ERGM-Based Empirical Study on Knowledge Connection Mechanisms in Interdisciplinary Fields: Postprint

Authors: Cao Yujie, Li Gang, Mao Jin, Yang Guancan

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

Abstract

[Purpose/Significance] This study aims to reveal the micro-level knowledge connection mechanisms in interdisciplinary fields by exploring the influencing factors and their operational mechanisms underlying the formation of co-word networks in interdisciplinary domains. [Method/Process] Grounded in network embeddedness theory, this study categorizes the influencing factors for establishing keyword co-occurrence relationships in interdisciplinary fields into network structural factors (endogenous variables) and keyword attribute factors (exogenous variables), and employs exponential random graph models to conduct an empirical study on the interdisciplinary field of "medical informatics". [Results/Conclusions] The findings indicate that network structure exerts a greater influence on co-occurrence relationship formation than keyword attributes; preferential attachment and transitivity mechanisms demonstrate significant positive effects; keyword nodes tend to connect with newer nodes; keywords from medical informatics tend to establish co-occurrence relationships with keywords from basic disciplines, whereas keywords from basic disciplines tend to connect with keywords within their own disciplines.

Full Text

Preamble

Volume 63, Issue 19, October 2019

An Empirical Study on Knowledge Connection Mechanisms in Interdisciplinary Fields Based on ERGM

Cao Yujie[1], Li Gang[1], Mao Jin[1], Yang Guancan[2]
[1]Center for Studies of Information Resources, Wuhan University, Wuhan 430072

[2]School of Information Resource Management, Renmin University of China, Beijing 100872

Abstract

[**Purpose/Significance**] This study aims to reveal the micro-level knowledge connection mechanisms in interdisciplinary fields by examining the influencing factors and their underlying mechanisms in the formation of co-word networks within such domains. [**Method/Process**] Grounded in network embedding theory, the factors influencing keyword co-occurrence relationships in interdisciplinary fields are categorized into network structural factors (endogenous variables) and keyword attribute factors (exogenous variables). Using the Exponential Random Graph Model (ERGM), an empirical investigation is conducted on the interdisciplinary field of Medical Informatics. [**Results/Conclusions**] The findings demonstrate that network structure exerts a greater influence on co-occurrence relationship formation than keyword attributes alone. Preferential attachment and transitivity mechanisms show significant positive effects. Keyword nodes tend to connect with newer nodes. Keywords from Medical Informatics show a propensity to establish co-occurrence relationships with those from foundational disciplines, while keywords from these foundational disciplines tend to connect with keywords within their own disciplines.

Keywords: Exponential Random Graph Model; Interdisciplinary; Homophily; Preferential Attachment; Time Effect
Classification Number: G250.2
DOI: 10.13266/j.issn.0252-3116.2019.19.013

Interdisciplinary research has become a crucial form of scientific activity, with interdisciplinary fields occupying a significant and important position in the scientific landscape [1]. Interdisciplinary research occurs not only within natural sciences but also across disciplinary boundaries, as seen in emerging fields such as social computing, digital humanities, and computational communication studies that have garnered increasing attention from researchers. The phenomenon of interdisciplinarity has attracted interest from fields like science of science and philosophy of science, while also generating relevant research in information science that employs bibliometric analysis to examine interdisciplinary domains [2-3]. The mainstream research direction involves using bibliometrics to reveal interdisciplinary characteristics of journals [4], scholars [5], and research fields [6].

From a temporal perspective, keywords in a field emerge sequentially, and the establishment of co-occurrence relationships between keywords represents a knowledge connection phenomenon driven by temporal ordering. Based on this observation, we employ co-word networks to investigate knowledge connections within disciplinary fields. Co-word networks, with keywords representing thematic and semantic content as nodes [9], reflect the knowledge structure of a field and reveal its internal knowledge relationships [10]. Current research

frequently utilizes the semantic information in co-word networks to uncover thematic structures and their evolution [11-13]. Interdisciplinary fields develop by crossing boundaries between multiple traditional disciplines [2]. Some studies have recognized this special characteristic, considering influences from different disciplines during thematic analysis to understand how interdisciplinary knowledge functions within these fields [14-15]. However, existing thematic analyses tend to focus on macro-level trends rather than micro-level processes. In fact, co-word networks inherently carry micro-level knowledge interrelationships, and studying the structural determinants of these networks provides an effective means to dissect the micro-level mechanisms operating within a field' s knowledge system.

Recent studies have employed scientific knowledge networks to describe scientific knowledge systems, involving specific network models such as citation networks, collaboration networks, and co-word networks, depending on the types of knowledge nodes and relationships [7-8]. Knowledge connections in scientific knowledge networks link different knowledge elements and constitute the micro-structural foundation of these networks. Investigating the formation processes and dynamic mechanisms of knowledge connections in scientific knowledge networks facilitates a deeper micro-level understanding of the generation, innovation, and evolution of scientific knowledge.

Several studies have utilized complex network and social network analysis methods to reveal structural characteristics of co-word networks, such as small-world phenomena and power-law degree distributions [16]. However, two primary limitations exist in current research: First, these analyses focus predominantly on network structural features while neglecting the influence of keyword attributes themselves. Second, few studies specifically examine the unique characteristics of interdisciplinary fields, with limited consideration of how multi-disciplinary features affect the micro-structure of co-word networks. Building on this analysis, the structural features of co-word networks in interdisciplinary fields can reflect interactions among multi-disciplinary knowledge, and studying the formation process of knowledge connections in these networks can reveal the micro-level mechanisms of interdisciplinary knowledge interaction.

The Exponential Random Graph Model (ERGM) is a statistical method developed from social network analysis models that quantitatively analyzes factors influencing relationship formation [17]. This model considers both endogenous network structures and node attributes, providing a comprehensive approach to revealing influencing factors and their mechanisms in network formation. ERGM is commonly used to explain the formation mechanisms of social networks. For instance, it has been applied to study reciprocity effects, transitivity mechanisms, and structural expansion mechanisms in adolescent peer networks [18]. Scholars have gradually applied ERGM to knowledge network research, exploring formation mechanisms of collaboration networks and citation networks from perspectives including network structural characteristics, social factors, and semantic factors. C. Zhang et al. [19] used this model to examine how two

network mechanisms—transitivity and preferential attachment—and homophily in author attributes such as productivity, influence, research topics, and gender affect the formation of author collaboration relationships. Yang Guancan et al. [20] employed ERGM to empirically explain patent citation relationship formation from perspectives of citation network edges, degree distribution, transitive closure, and attributes including geographic location, field, discipline, ownership mechanism, and examiner type. However, no studies have yet used this model to investigate co-word network formation processes. Given ERGM' s advantages and current research gaps, we utilize the Exponential Random Graph Model to reveal the formation mechanisms of co-word networks in interdisciplinary fields, comprehensively analyzing the influences of network structure, disciplinary attributes, time, and other factors on the formation of interdisciplinary knowledge systems.

2 Research Methods

2.1 Analytical Framework and Research Hypotheses

Early research on social network formation mechanisms often employed linear regression approaches, focusing on how dyadic attributes affect relationship formation without considering other network structures [21]. Network embedding theory posits that actors' behaviors and influences are embedded within network environments, and network structures can be used to understand actors' behaviors [22]. Based on this perspective, factors influencing relationship formation in interdisciplinary co-word networks include not only dyadic factors but also structural information embedded within the overall network. From a network systems perspective, factors influencing keyword edge formation in co-word networks primarily derive from two sources: (1) endogenous variables, or structural embedding, referring to structural properties nodes possess by virtue of their position in the network; and (2) exogenous variables, or node attributes, referring to properties of nodes or edges themselves that influence network edge formation. Figure 1 [Figure 1: see original paper] presents the analytical framework for examining the formation mechanisms of interdisciplinary co-word networks.

From an endogenous perspective, network edge formation is influenced by preferential attachment and transitivity mechanisms. Preferential attachment is a common dynamic mechanism in complex networks where nodes with more connections are more likely to acquire new connections [23]. In co-word networks, keyword node degree centrality measures co-occurrence frequency with other keywords. Keywords with higher degree centrality may represent research hotspots and are more likely to generate new knowledge; consequently, new research topics (which may contain multiple newly emerging keywords) tend to connect with these keywords, manifesting preferential attachment [16, 24]. Keyword betweenness centrality measures whether a node lies on the shortest paths between other nodes. Nodes in intermediary positions serve as knowledge brokers, connecting different knowledge clusters. During field development, new topics may connect with intermediary nodes, thereby linking different knowl-

edge clusters more tightly and strengthening the field' s knowledge structure. Based on this analysis, we propose the following hypotheses:

H1: The greater a keyword node' s degree centrality, the higher the probability that other nodes will form co-occurrence relationships with it.

H2: The greater a keyword node' s betweenness centrality, the higher the probability that other nodes will form co-occurrence relationships with it.

Network clustering coefficient measures the degree to which nodes cluster together through the number of triads in the network. The structure where three nodes establish relationships with each other is called "triadic closure," a common dynamic in network evolution. The popular explanation is that "a friend of my friend tends to become my friend," a mechanism also known as transitivity. In co-word networks, triadic closure structures facilitate knowledge clustering effects [25], and keywords with potential for triadic closure formation possess intrinsic knowledge relatedness, making them more likely to establish knowledge connections. Therefore, we propose:

H3: The greater a keyword node' s clustering coefficient, the higher the probability that other nodes will form co-occurrence relationships with it.

From an exogenous perspective, keyword attributes may also influence co-occurrence relationship formation. Here, we primarily consider time effects and disciplinary heterogeneity. In knowledge growth processes, early-emerging keywords are more likely to represent a field' s knowledge foundation, while emerging knowledge nodes develop associations based on existing foundations. Influenced by the emergence of research hotspots, knowledge nodes tend to connect more with nodes appearing in similar time periods [26]. In interdisciplinary fields, multi-disciplinary knowledge integrates and interacts, and knowledge from different disciplines may tend to connect with each other. To verify this analysis, we propose:

H4: The smaller the difference in keywords' earliest appearance times, the more conducive to co-occurrence relationship formation.

H5: Heterogeneity in keywords' disciplinary attributes facilitates co-occurrence relationship formation.

2.2 Factors Influencing Knowledge Connections in Co-word Networks

Based on the theoretical framework above, we examine five influencing factors listed in Table 1 , comprising three network structural variables and two keyword attribute variables. Node degree centrality, betweenness centrality, and clustering coefficient are measured using social network analysis methods. The earliest year a keyword appears in the field, min_{year}, is determined based on the keyword' s associated papers. We employ a keyword disciplinary affiliation calculation method [27] to obtain keywords' disciplinary classification attributes. The basic principle is that a keyword' s degree of affiliation with

a discipline is proportional to its frequency in that discipline' s literature and inversely proportional to the discipline' s publication volume. For keyword K, if it appears C_T times in discipline T' s publications, and the total number of publications in that discipline is P_T, then keyword K' s degree of affiliation with the discipline, S_{KT}, is:

$$S_{KT} = \frac{C_T}{\sum_n}$$

where N is the total number of disciplines. We calculate keyword K' s affiliation degree across all N disciplines and assign it to the discipline with the maximum value.

Table 1 Factors Influencing Knowledge Connections in Co-word Networks

Hypothesis	Variable Type	Variable	Type	Description
H1	Endogenous	degree	Continuous	Node degree centrality
H2	Endogenous	betweenness	Continuous	Node betweenness centrality
H3	Endogenous	cluster_{coef}	Continuous	Node clustering coefficient
H4	Exogenous	min_{year}	Continuous	Keyword' s earliest appearance year in the field
H5	Exogenous	wos_{category}	Categorical	Keyword' s disciplinary classification (Medical Informatics, Medicine, Health Care, Computer Science, Statistics)

2.3 Exponential Random Graph Model Construction

We employ the Exponential Random Graph Model method to investigate the formation mechanisms of keyword co-occurrence relationships in interdisciplinary co-word networks. ERGM, based on relational data and dependency assumptions, selects local network structures as network statistics to observe complex networks' overall structural features, thereby obtaining holistic understanding

of network complexity, relatedness, and randomness [28]. The fundamental ERGM formula is:

$$\Pr(Y = y) = \frac{1}{k} \exp\left\{\sum_A \eta_A g_A(y)\right\}$$

where the summation includes all statistical variables A, _A represents the parameter corresponding to statistical variable A, g_A(y) = {y{ij} A} y_{ij} is the network statistic for the corresponding variable, and k is a normalization constant ensuring the formula represents a proper probability distribution between 0 and 1 [29]. In specific research, the set of statistical variables A can be designed according to research needs. This study' s model variable set is shown in Table 1.

In essence, ERGM' s core task is a process of assigning weights to networks with specific mechanism combinations. Therefore, the formula can also be expressed in conditional Logit form [28, 30]:

$$\text{Logit}\frac{\Pr(Y_{ij} = 1|Y_{ij}^C)}{\Pr(Y_{ij} = 0|Y_{ij}^C)} = \sum_A \eta_A[g_A(Y_{ij}^+, X) - g_A(Y_{ij}^-, X)]$$

where Y_{ij} represents a new co-occurrence relationship in the co-word network, Y_{ij}^C denotes all other co-occurrence relationships in the network except Y_{ij}, and Logit calculates the log-odds ratio of the probability of a relationship forming versus not forming; g_A(Y_{ij}^+, X) - g_A(Y_{ij}^-, X) represents the change in network statistics; _A is the corresponding estimated parameter. Through this formula, we can determine how many times more likely a co-occurrence relationship is to form than not to form, given a one-unit change in specific co-occurrence network statistical variables (including node attributes and network structural attributes) as the relationship formation probability changes from 0 to 1. In ERGM, this coefficient is also known as the log-odds. When a co-occurrence relationship' s log-odds is calculated as β, it can be understood that, as the relationship formation probability ranges from 0 to 1, influenced by a one-unit change in specific co-occurrence network statistical variables, the probability of the relationship forming is e^β times the probability of it not forming.

Based on the above design, we construct and solve the ERGM using the statnet package in R [28]. During empirical analysis, we examine the effects of exogenous and endogenous variables on network formation by constructing four models for comparison: a null model, a network structure model, a node attribute model, and a comprehensive model. The null model serves as a baseline, examining only the effect of network edge count. The network structure model adds endogenous variables to the null model. The node attribute model adds exogenous variables to the null model, examining both unary attributes and

homophily between binary nodes for disciplinary attributes. The comprehensive model includes both endogenous and exogenous variables in the null model. Table 2 provides model details.

Table 2 Details of Four Empirical ERGMs

Model Type	Model Specification	Variables
Null Model	keyword_{network} ~ edges	1. Number of edges in network structure
Network Structure Model	keyword_{network} ~ edges + nodecov(“degree”) + nodecov(“betweenness”) + nodecov(“cluster_{coef}”)	1. Number of edges in network structure; 2. Node degree centrality in network structure; 3. Node betweenness centrality in network structure; 4. Node clustering coefficient in network structure
Node Attribute Model	keyword_{network} ~ edges + nodefactor(“wos_{category}” , base=4) + nodematch(“wos_{category}” , diff=TRUE) + absdiffcat(“min_{year}”)	1. Number of edges in network structure; 2. Keyword’ s earliest appearance year - difference effect; 3. Keyword’ s disciplinary classification attribute; 4. Homophily in disciplinary classification between node pairs

Model Type	Model Specification	Variables
Comprehensive Model	keyword_{network} ~ edges + nodefactor(“wos_{category}” , base=4) + nodematch(“wos_{category}” , diff=TRUE) + absdiffcat(“min_{year}”) + nodecov(“degree”) + nodecov(“betweenness”) + nodecov(“cluster_{coef}”)	1. Number of edges in network structure; 2. Node degree centrality in network structure; 3. Node betweenness centrality in network structure; 4. Node clustering coefficient in network structure; 5. Keyword’ s earliest appearance year - difference effect; 6. Keyword’ s disciplinary classification attribute

3 Empirical Analysis

3.1 Interdisciplinary Field Selection and Data Acquisition

We selected Medical Informatics as our case study for empirical investigation of an interdisciplinary field. Medical Informatics exhibits clear interdisciplinary characteristics [31] and has developed over a long period with substantial literature. To define the field’ s scope, we employed Bradford’ s Law principle, which states that the majority of a discipline’ s papers come from a small number of core journals [32], using core journals to define the disciplinary boundary. Using the Web of Science (WOS) database as our data source for Medical Informatics, we identified 24 journals from the 2016 WOS Medical Informatics category as source journals, retrieving 37,650 article-type bibliographic records from 1900 to 2016, including metadata such as titles, author keywords, system keywords, journals, publication dates, and disciplinary classifications.

To calculate keywords’ disciplinary attributes, we first identified Medical Informatics’ key foundational disciplines by analyzing the disciplinary classifications of cited journals in Medical Informatics references. We selected four disciplines as related foundational disciplines: Medicine, Health Care, Computer Science, and Statistics. We collected bibliographic information from journals in these four disciplines for the same 1900-2016 period. During preprocessing, we identi-

fied keywords across all five disciplines and calculated their disciplinary affiliation degrees. Due to substantial missing author keywords, we also used system keywords and extracted noun phrases from titles as title keywords, integrating three keyword types to improve coverage. After data cleaning, we obtained our final keyword set. Table 3 lists basic information for the five disciplines, including numbers of journals, papers, keywords, and total keyword frequencies.

Table 3 Basic Information for Five Disciplines

Discipline	Papers	Keywords	Total Keyword Frequency
Medical Informatics	37,650	101,013	1,401,86
Medicine	-	-	-
Health Care	-	-	-
Computer Science	-	-	-
Statistics	-	-	-

3.2 Co-word Network Construction

As Table 3 shows, Medical Informatics contains a large number of keywords. Since statnet supports only limited network sizes, and keywords with very low frequency are more likely to be influenced by random factors, keyword selection is necessary. Following J.C. Donohue' s high-frequency/low-frequency word demarcation formula [33], we selected the top 1,000 high-frequency words. Since words ranked 995 to 1,013 all had a frequency of 46, we included the top 1,013 high-frequency words for network construction. If two keywords appeared together in an article, we considered a co-occurrence relationship between them, temporarily ignoring co-occurrence strength.

The Medical Informatics co-word network contains 1,013 nodes and 140,186 edges, with network density of 0.273 and clustering coefficient of 0.138. This indicates that, compared to typical social networks, co-word networks constructed from high-frequency words are relatively dense. We then calculated appearance frequency and earliest appearance year for each of the 1,013 keywords and determined their disciplinary attributes. Among these keywords, minimum frequency was 46, maximum frequency was 2,616, and average frequency was 161. The earliest keyword appeared in 1964 and the latest in 2013, representing a time span of 0-49 years. Based on the temporal distribution of keywords (Figure 2 [Figure 2: see original paper]), most keywords emerged between 1984 and 1998, suggesting that the field introduced or produced substantial new knowledge during this period, with fewer new keywords after 1999, indicating the field' s entry into a mature stage. Figure 3 [Figure 3: see original paper] presents the disciplinary distribution of these 1,013 keywords, showing Medical Informatics has the most keywords (424), followed by Health Care (269), while Medicine has the fewest (83).

3.3 Model Results Analysis and Discussion

We estimated parameters for the null, network structure, node attribute, and comprehensive models using the statnet package with Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMCMLE). Model fit was evaluated using AIC and BIC, where smaller values indicate better fit. Table 4 presents parameter estimates for the four models. Based on AIC and BIC values, the network structure model outperforms the node attribute model, indicating that endogenous factors have greater influence on relationship formation than exogenous factors. From a systems perspective, when describing an interdisciplinary knowledge system as a co-word network, the network' s structural properties become the primary drivers of relationship formation, reflecting the advantage of using co-word networks to represent domain knowledge systems.

The comprehensive model shows the best fit, indicating that endogenous and exogenous factors do not operate independently but interactively. This model' s fitted data most closely approximates the real network, and subsequent analysis primarily relies on the comprehensive model' s parameters to examine specific effects of each influencing factor.

3.3.1 Endogenous Variable Mechanisms In the comprehensive model, degree centrality shows a significant positive correlation. This indicates that, holding other factors constant, keywords with higher degree centrality are more likely to form co-occurrence relationships with other keywords. This result confirms that the universal preferential attachment mechanism in complex networks significantly influences co-word network formation [34], supporting H1. This conclusion aligns with findings from power-law degree distributions in co-word networks [16]. However, betweenness centrality shows no significant correlation (0.00, p-value < 0.001), indicating that a keyword' s intermediary position does not affect its likelihood of being linked by other keywords, thus H2 is not supported. Clustering coefficient shows significant positive correlation (0.52, p-value < 0.001) with a relatively large value. This demonstrates that, all else being equal, keywords with higher clustering coefficients are more likely to form co-occurrence relationships with other keywords. A larger clustering coefficient indicates greater capacity for a node to cluster with other nodes, making new nodes more likely to establish connections with it, confirming H3.

3.3.2 Exogenous Variable Mechanisms The comprehensive model examines how two exogenous variables—keywords' earliest appearance year and disciplinary classification—affect co-occurrence relationship formation. For the earliest appearance year variable, we used difference analysis to examine how temporal differences influence co-occurrence formation. The time difference range is [1-49] years; due to space limitations, Table 4 does not show all results, which are instead presented in Figure 4 [Figure 4: see original paper]. Figure 4 displays only the 44 significant parameter estimates from the comprehensive model, omitting 5 non-significant results that indicate unclear effects of certain year dif-

ferences.

All parameter estimates for keyword earliest appearance year differences are negative, indicating that temporal difference significantly negatively affects co-occurrence formation. Compared to keywords appearing in the same year, keywords from different years are less likely to form co-occurrence relationships. Moreover, as the time difference increases, the parameter estimates decrease, indicating progressively lower likelihood of co-occurrence formation. This phenomenon suggests that in Medical Informatics, keywords tend to connect with newer keywords in the field, indicating that new knowledge is more likely to generate knowledge connections and derive emergent knowledge. This result confirms H4, which can be explained by the research hotspot emergence process in scientific research: scientific research builds continuously upon existing discoveries, with new findings stimulating more related scientific questions and forming new research hotspots. In co-word networks, this manifests as new keywords tending to connect with keywords appearing in similar time periods.

Keyword disciplinary classification is a categorical variable for which we designed two tests: (1) main effects, examining how a single node' s disciplinary category affects edge formation; and (2) homophily between binary node pairs, testing how disciplinary consistency between connected nodes affects edge formation. For main effects, using Medical Informatics as the reference category, results show significant negative correlations for all other disciplines. This indicates that compared to Medical Informatics keywords, keywords from Medicine, Health Care, Computer Science, and Statistics have lower probabilities of establishing edges with other keywords. Conversely, Medical Informatics keywords have the highest probability of being selected as co-occurrence partners. Across the entire network, keywords tend to preferentially establish co-occurrence relationships with Medical Informatics keywords.

Table 4 shows homophily test results for disciplinary classification: parameter estimates for Medicine, Health Care, Computer Science, and Statistics are 0.86, 1.49, 0.42, and 2.28 respectively, all significantly positive. This indicates that keywords from these four disciplines tend to form co-occurrence relationships with keywords from their own disciplines. Meanwhile, Medical Informatics shows significant negative homophily (-0.13), indicating that Medical Informatics keywords are less likely to form co-occurrence relationships within their own discipline and instead tend to connect with keywords from the other four disciplines. This finding refines H5.

In summary, keywords from Medicine, Health Care, Computer Science, and Statistics preferentially tend to form co-occurrence relationships with keywords from their own disciplines, secondarily with Medical Informatics keywords, showing certain path dependencies. However, Medical Informatics keywords preferentially tend to form co-occurrence relationships with keywords from other disciplines, reflecting the interdisciplinary nature of this field in borrowing, transplanting, and integrating knowledge from other disciplines.

The co-word network in interdisciplinary fields represents a concrete manifestation of multi-disciplinary knowledge connections, where keyword co-occurrence relationships embody micro-level knowledge connections. Starting from interdisciplinary co-word networks and employing ERGM, we examined influencing factors and their specific effects on keyword co-occurrence relationship formation from both endogenous and exogenous perspectives to understand micro-level knowledge connection mechanisms in interdisciplinary fields, particularly the roles of different disciplinary knowledge. Using Medical Informatics as a case study, we found: (1) From a network systems perspective, network structure influences co-occurrence formation more than keyword attributes, with preferential attachment and transitivity mechanisms substantially affecting co-word network formation. This demonstrates the advantage of using complex networks to represent domain knowledge systems and reveals internal structural patterns through network embedding theory. (2) Keyword nodes tend to connect with newer knowledge nodes, with new knowledge nodes more likely to derive new knowledge, a pattern explainable by the research hotspot emergence process. (3) Medical Informatics keywords tend to establish co-occurrence relationships with keywords from related foundational disciplines, reflecting the field' s interdisciplinary characteristic of absorbing foundational disciplinary knowledge. Additionally, we found that in this interdisciplinary field, keywords from foundational disciplines tend to connect with keywords within their own disciplines, indicating that interdisciplinary application of foundational knowledge follows certain paths rather than comprehensive integration.

This study has several limitations. We examined only a few factors influencing co-word relationship formation; additional factors such as keyword semantics may exist. We used a simple voting-based disciplinary affiliation index to determine keywords' disciplinary classification, which does not account for keywords belonging to multiple disciplines or the fuzziness and interdisciplinary nature of keyword disciplinary 归属 [14, 35]. We studied only Medical Informatics; whether our conclusions are affected by disciplinary differences remains untested. Future research should address these issues, conduct comparative analyses across multiple interdisciplinary fields, and employ statistical physics models to corroborate our findings.

References

[1] Liu Zhonglin, Zhao Xiaochun. Interdisciplinary research: The driving force of original scientific achievements—Taking Nobel Prize-winning achievements in Physiology and Medicine as examples [J]. Science Technology and Dialectics, 2005, 22(6): 105-109.

[2] Xu Haiyun, Yin Chunxiao, Guo Ting, et al. A review of interdisciplinary research [J]. Library and Information Service, 2015, 59(5): 119-127.

[3] Zhang Chengzhi, Wu Xiaolan. A review of interdisciplinary research [J]. Journal of the China Society for Scientific and Technical Information, 2017,

36(5): 523-535.

[4] Leydesdorff L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals [J]. Journal of the American Society for Information Science and Technology, 2007, 58(9): 1303-1319.

[5] Porter A, Cohen A, David Roessner J, et al. Measuring researcher interdisciplinarity [J]. Scientometrics, 2007, 72(1): 117-147.

[6] Morillo F, Bordons M, Gómez I. An approach to interdisciplinarity through bibliometric indicators [J]. Scientometrics, 2001, 51(1): 203-222.

[7] Ma Feicheng, Liu Xiang. Evolution model of scientific knowledge networks (I) [J]. Systems Engineering—Theory & Practice, 2013, 33(2): 437-443.

[8] Yan E, Ding Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other [J]. Journal of the American Society for Information Science and Technology, 2012, 63(7): 1313-1326.

[9] Callon M, Courtial J P, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry [J]. Scientometrics, 1991, 22(1): 155-205.

[10] Wang Xiaoguang. Formation and evolution of scientific knowledge networks (I): Proposal of co-word network method [J]. Journal of the China Society for Scientific and Technical Information, 2009(4): 599-605.

[11] Ding Y, Chowdhury G G, Foo S. Bibliometric cartography of information retrieval research by using co-word analysis [J]. Information Processing & Management, 2001, 37(6): 817-842.

[12] Hu Jiming, Zhang Xiaojuan, Tan Jing. Thematic structure and evolution of Chinese government information resources research [J]. Journal of Information Resources Management, 2018, 8(3): 54-63, 36.

[13] Li Gang, Ba Zhichao. Research on several issues in co-word analysis process [J]. Journal of Library Science in China, 2017, 43(4): 93-113.

[14] Xu Haiyun, Guo Ting, Yue Zenghui, et al. Research on interdisciplinary topics in information science based on TI indicator series [J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(10): 1067-1078.

[15] Hu J, Zhang Y. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization [J]. Scientometrics, 2017, 112(1): 91-109.

[16] Wang Xiaoguang. Formation and evolution of scientific knowledge networks (II): Co-word network visualization and growth dynamics [J]. Journal of the China Society for Scientific and Technical Information, 2010(2): 314-322.

[17] Robins G, Pattison P, Kalish Y, et al. An introduction to exponential random graph (p*) models for social networks [J]. Social Networks, 2007, 29(2):

173-191.

[18] Jiao C, Wang T, Liu J, et al. Using exponential random graph models to analyze the character of peer relationship networks and their effects on the subjective well-being of adolescents [J]. Frontiers in Psychology, 2017, 8: 583.

[19] Zhang C, Bu Y, Ding Y, et al. Understanding scientific collaboration: Homophily, transitivity, and preferential attachment [J]. Journal of the Association for Information Science and Technology, 2018, 69(1): 72-86.

[20] Yang Guancan, Chen Liang, Zhang Jing, et al. An explanatory framework for patent citation relationship formation: An exponential random graph model perspective [J]. Library and Information Service, 2019, 63(5): 100-109.

[21] Pol J V D. Introduction to network modeling using Exponential Random Graph Models (ERGM): Theory and application using R-Project [J/OL]. Computational Economics, 2018: 1-31 [2019-03-20]. https://doi.org/10.1007/s10614-018-9853-2.

[22] Peng T Q. Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks [J]. Journal of Informetrics, 2015, 9(2): 250-262.

[23] Barabási A L, Ravasz E, Vicsek T. Deterministic scale-free networks [J]. Physica A: Statistical Mechanics and its Applications, 2001, 299(3-4): 559-564.

[24] Ma Feicheng, Liu Xiang. Evolution of knowledge networks (III): Connection mechanisms [J]. Journal of the China Society for Scientific and Technical Information, 2011, 30(10): 1015-1021.

[25] Bianconi G, Darst R K, Iacovacci J, et al. Triadic closure as a basic generating mechanism of communities in complex networks [J]. Physical Review E, 2014, 90(4): 042806.

[26] Ma Feicheng, Liu Xiang. Evolution of knowledge networks (II): Growth, aging, and knowledge generation timing [J]. Journal of the China Society for Scientific and Technical Information, 2011, 30(9): 916-921.

[27] Lü Shuang. Domain analysis of international knowledge management research II: In-depth mining of disciplinary distribution [J]. Journal of Intelligence, 2012, 31(3): 118-123.

[28] Handcock M S, Hunter D R, Butts C T, et al. statnet: Software tools for the representation, visualization, analysis and simulation of network data [J]. Journal of Statistical Software, 2008, 24(1): 1-9.

[29] Roseki M, Jiyou N, Howard M, et al. Understanding network formation in strategy research: Exponential random graph models [J]. Strategic Management Journal, 2016, 37(1): 22-44.

[30] Wasserman S, Pattison P. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p* [J]. Psychometrika, 1996,

61(3): 401-425.

[31] Qi Yan, Xu Haiyun, Fang Shu. Research on the interdisciplinary development trend of medical informatics based on WOS data [J]. Chinese Journal of Medical Library and Information Science, 2016, 25(11): 30-41.

[32] Qiu Junping. Informetrics [M]. Wuhan: Wuhan University Press, 2007.

[33] Donohue J C. Understanding scientific literature: A bibliographic approach [M]. Massachusetts: The MIT Press, 1973.

[34] Barabási A L. Scale-free networks: A decade and beyond [J]. Science, 2009, 325(5939): 412-413.

[35] Kwon S. Characteristics of interdisciplinary research in author keywords appearing in Korean journals [J]. Malaysian Journal of Library & Information Science, 2018, 23(2): 77-93.

Author Contributions

Cao Yujie: Empirical analysis, paper writing
Li Gang: Research conceptualization, paper writing
Mao Jin: Research conceptualization, paper writing, data collection
Yang Guancan: Empirical analysis

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