially validated.

Classification Number: G250
Keywords: interdisciplinary, trend analysis, network analysis, subject categories

1. Introduction

Interdisciplinarity refers to practical activities that transcend the boundaries of a known discipline and involve two or more disciplines [1]. Modern scientific research is increasingly characterized by high-level disciplinary integration, with interdisciplinary research becoming an important development trend. Many significant scientific breakthroughs, knowledge innovations, and solutions to major social problems are closely linked to interdisciplinary research [2]. Each scientific research question is typically manifested through specific subject research fields, which this paper defines as concrete research objects in scientific inquiry. Currently, researchers both domestically and internationally have conducted extensive studies on interdisciplinarity.

Citation analysis primarily reveals interdisciplinary characteristics from a meso-level perspective, but analyzing interdisciplinary development trends in subject research fields requires micro-level examination of literature content. Existing studies, whether based on citation analysis or keyword analysis, tend to focus on analyzing the interdisciplinary characteristics of specific disciplines rather than providing actionable insights for interdisciplinary development in concrete subject research fields. When analyzing interdisciplinary knowledge content, researchers often lock onto specific disciplines and extract keywords for clustering, but such results are frequently influenced by the inherent development trajectories of individual disciplines and may not authentically reflect interdisciplinary characteristics.

The Institute for Scientific Information (ISI) assigns subject categories to journals in the Web of Science database based on journal-to-journal citation patterns and expert judgment, thereby categorizing the literature published in those journals. Although ISI's subject classification system has been controversial [14], it is widely used and easily accessible [9]. This paper therefore adopts ISI's Subject Categories (SC) as its foundation, using keywords from literature within these categories as the unit of analysis to address two primary questions: (1) at the meso-level of subject categories, identifying key subject categories with high influence to facilitate interdisciplinary collaboration for subject field development; and (2) at the micro-level of knowledge content, identifying major interdisciplinary research topics to guide development directions.

2. Research Methods and Content

This research focuses on a specific subject research field, aiming to identify both key influential subject categories and relevant interdisciplinary knowledge content. The specific research framework is shown in Figure 1 [Figure 1: see original paper].

2.1 Interdisciplinary Category Analysis

Subject research fields typically involve multiple disciplines. To study interdisciplinary development, it is necessary to dynamically identify all relevant subject categories and, more importantly, recognize key categories that can provide accurate targets for interdisciplinary collaboration. This analysis primarily aims to identify highly influential core subject categories. Using keywords that represent disciplinary knowledge content as entry points, this paper defines the Discipline Influence Index to construct a discipline influence network, then applies network centrality, structural holes, and visualization analyses to identify core subject categories with significant influence on the subject research field.

2.2 Interdisciplinary Knowledge Content Analysis

Keyword-subject category co-occurrence network analysis is currently a popular method for interdisciplinary knowledge content analysis [10-11]. However, different subject categories often have vastly different keyword volumes, which directly affects clustering effectiveness. Moreover, when the range of involved disciplines is large, clustering results struggle to associate knowledge content with specific subject categories. This paper first constructs a keyword-subject category co-occurrence network encompassing all involved disciplines and their keywords, then performs centrality analysis to identify major knowledge content in the subject research field. This is combined with key subject categories and domain expert opinions to identify interdisciplinary knowledge content relevant to the subject field.

2.3 Key Components

2.3.1 Discipline Influence Index Analyzing high-citation networks of inter-subject citation relationships can identify influential subject categories to some extent, but such relationships have weak relevance to specific subject research fields. Whether identifying high-influence subject categories based on Hub nodes [8], detecting disciplinary links using improved BGL algorithms [7], or calculating similarity based on inter-subject citation relationships [9], these approaches analyze high-citation networks between subject categories. While citation networks are dense, they are loose in terms of knowledge units contained in network nodes, making it difficult to reflect disciplinary influence at the thematic knowledge content level.

Keywords are important indicators representing literature research topics and can simply and directly reflect subject content. Since journals have subject category assignments, the literature they publish belongs to one or more subject categories. This paper explores inter-subject influence from the perspective of keywords belonging to specific subject categories. Generally, if two subject categories share more keywords, they exert stronger mutual influence. However, this influence is asymmetric. Influence magnitude is affected by both the number of shared keywords and the scale of the influencing party's keyword set. With a fixed number of shared keywords, influence is weaker when the influencing party has a larger keyword scale, and stronger when smaller.

Therefore, the influence of subject category i on another subject category j is proportional to their number of shared keywords and inversely proportional to the keyword scale of subject category i. Accordingly, this paper defines the Discipline Influence Index *P_{ij}* to determine the influence strength of subject category i on subject category j:

$$P_{ij} = \frac{\sum(W_i \cap W_j)}{\sum W_i - \sum(W_i \cap W_j)}$$

where *W_i* is the keyword set of subject category i, *W_j* is the keyword set of subject category j, *(W_i W_j)* is the number of shared keywords between categories i and j, and *W_i - (W_i W_j)* is the number of keywords in category i that differ from those in category j. When *W_i - (W_i W_j) = 0*, meaning all keywords in category i appear in category j, the influence index *P_{ij}* equals 1.

2.3.2 Network Centrality Analysis Centrality measures the central position of individual nodes in a network, reflecting differences in location or advantage within the social network structure [15]. Centrality analysis includes degree centrality, betweenness centrality, and closeness centrality. Degree centrality measures the number of nodes directly connected to a given node. Betweenness centrality measures a node's influence on other nodes in the network. Closeness centrality measures the degree to which a node is controlled by other nodes.

This paper prioritizes betweenness centrality in analyzing the discipline influence network, followed by closeness centrality, and finally degree centrality. Betweenness centrality measures a node's control capability—the extent to which it lies between other nodes—while closeness centrality measures a node's independence from control by other nodes [16].

2.3.3 Network Structural Holes Analysis In network influence strength research, R.S. Burt proposed structural holes theory in 1992, arguing that individuals in structural hole positions gain competitive advantages and innovation capabilities through information filtering [17]. H.J. Raider's 1998 empirical study confirmed that structural hole spanners play important roles in information control, identification, and transaction [18]. Structural holes have four measurement indicators: effective size, efficiency, constraint, and hierarchy. Effective size equals non-redundant factors in the network. Efficiency is related to effective size. Constraint refers to an actor's ability to utilize structural holes—lower constraint indicates greater influence over other subject categories. Hierarchy is related to constraint.

This paper calculates structural hole indicators for subject category nodes in the discipline influence network to observe status changes and further identify influential core subject categories. Network centrality and structural holes analyses are common methods in social network analysis. For example, Ye Chunlei et al. used social network analysis for knowledge recommendation in user retrieval [20], and Wang Xia used it to analyze characteristics of interdisciplinary knowledge exchange networks in China over 35 years [21]. Additionally, this paper employs NetDraw [19] for visualizing the subject category influence network to reveal more intuitive identification results.

---

3. Experiments and Results Analysis

3.1 Experimental Data

Urban agriculture is an agricultural phenomenon closely related to urban economy, society, culture, science, and technology—a developed modern agriculture that emerges when urban economic development reaches higher levels. It is a research field with strong interdisciplinary characteristics [20]. This study selected the Web of Science Core Collection database with the search query: TI=(urban agriculture OR urban modern agriculture OR modern urban agriculture OR metropolis modern agriculture) OR AK=(urban agriculture OR urban modern agriculture OR modern urban agriculture OR metropolis modern agriculture), limited to Article document types. The search was conducted in April 2017, covering 2007-2016, retrieving 487 documents. Table 1 shows some subject categories and their document counts in this dataset.

From 2007 to 2016, the number of subject categories involved in urban agriculture research increased annually (17 categories in 2007, 41 in 2015), indicating expanding research scope. Meanwhile, the average number of keywords per subject category also rose yearly, from 5.6 in 2008 to 21.9 in 2016 (see Table 2), suggesting increasingly rich research content.

3.2 Core Subject Category Analysis in the Research Field

While Table 2 shows macro-level trends in research scope and content, interdisciplinary development analysis requires identifying high-influence core subject categories. Using 2007 data as an example, the identification process is demonstrated below.

First, the Discipline Influence Index (Formula 1) calculates pairwise influence between subject categories to construct a discipline influence matrix. Table 3 shows partial 2007 data.

3.2.1 Centrality Analysis Centrality analysis of the 2007 urban agriculture discipline influence network yields the results shown in Table 4 . Betweenness centrality measures a node's influence on others—higher values indicate core network positions with stronger influence. Table 4 shows "Environmental Sciences" has the highest betweenness centrality, followed by "Water Resources," "Environmental Studies," "Agricultural Engineering," and "Agronomy." Closeness centrality operates inversely: higher values indicate less central positions. Table 4 data reveals relatively low closeness centrality for "Environmental Sciences," "Environmental Studies," and "Water Resources," suggesting these are likely core network nodes. "Environmental Sciences" and "Environmental Studies" also show high degree centrality, reflecting numerous direct connections to other disciplinary nodes and indicating their prominent network status as potential core subject categories.

3.2.2 Structural Holes Analysis Structural holes analysis of the 2007 urban agriculture discipline influence network (using "whole network model") produces the results in Table 5 . Lower constraint indicates greater influence over other categories, while hierarchy reflects how concentrated constraints are on a single node. Table 5 shows low constraint and relatively low hierarchy for "Planning & Development," "Microbiology," "Public, Environmental & Occupational Health," "Environmental Sciences," and "Environmental Studies." Effective size represents non-redundant factors—higher values indicate greater influence. Table 5 reveals high effective size for "Public, Environmental & Occupational Health," "Environmental Sciences," and "Environmental Studies." Combined analysis of effective size, constraint, and hierarchy suggests these three categories may be core network nodes.

3.2.3 Visualization Analysis Inputting the 2007 urban agriculture discipline influence network data into NetDraw with an edge (influence index) threshold >0.03 yields the visualization in Figure 2 [Figure 2: see original paper]. In this network, subject categories are nodes, edges represent inter-category influence, and edge weights (influence strength) are indicated by line thickness. Figure 2 shows "Environmental Sciences," "Environmental Studies," "Public, Environmental & Occupational Health," "Microbiology," and "Planning & Development" are prominent in both link quantity and strength.

Integrating centrality, structural holes, and visualization analyses identifies the 2007 urban agriculture core subject categories as: "Environmental Sciences," "Environmental Studies," and "Public, Environmental & Occupational Health."

3.3 Interdisciplinary Knowledge Content Analysis

The 2007 urban agriculture field involved 17 subject categories with 110 unique keywords. A keyword-subject category co-occurrence matrix was constructed (partial data shown in Table 6). Two-mode network centrality analysis of this matrix yields the results in Table 7 .

Table 7 shows that besides the subject field keyword "urban agriculture," keywords like "Irrigation," "Wastewater," "Helminthes," "Coliforms," and "food security" have high betweenness and degree centrality, indicating their core positions and influence on other nodes. These keywords thus reveal major 2007 urban agriculture research topics.

3.4 Urban Agriculture Interdisciplinary Trend Analysis Results

The identified core subject categories ("Environmental Sciences," "Environmental Studies," and "Public, Environmental & Occupational Health") encompass 8 unique documents and 57 unique keywords, covering all major knowledge content identified in Table 7. Combining these keywords with citation counts and domain expert opinions identified five core documents, detailed in Table 8 .

These five documents represent interdisciplinary research: four belong to "Public, Environmental & Occupational Health," two to "Environmental Sciences," and one to "Environmental Studies." Research topics include wastewater irrigation methods, microbial contamination, relationships between contamination and health, microbial contamination in farms and markets and associated consumer risk groups, and economic evaluation of peri-urban agriculture profitability and sustainability.

The study also used VOSviewer with its built-in SLM algorithm to visualize co-citation clustering relationships, forming a knowledge map of highly cited papers [23]. Applied to the 17 subject categories in the 2007 dataset, this produced the visualization in Figure 3 [Figure 3: see original paper]. VOSviewer identified four clusters, with "tropical medicine" recognized as an influential category due to strong associations with other high-impact disciplines. While "Public, Environmental & Occupational Health" did not prominently appear, replacing it with "tropical medicine" would have omitted two key documents

(Nos. 498 and 504) that domain experts confirmed as constituting 2007 research hotspots.

---

4. Conclusion and Discussion

In interdisciplinary research, scientific questions are manifested through specific subject research fields. Analyzing interdisciplinary trends requires dynamically identifying all involved subject categories, recognizing core categories, and then identifying major interdisciplinary research content to effectively guide development directions.

This paper proposes a comprehensive interdisciplinary trend analysis method that identifies core subject categories and reveals major interdisciplinary research content. Using urban agriculture as a case study, Web of Science data from 2007-2016 was retrieved and analyzed through: (1) a discipline influence network built using the Discipline Influence Index, and (2) a keyword-subject category co-occurrence network, both analyzed with integrated network methods. Alternative methods like LDA and clustering were tested but proved less effective for urban agriculture's limited disciplinary scope.

Limitations include: (1) The keyword-based Discipline Influence Index and network analysis method, while addressing existing research gaps, requires further validation of practical effectiveness and accuracy. (2) Domain expert participation is needed when applying network centrality, structural holes, and visualization to identify core disciplines. (3) Determining core documents based on core categories, keywords, citation counts, and expert opinions involves substantial manual intervention, making it cumbersome for large networks.

References

[1] Liu Zhonglin. Interdisciplinary research in the era of interdisciplinary science [J]. Science Research, 1993, 11(2): 9-13.
[2] Wagner CS, Roessner JD, Bobb K, et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature [J]. Journal of Informetrics, 2011, 5(1): 14-26.
[3] Li Jiang. The conceptual framework and measurement of "interdisciplinarity" [J]. Library and Information Knowledge, 2014(3): 87-93.
[4] Stirling A. On the economics and analysis of diversity [EB/OL]. [2017-04-30]. http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp28/sewp28.pdf.
[5] Stirling A. A general framework for analyzing diversity in science, technology and society [J]. Journal of Royal Society Interface, 2007, 4(15): 707-719.
[6] Huang Ying, Gao Tianshu, Wang Zhinan, et al. Interdisciplinary measurement research based on Web of Science classification [J]. Science Research Management, 2016(3): 124-132.
[7] Chen Shiji, Chen Chen, Han Tao, et al. Detection and analysis of interdis-

ciplinary links based on overlapping structures [J]. Library and Information Service, 2016, 60(15): 94-100.
[8] Small H. Maps of science as interdisciplinary discourse: co-citation contexts and the role of analogy [J]. Scientometrics, 2009, 83(3): 835-849.
[9] Rafols I, Porter AL, Leydesdorff L. Science overlay maps: a new tool for research policy and library management [J]. Journal of the American Society for Information Science and Technology, 2010, 61(9): 1871-1887.
[10] Wei Jianxiang, Sun Yuehong, Su Xinning. Research on interdisciplinary knowledge mining model [J]. Information Theory and Practice, 2012(4): 76-80.
[11] Yue Zenghui, Xu Haiyun, Guo Ting, et al. Comparative study on interdisciplinary applications between "information science" and "computer science" [J]. Information and Documentation Services, 2016(2): 16-22.
[12] Buter RK, Noyons ECM, Van Raan AFJ. Identification of converging research areas using publication and citation data [J]. Research Evaluation, 2010, 19(1): 19-27.
[13] Buter RK, Noyons EC, Van Raan AF. Searching for converging research using field-to-field citation [J]. Scientometrics, 2011, 86(2): 325-338.
[14] Pudovkina AI, Garfield E. Algorithmic procedure for finding semantically related journals [J]. Journal of the American Society for Information Science and Technology, 2002, 53(11): 1113-1119.
[15] Lin Juren. Social Network Analysis: Theory, Method and Application [M]. Beijing: Beijing Normal University Press, 2009.
[16] Liu Jun. Introduction to Social Network Analysis [M]. Beijing: Social Sciences Academic Press, 2004.
[17] Burt RS. Structural Holes: The Social Structure of Competition [M]. Cambridge, MA: Harvard University Press, 1992.
[18] Raider HJ. Market structure and innovation [J]. Social science research, 1998, 27(1): 1-21.
[19] NetDraw [EB/OL]. [2017-04-15]. http://www.analytictech.com/downloadnd.htm.
[20] Ye Chunlei, Xing Yanli. Research on personalized retrieval recommendation model for graduate students based on LDA and social network centrality [J]. Library and Information Service, 2015, 59(13): 104-110.
[21] Wang Xia, Zhao Bingjun. Research on structural characteristics of interdisciplinary knowledge exchange networks [J]. Information Science, 2016, 34(5): 46-50.
[22] Miao Runlian. Development and Practice of Urban Agriculture Abroad [M]. Beijing: China Agriculture Press, 2014.
[23] Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping [J]. Scientometrics, 2010, 84(2): 523-538.

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