Research on Data Model Theory and Knowledge Discovery Application Examples Based on Multiple Co-occurrence in Scientific Literature (Postprint)

Authors: Pang Hongshen

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] In scientific literature, various feature items and their interrelationships constitute the basic units of diverse co-occurrence phenomena. By mining the associations among co-occurring feature items, co-occurrence analysis can detect various aspects of the patterns in scientific and technological activities from different perspectives, providing research managers and researchers with a new viewpoint for observing scientific development in a comprehensive and multi-angle manner.

[Method/Process] Through fundamental theoretical research on multiple co-occurrence, we construct a unique foundational theoretical system for multiple co-occurrence data models. This theoretical system includes: the definition of multiple co-occurrence, the research scope of multiple co-occurrence, variable symbols for multiple co-occurrence, matrix definitions for multiple co-occurrence, data organization forms for multiple co-occurrence, as well as calculation formulas for multiple co-occurrence extension coefficients and their application scope. Furthermore, based on the cross-graph visualization method for multiple co-occurrence, we construct knowledge discovery methods that can be used to analyze co-occurrence relationships among three or more feature items, including analytical methods for co-occurrence association strength, cited association strength, and co-occurrence burst strength.

[Results/Conclusion] Through the construction of this foundational theoretical system, we expand the research scope of co-occurrence phenomena and provide theoretical support for co-occurrence analysis to evolve toward multi-angle and multi-dimensional multiple co-occurrence analysis. Through empirical research and by selecting different application cases of multiple co-occurrence, we demonstrate that this method can be applied to analyses of research fields, research

institutions, inter-institutional comparisons, research scholars, etc., while achieving favorable analytical results. Due to the characteristics of multi-dimensional analytical perspectives and diversified analytical methods in this methodological system, analyses using this method can not only achieve the analytical effects of single and double co-occurrence but also reveal more extensive and in-depth knowledge content than general co-occurrence analysis.

Full Text

Preamble

A Study on Data Model Theory and Knowledge Discovery Application Paradigms Based on Multiple Co-occurrence in Scientific Literature

Pang Hongqian
Library of Shenzhen University, Shenzhen 518060

Abstract

[**Purpose/Significance**] In scientific literature, various entities and their associations constitute the basic units that form diverse co-occurrence phenomena. By mining the associations among co-occurring entities, co-occurrence analysis can detect all aspects of the laws governing scientific and technological activities from different perspectives, providing researchers and research managers with a new, comprehensive, and multi-angle view of scientific development. [**Method/Process**] Through foundational theoretical research on multiple co-occurrence, this paper constructs a unique theoretical system for multiple co-occurrence data models, including: definitions of multiple co-occurrence, research scope, variable symbols, matrix definitions, data organization forms, and calculation formulas for extension coefficients along with their application domains. Additionally, based on cross-graph visualization for multiple co-occurrence, the paper builds knowledge discovery methods for analyzing relationships among three or more entities, encompassing methods for analyzing co-occurrence relevance strength, citation relevance strength, and co-occurrence burst strength. [**Result/Conclusion**] This theoretical framework expands the research scope of co-occurrence phenomena and provides foundational theoretical support for advancing co-occurrence analysis toward multi-angle, multi-dimensional multiple co-occurrence analysis. Through empirical studies using different application cases, the method demonstrates applicability in analyzing research fields, institutions, inter-institutional comparisons, and scholars, yielding effective analytical results. Due to its multi-dimensional analysis angles and diverse methods, this approach not only achieves the effects of single and pairwise co-occurrence analysis but also reveals more extensive and in-depth knowledge than conventional co-occurrence analysis.

Keywords: multiple co-occurrence; multi-feature co-occurrence; multi-source data; data model; knowledge discovery

1. Scope of Co-occurrence Analysis

Co-occurrence in scientific literature refers to the phenomenon where identical or different types of features appear together in papers, patents, and other documents. Examples include: keywords that appear together across multiple journal articles; co-authors and collaborating institutions; simultaneous appearance of papers with keywords or institutions with authors; and in patent literature, co-occurring inventors or inventor-IPC classification co-occurrences. All these fall within the scope of co-occurrence research [1].

Co-occurrence analysis is a quantitative method for analyzing co-occurring information in various information carriers, grounded in the psychological principle of contiguity association [2] and the mapping principle of knowledge structures. Through co-occurrence analysis, researchers can reveal patterns of scientific activity from multiple perspectives. For instance, keyword co-occurrence in journal papers directly reflects research themes, details, methods, and technologies, enabling examination of scientific structure in terms of knowledge, methodology, and dimensions. Patent inventor co-occurrence directly represents technological collaboration at the individual level. Citations among papers and patents serve as tangible markers of invisible colleges, and analyzing these citation phenomena reveals patterns, laws, and characteristics of scientific communication.

In bibliometric data, co-occurrence is not an isolated case but a widespread phenomenon. Various types of feature co-occurrence connect discrete paper data into an organic whole, revealing relationships among research objects, mining implicit or potential useful knowledge, and disclosing the structure and changes of disciplines or research entities. With computer technology assistance, co-occurrence analysis demonstrates unique capabilities in building concept spaces and ontologies for semantic retrieval, improving text classification in knowledge organization, analyzing knowledge content associations in literature, and mining knowledge value, becoming an important tool for knowledge mining and knowledge services. In knowledge representation, both content features and external features that reflect information have semantic connotations and interconnections, forming the foundation for text knowledge association revelation and knowledge mining [3].

Many scholars have studied co-occurrence analysis methods and tools for features in academic papers. R. Fano [3] first proposed the concept of bibliographic coupling in 1956; H. Small [4] introduced co-citation analysis; H. White and B. Griffith [5] proposed author co-citation analysis; M. Callon et al. [6] first introduced co-word analysis; Chinese scholar Zheng Huachuan et al. [7] proposed co-paper analysis; D. Zhao et al. [8] introduced author-bibliographic coupling analysis; Liu Zhihui et al. [9] proposed author-keyword coupling analysis; and L. Yang et al. [10] applied institution-keyword co-occurrence analysis. In terms of visualization tools for co-occurrence analysis, over ten software tools have been involved, including bibliometric research software BibExcel, statistical software SPSS, citation network visualization software CiteSpace, social network analysis software Ucinet and Pajek, and other co-occurrence analysis tools such as SCI-map (citation network browsing), HistCite (citation analysis), and DIVA (bibliographic coupling, co-authorship analysis). These software tools can visualize different types of co-occurrence analysis.

However, current domestic and international research methods and tools for feature co-occurrence primarily focus on pairwise co-occurrence between two features [16-20], mostly achieved by fusing multiple pairwise co-occurrence methods to reveal multi-feature relationships [21]. Direct research on analysis methods and visualization for co-occurrence among three or more features is rare. Pang Hongqian et al. [22-24] used multiple (multi-feature) co-occurrence analysis methods and developed corresponding visualization mapping tools to analyze research institutions and fields, significantly expanding analytical vision and scope to discover more extensive and in-depth knowledge than general co-occurrence analysis. Therefore, systematic knowledge discovery research on co-occurrence relationships among three or more features can add multiple analytical angles and information sources for reflecting scientific and technological activity patterns and knowledge domains, offering substantial value for knowledge mining and exploration. Furthermore, integrating multi-source data from scientific literature databases to mine associations among co-occurring features will facilitate the development of personalized value discovery methods that fuse multi-source big data, and by studying data fusion and associations across multi-disciplinary scientific literature, can reveal value paradigms for general and interdisciplinary development.

2. Definition and Research Scope of Multiple Co-occurrence

This paper defines the repeated appearance of a single feature across multiple documents as single co-occurrence, co-occurrence of two features as double co-occurrence, and so on. Co-occurrence of three or more features is termed multiple co-occurrence. Therefore, multiple co-occurrence is defined as the phenomenon where three or more identical or different types of features appear together, such as author-keyword-journal triples appearing simultaneously in multiple papers, inventor-IPC classification-keyword combinations, or author-cited author-keyword-cited keyword combinations.

Compared with double co-occurrence, multiple co-occurrence like author-keyword-journal can reveal deeper knowledge than pairwise combinations like author-keyword or author-journal. Analyzing author-keyword-journal multiple co-occurrence is equivalent to simultaneously analyzing three double co-occurrence phenomena (author-keyword, author-journal, keyword-journal) and their relationships. As shown in Table 1 , multiple co-occurrence has unique advantages in revealing in-depth knowledge, particularly when analyzing cross co-occurrence of multiple features between papers and patents, which can further reflect evolution paths in industry-academia-research relationships. Figure 1 [Figure 1: see original paper] visually demonstrates the difference between multiple co-occurrence and conventional feature co-occurrence analysis objects, while Figure 2 [Figure 2: see original paper] shows examples of potential relationships in multiple co-occurrence across multi-source scientific literature (papers, patents, monographs, etc.).

3. Variable Symbols for Multiple Co-occurrence Features

In S. Morris's doctoral dissertation, Figure 3 [Figure 3: see original paper] (compiled in this paper) vividly depicts relationships among features in journal articles. Arrows and accompanying text represent interactions between two features; for example, paper → keyword indicates different keywords can appear multiple times across papers, while keyword → paper indicates each paper can contain multiple keywords. Morris also used feature name abbreviations as variable names, a convention this paper adopts and extends for application to scientific literature (including papers and patents) as shown in Table 2 .

4. Matrix Definition and Data Organization for Multiple Co-occurrence

In bibliometric research, quantitative analysis of associations among co-occurring features requires mathematical processing to convert data into various co-occurrence matrices, upon which data mining and visualization methods can identify implicit relationships. Although different co-occurrence applications reveal different scientific phenomena, matrix analysis techniques are largely similar [1].

Co-citation, co-word, and other same-feature co-occurrence matrices are widely used in informetrics. In early co-occurrence research, limitations in computer storage, processing speed, and matrix-based data mining techniques meant much analysis relied on same-feature co-occurrence matrices. With advancing computer technology and increasing demand for multi-matrix transformation in visualization, researchers gradually recognized the importance of matrix transformation research. Dutch scholars E. Engelsman and A. van Raan discovered that original binary co-occurrence matrices could be converted into corresponding symmetric co-occurrence matrices through matrix multiplication [25]. American informetrics expert S. Morris systematically and comprehensively studied mathematical transformation relationships among various co-occurrence matrices in his doctoral dissertation [11].

The general matrix definition is: an m×n table of numbers a_{ij} (i = 1, 2, …, m; j = 1, 2, …, n) arranged in m rows and n columns:

```
a_{11} a_{12} ... a_{1n}
a_{21} a_{22} ... a_{2n}
...
a_{m1} a_{m2} ... a_{mn}
```

This is called an m×n matrix, denoted as A = (a_{ij})*m×n. The element a*{ij} is called the element in row i, column j of matrix A. Specifically, an n×n matrix is also called an n-order square matrix.

In social network analysis, asymmetric matrix rows and columns represent actors and indicators respectively; for symmetric square matrices, rows and columns represent identical actors. In bibliometric research, matrix rows and columns represent co-occurring features, with elements indicating whether corresponding features are related or the strength of their relationship.

In S. Morris's dissertation [11], the matrix definition for two-feature co-occurrence is:

O_{ij}[x1; x2] = n (features i and j co-occur n times)
O_{ij}[x1; x2] = 0 (features i and j do not co-occur)

where x1, x2 represent two different feature types (e.g., keywords, authors, journals), and i, j are specific objects of x1, x2 respectively. The corresponding graph structure is shown in Figure 5 [Figure 5: see original paper].

Building on Morris's work, this paper extends his theoretical framework from double to multiple co-occurrence, including matrix definitions and data organization forms from two-dimensional matrices to multi-tuples for multiple co-occurrence analysis.

The extended matrix definition for multiple co-occurrence features is:

O_{ijk}…[x1; x2; x3…] = n (features i, j, k, etc. co-occur n times)
O_{ijk}…[x1; x2; x3…] = 0 (features i, j, k, etc. do not co-occur)

The graph structure for this multi-dimensional matrix definition is shown in Figure 6 [Figure 6: see original paper], where lines of the same type represent co-occurrence frequencies among several features. For example, o_{111}[x1; x2; x3] represents the frequency of co-occurrence of feature 1 from set x1 with feature 1 from sets x2 and x3.

For data organization, Morris used traditional two-dimensional matrices to represent relationships between two features [1]:

O[x1; x2] = [O_{ij}]

Since multiple co-occurrence involves relationships among three or more features, traditional two-dimensional matrices are inadequate. This paper uses multi-tuples R(x1, x2, x3…, value) to represent multi-dimensional data, where value_{ijk}… represents the co-occurrence frequency of feature i from x1, feature j from x2, feature k from x3, etc., i.e., value_{ijk}… = O_{ijk}…[x1; x2; x3…]. This extends from two-dimensional matrices to multi-tuples (Figure 7 [Figure 7: see original paper]) to accommodate multiple co-occurrence data organization and analysis.

The matrix definitions above use co-occurrence frequency as element values. This paper also defines binary versions where all elements are 0 or 1:

O'*{ij}[x1; x2] = 1 (features i and j co-occur one or more times)*
O'{ij}[x1; x2] = 0 (features i and j do not co-occur)

O'*{ijk}…[x1; x2; x3…] = 1 (features i, j, k, etc. co-occur one or more times)*
O'{ijk}…[x1; x2; x3…] = 0 (features i, j, k, etc. do not co-occur)

Table 3 provides an example dataset D1 to illustrate these definitions:

Table 3. Example Dataset for Multiple Co-occurrence Matrix Definition

Dataset D1	Inventor (pi)	Application Country (pc)	Application Year (py)	Keywords (pkw)
Patent 1	Inventor 1, Inventor 2	Country 1	Year 1	Keyword 1, Keyword 2, Keyword 3
Patent 2	Inventor 2, Inventor 3	Country 2	Year 2	Keyword 4, Keyword 5
Patent 3	Inventor 1	Country 2	Year 3	Keyword 3, Keyword 4

For dataset D1:
- O[inventor 1; keyword 3] = 2 (co-occurs twice)
- O'[inventor 1; keyword 3] = 1 (co-occurs one or more times)
- O[inventor 1; country 2; keyword 3] = 1 (triple co-occurs once)
- O'[inventor 1; country 2; keyword 5] = 0 (no triple co-occurrence)

5. Extension Coefficients for Multiple Co-occurrence

Based on the above matrix definitions, extension coefficients EX_n and E'X_n for multiple co-occurrence can be calculated.

Let m_i, …, m_j, m_k represent the numbers of distinct objects in features x1, …, x_{n-1}, x_n respectively. The formulas are:

Formula (1):
EX_n(i, …, j) = (_{k=1}^{m_k} O_i,…,jk[x1; …; x_n]) / O_i…j[x1; …; x_{n-1}]

Formula (2):
EX_n(x1; …; x_{n-1}) = (*{i=1}^{m_i}* … {j=1}^{m_j} *{k=1}^{m_k} O_i…jk[x1; …; x_n]) / (* {i=1}^{m_i} … _{j=1}^{m_j} O_i…j[x1; …; x_{n-1}])

Formula (3):
E'X_n(i, …, j) = (_{k=1}^{m_k} O'_i…jk[x1; …; x_n]) / O'_i…j[x1; …; x_{n-1}]

Formula (4):
E'X_n(x1; …; x_{n-1}) = ({i=1}^{m_i} … {j=1}^{m_j} {k=1}^{m_k} O'i…jk[x1; …; x_n]) / ({i=1}^{m_i} … {j=1}^{m_j} O'_i…j[x1; …; x_{n-1}])

Application domains:
- **EX_n**: Analyzes average quantity distribution of features per document, such as average keywords per paper, average inventors per patent in a given year, or average authors and keywords per paper in a specific journal.
- **E'X_n**: Analyzes distribution of feature types across the entire dataset, such as the number of different journals in which an author has published, types of patents an inventor applied for in a year, or years in which a journal published papers by a specific author.

Table 4 provides example dataset D2 to illustrate these coefficients:

Table 4. Example Dataset for Multiple Co-occurrence Extension Coefficients

Dataset D2	Author (ap)	Journal (jp)	Year (yp)	Keywords (kwp)
Paper 1	Author 1, Author 2	Journal 1	Year 1	Keyword 1, Keyword 2, Keyword 3
Paper 2	Author 2, Author 3	Journal 2	Year 2	Keyword 4, Keyword 5
Paper 3	Author 1	Journal 1	Year 1	Keyword 1, Keyword 2

For Paper 1:
- E_{kwp}(ap) = 3 (each author used 3 keywords)
- E_{kwp}(ap; jp) = 3 (each author used 3 keywords in each journal)

For Author 1 in D2:
- E_{kwp}(author 1) = 2.5 (average 2.5 keywords per paper)
- E_{kwp}(author 1, journal 1) = 3 (average 3 keywords per paper in journal 1)

For entire dataset D2:
- E'*{jp}(author 1) = 2 (published in 2 different journals)*
- *E'*{yp}(author 2) = 1 (published in only 1 year)
- E_{kwp}(ap) = 2.4 (average 2.4 keywords per author per paper)
- E_{kwp}(ap, jp) = 2.4 (average 2.4 keywords per author per journal per paper)
- E'*{jp}(ap) = 1.67 (average 1.67 journals per author)*
- *E'*{yp}(ap) = 1.33 (average 1.33 publication years per author)
- E'_{jp}(ap, yp) = 1.25 (average 1.25 journals per author-year combination)

6. Design of Knowledge Discovery Methodology for Multiple Co-occurrence

This paper integrates knowledge discovery concepts, patterns, and general processes with multiple co-occurrence analysis, following these steps: multi-source scientific literature collection and cleaning → data processing (using matrix transformation, dimensionality reduction, clustering) → generating multiple co-occurrence cross-graphs → analyzing cross-graph characteristics → summarizing knowledge discovery conclusions. The methodology framework is shown in Figure 8 [Figure 8: see original paper], comprising three aspects: co-occurrence relevance strength analysis, citation relevance strength analysis, and co-occurrence burst strength analysis.

- **Co-occurrence relevance strength analysis** reveals potential co-occurrence relationships through frequency analysis among multiple features.

- **Citation relevance strength analysis** reveals attention patterns through analysis of joint citation frequencies.

- **Co-occurrence burst strength analysis** reveals changes and emerging hotspots through analysis of burst weights.

The methodology can analyze relationships among three or more paper features, though co-occurrence frequencies for more than three features tend to be low with high data dispersion, making triple co-occurrence (three features) the primary research sample. Methods for more than three features can be extrapolated accordingly. Data sources may include papers, patents, monographs, and other scientific literature, with different combinations revealing different knowledge content depending on research purposes.

The visualization approach uses multiple co-occurrence cross-graphs (improved from Morris's cross-graphs), which can simultaneously display four types of co-occurrence relationships in one triple co-occurrence cross-graph: three pairwise relationships (2-mode networks) and one triple relationship (3-mode network). This is more intuitive and convenient than multi-modal network graphs and offers superior display effects and data representation.

7. Case Studies of Triple Co-occurrence Applications

This methodology has broad application scope, enabling analysis of research fields, institutions, inter-institutional comparisons, and scholars. Figure 9 [Figure 9: see original paper] shows an institution-journal-keyword triple co-occurrence cross-graph for the embryonic stem cell research field, revealing major research institutions and their thematic distributions across mainstream journals. The top and bottom regions show leading journals by publication volume, with distinct thematic focuses. The central region shows how major institutions distribute their research themes across these journals. For example, Sun Yat-sen University publishes papers on "cell differentiation" and "hematopoietic stem cells" primarily in *Chinese Journal of Pathophysiology*, while focusing on "epidermal stem cells" in *Journal of Sun Yat-sen University (Medical Sciences)*.

Figure 10 [Figure 10: see original paper] presents a year-journal-author triple co-occurrence citation strength cross-graph for the Institute of Scientific and Technical Information of China (ISTIC), showing highly cited authors and citation distribution trends across journals and years. Figure 11 [Figure 11: see original paper] shows a burst strength cross-graph for ISTIC, revealing emerging features and hotspot content. Early in 2001-2010, researchers like Wen Rongsheng and Bai Guoying showed high publication growth; later in this period, ISTIC graduate students appeared frequently, indicating emerging research strength. In author-keyword combinations, early hotspots focused on classification studies by Wen Rongsheng and Bai Guoying, later diversifying into multiple research themes with rapid growth in different author-topic combinations.

Conclusion

This paper outlines concepts of multiple co-occurrence, defines its scope, clarifies variable symbols, and extends Morris's foundational research to multiple co-occurrence domains, including matrix definitions, data organization forms, and extension coefficient calculations. The constructed theoretical system expands co-occurrence research scope and provides theoretical support for multi-angle, multi-dimensional analysis. The knowledge discovery methodology based on cross-graph visualization can analyze relationships among three or more features through co-occurrence relevance, citation relevance, and burst strength analyses. Empirical studies demonstrate applicability across research fields, institutions, and scholars, revealing more extensive and in-depth knowledge than conventional methods.

Future research will incorporate additional knowledge discovery theories such as data mining, decision trees, association rules, and neural networks to explore general and special patterns among multi-features and multi-source literature. Further empirical studies across different scientific domains are needed to validate feasibility and applicability. While cross-graph visualization is adopted here, other methods like multi-modal network graphs remain viable alternatives. The current analysis of various feature combinations (institution-journal-keyword, year-keyword-journal, author-journal-keyword, etc.) shows promise, though knowledge discovery effects for other combinations (author-year-reference, author-citing author-citing year) and multi-source literature (papers, patents, monographs) require further investigation.

References

[1] Yang Liying. Theory and Application of Co-occurrence in Scientific Papers

[D]. Beijing: Institute of Scientific and Technical Information of China, 2007.
[2] Wang Yuefen, Song Shuang, Miao Lu. Research on Co-occurrence Analysis Application in Knowledge Services [J]. New Technology of Library and Information Service, 2006(4): 29-34.
[3] Fano R. Information theory and the retrieval of recorded information [M]//Documentation in Action. New York: Reinhold Pub. Co., 1956: 238-244.
[4] Small H. Macro-level changes in the structure of co-citation clusters: 1983-1989 [J]. Scientometrics, 1993, 26(1): 5-20.
[5] White H, Griffith B. Author co-citation: A literature measure of intellectual structure [J]. Journal of the American Society for Information Science, 1981, 32(3): 163-169.
[6] Callon M, Law J, Rip A. Mapping the dynamics of science and technology: Sociology of science in the real world [M]. New York: Sheridan House, 1986.
[7] Zheng Huachuan, Cui Lei. Co-word and co-paper cluster analysis of low-frequency cited papers on gastric precancerous lesions [J]. Chinese Journal of Medical Library and Information Science, 2002, 11(3): 1-3.
[8] Zhao D, Andreas S. Evolution of research activities and intellectual influences in information science 1996-2005: Introducing author bibliographic-coupling analysis [J]. Journal of the American Society for Information Science and Technology, 2008, 59(13): 2070-2086.
[9] Liu Zhihui, Zhang Zhiqiang. Author-keyword coupling analysis method and empirical research [J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(2): 268-275.
[10] Yang L, Morris S, Barden E. Mapping institutions and their weak ties in a specialty: A case study of cystic fibrosis body composition research [J]. Scientometrics, 2009(2): 421-434.
[11] Morris S. Unified mathematical treatment of complex cascaded bipartite networks: The case of collections of journal papers [D]. Oklahoma: Oklahoma State University, 2005.
[12] Morris S, Deyong C, Wu Z, et al. DIVA: A visualization system for exploring document databases for technology forecasting [J]. Computers & Industrial Engineering, 2002, 43(4): 841-862.
[13] Leng Fuhai, Wang Lin, Li Yong. A triple co-word analysis method based on literature keywords: A case study of knowledge discovery [J]. Journal of the China Society for Scientific and Technical Information, 2011(10): 1072-1077.
[14] Zhang Zili, Zhang Ziqiong, Li Xiangyang. Research institution and keyword co-occurrence analysis method based on 2-mode networks [J]. Journal of the China Society for Scientific and Technical Information, 2011(12): 1249-1260.
[15] Leydesdorff L. What can heterogeneity add to the scientometric map? Steps towards algorithmic historiography [EB/OL]. [2018-01-30]. http://arxiv.org/abs/1002.0532.
[16] Leydesdorff L, Vaughan L. Co-occurrence matrices and their applications in information science: Extending ACA to the web environment [J]. Journal of the American Society for Information Science and Technology, 2006, 56(12): 1616-1628.
[17] Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of

co-citation clusters: A multiple-perspective co-citation analysis [J]. Journal of the American Society for Information Science and Technology, 2010, 61(7): 1386-1409.
[18] Zhang Ting. Visual analysis of science communication research [D]. Dalian: Dalian University of Technology, 2009.
[19] Liu Zeyuan, Chen Yue, Hou Haiyan, et al. Methods and applications of scientific knowledge mapping [M]. Beijing: People's Publishing House, 2008.
[20] Feng Lu, Leng Fuhai. Theoretical progress in co-word analysis methods [J]. Journal of Library Science in China, 2006(2): 88-92.
[21] Hu Qiongfang, Zeng Jianxun. Research on literature relevance determination based on multi-co-occurrence [J]. Information Studies: Theory & Application, 2010, 33(8): 77-80.
[22] Pang H. A knowledge discovery method based on analysis of multiple co-occurrence relationships in collections of journal papers [J]. Chinese journal of library and information science, 2012, 5(4): 9-20.
[23] Pang Hongqian, Fang Shu, Fan Wei, et al. Research on institutional research status analysis method based on multiple co-occurrence: A case study of National Science Library, Chinese Academy of Sciences [J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(11): 1140-1152.
[24] Pang Hongqian. Algorithm implementation and visualization mapping research on burst strength analysis method based on multi-feature co-occurrence in scientific papers [J]. Library and Information Service, 2015, 59(24): 115-122.
[25] Engelsman E, van Raan A. Mapping of technology: A first exploration of knowledge diffusion amongst fields of technology [R]. Bangalore: CWTS report, 1991.

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