天津大学

本科生毕业设计（论文）外文资料

外文原文题目：Feature analysis of multidisciplinary scientific collaboration patterns based on PNAS

中文译文题目：基于 PNAS 的多学科科学合作模式的特色分析

毕业设计（论文）题目：美国科学研究系统建模及合作模式挖掘

学　　院　管理与经济学部

专　　业　信息管理与信息系统

年　　级　2019 级

姓　　名　蒋世华

学　　号　3019209018

指导教师　王文俊 教授

Noname manuscript No.
(will be inserted by the editor)

Feature analysis of multidisciplinary scienti c collaboration patterns based on PNAS

Zheng Xie Miao Li Jianping Li
Xiaojun Duan Zhenzheng Ouyang

Received: date / Accepted: date

Abstract The features of collaboration patterns are often considered to be di erent from discipline to discipline. Meanwhile, collaborating among dis-ciplines is an obvious feature emerged in modern scienti c research, which incubates several interdisciplines. The features of collaborations in and among the disciplines of biological, physical and social sciences are analyzed based on 52,803 papers published in a multidisciplinary journal PNAS during 1999 to 2013. From those data, we found similar transitivity and assortativity of collab-oration patterns as well as the identical distribution type of collaborators per author and that of papers per author, namely a mixture of generalized Pois-son and power-law distributions. In addition, we found that interdisciplinary research is undertaken by a considerable fraction of authors, not just those with many collaborators or those with many papers. This case study provides a window for understanding aspects of multidisciplinary and interdisciplinary collaboration patterns.

Keywords Collaboration pattern Interdiscipline Hypergraph Complex network

1 Introduction

Natural and social sciences provide methodical approaches to study, predict and explain natural phenomena and sociality (human behaviors and psy-

Z. Xie Correspondence: E-mail: xiezheng81@nudt.edu.cn, J. Li E-mail: jian-pingli65@nudt.edu.cn, X. Duan E-mail: xjduan@nudt.edu.cn, Z. Ouyang E-mail: zzouyang@nudt.edu.cn
College of Science, National University of Defense Technology, Changsha, 410073, China

M. Li E-mail: limiaojoy@sjtu.edu.cn
School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, 200240, China The rst three authors have contributed equally to this work.

chological states) respectively [1]. The specialization of knowledge in these sciences forms various disciplines. Meanwhile, to solve problems whose solu-tions are beyond the scope of a single discipline, researchers need to integrate data, techniques, concepts, and theories from several disciplines [2-5]. Interac-tions between disciplines incubate several interdisciplines, fuzz the boundary of natural and social sciences, and produce many important scienti c break-throughs [6-8].

Studying collaboration patterns within and across disciplines or sciences contributes to understand the diversity of cooperative behaviors and fusion modes of knowledge. Papers of multidisciplinary journals provide an informa-tive and reliable platform for this studying, because the media of natural and social sciences mainly count on papers [9-12]. Here we investigated the pat-terns based on 52,803 papers published in Proceedings of the National Academy of Sciences (PNAS) over the years 1999-2013.[a] The content of dataset spans three science categories: social sciences and two principal sub-sciences in nat-ural sciences, viz. biological and physical sciences.

Collaboration relationship can be expressed by graphs, termed as coau-thorship networks. Hence the patterns can be studied in network perspective. Coauthorship networks from di erent scienti c elds appear speci c similari-ties, such as partial transitivity of coauthorship, homophily on the number of collaborators, the right-skewed distribution of collaborators per author [13-19]. These commonalities also appear in the collaboration networks of three author sets (which come from the three science categories of PNAS respectively). We dived more into the rule and reason of these commonalities. We found that the distribution of collaborators per author and that of papers per author follow the same distribution type: a mixture of a generalized Poisson distribution and a power-law. We provided a possible explanation for the distribution type and these commonalities through the diversity of author abilities to attract collaborations.

A range of previous works discussed quantitative indexes of interdisciplinar-ity for sciences [20-22], for disciplines [23-26], for universities [27], for jour-nals [28,29], and for research teams [30]. Some works addressed the correlation between interdisciplinarity and scienti c impact [31-34] (e. g. citation catch-ing ability [35-37]). Based on speci c general ideas of these references, we studied interdisciplinary activities of PNAS through paper co-occurrence of disciplines, and through some indexes calculated based on the co-occurrence, such as Rao-Sterling diversity [38], and betweenness centrality [39].

We further studied the collaboration patterns across disciplines, and found that a considerable proportion of authors and papers in physical and social sciences involved in interdisciplinary research. The multidisciplinary coauthor-ship network extracted from the data has a giant component, which contains more than 88%, 80% and 71% authors in biological, physical and social sciences respectively. A considerable number of authors contribute to the formation of giant component. The contributions of author activity and productivity to the formation increase over time. The high extent of interdisciplinarity shown by the case study might not be representative of general collaboration patterns,

because authors could submit more interdisciplinary work to multidisciplinary journals than domain speci c ones.

This report is structured as follows: the data processing is described in Section 2; the similarities and interactions are analyzed in Section 3; and the discussion and conclusion are drawn in Section 4.

2 The Data

2.1 Reason for using the data

The case study involves two concepts, namely multidisciplinarity (researchers from di erent disciplines study within their disciplines) and interdisciplinar-ity (study beyond disciplinary boundaries) [40]. Multidisciplinarity could be viewed as a combination of disciplines, and interdisciplinarity as a merging of them. A multidisciplinary journal with the scope covering natural and social sciences can be utilized to analyze the interactions between science categories. Such journal can be also utilized to compare the collaboration patterns of multi-disciplines and nd similarities. PNAS publishes high quality research papers, and provides reliable discipline information of those papers. The jour-nal also provides a high quality data platform for analyzing worldwide collab-oration patterns, because nearly half of its papers come from authors outside the United States.

Multidiscipline journals: Science, Nature and Nature Communications do not provide discipline information of papers. Journal of the Royal Society In-terface focuses on the cross-disciplinary research at the interface between the physical and life sciences, but does not involve social sciences. Our analysis is restricted to PNAS, which brings limitations to our ndings. For example, the media of social sciences not only count on papers, but also on books [11, 12]. Hence the results obtained must be carefully interpreted as being the patterns of researchers who publish papers in the chosen journal. However, due to the in uence and representability of PNAS, the case study could contribute to understanding aspects of multidisciplinary and interdisciplinary collaboration patterns.

2.2 Discipline information

Most papers of the dataset have been classi ed into three rst-class disciplines (biological, physical, and social sciences) and 39 second-class disciplines (Ta-ble 1). Interdisciplinary papers are classi ed into several disciplines. The data contain 43,304 biological papers (including 3,957 papers of biophysics), which account for 82.01% of the total. The data also contain 5,987 physical papers and 1,310 social papers. There are 2,961 interdisciplinary papers belonging to more than one of the second-class disciplines, which account for 5:61% of the total. The signi cant di erence of discipline proportion does not mean the

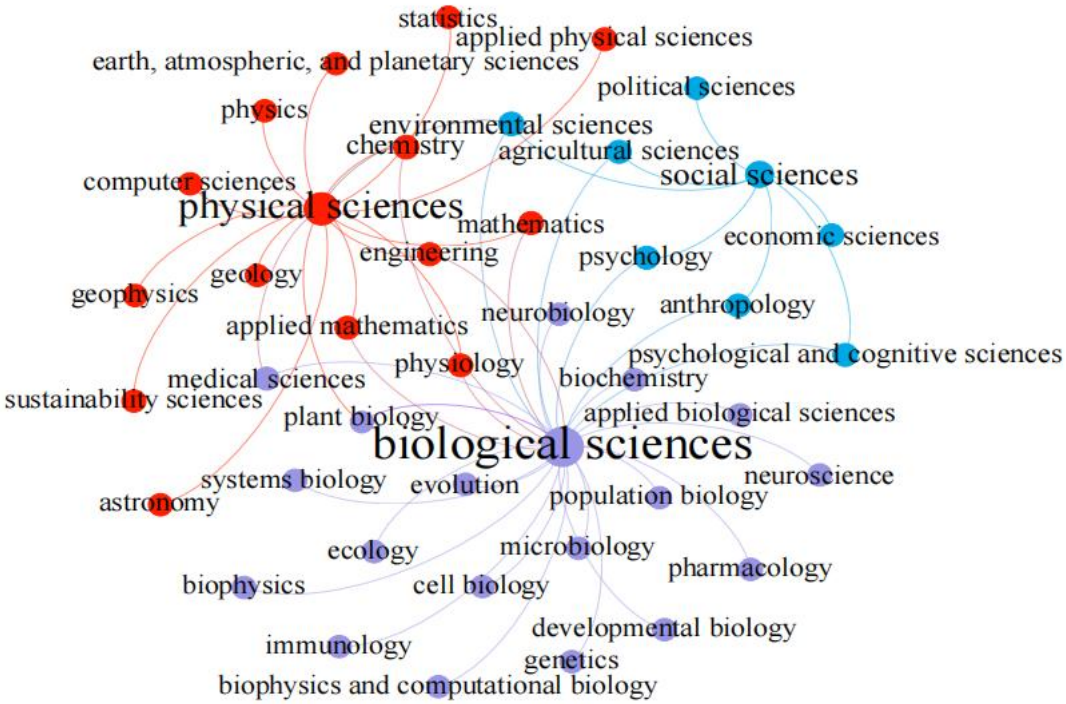

Figure 1 The relationship between the rst-class and the second-class disci-plines. The network is built based on the discipline information of papers in PNAS 1999-2013. Two disciplines are connected if they are the rst-class and the second-class disciplines of a paper. The node size indicates node degree.

preference for PNAS. In reality, the number of researchers involved in natural sciences (especially, biological sciences) is far more than that of researchers involved in social sciences [41]. There are 1,842 papers that are only classi ed into the rst-class disciplines. For these papers, their second-class discipline are regarded to be missing, but which have been regarded to be the same as their rst-class disciplines in our pervious work [42]. Hence the data in Table 1 are di erent from those in Reference [42].

Based on the discipline information of papers, we constructed a network to express the relationship between the rst-class and the second-class disci-plines (Fig. 1), where two disciplines are connected if they are the rst-class and the second-class disciplines of a paper. We can also construct a network to express the interactions between the second-class disciplines (Fig. 2), where each node is a discipline and two nodes are connected if there is a paper belong-ing to them simultaneously. These networks could evolve with the discipline information of newly published papers. So using the latest data, one may have a more comprehensive view.

2.3 Coauthorship

Identifying ground-truth authors, termed as disambiguating author names, is an important, time-consuming, but a necessary procedure of coauthorship

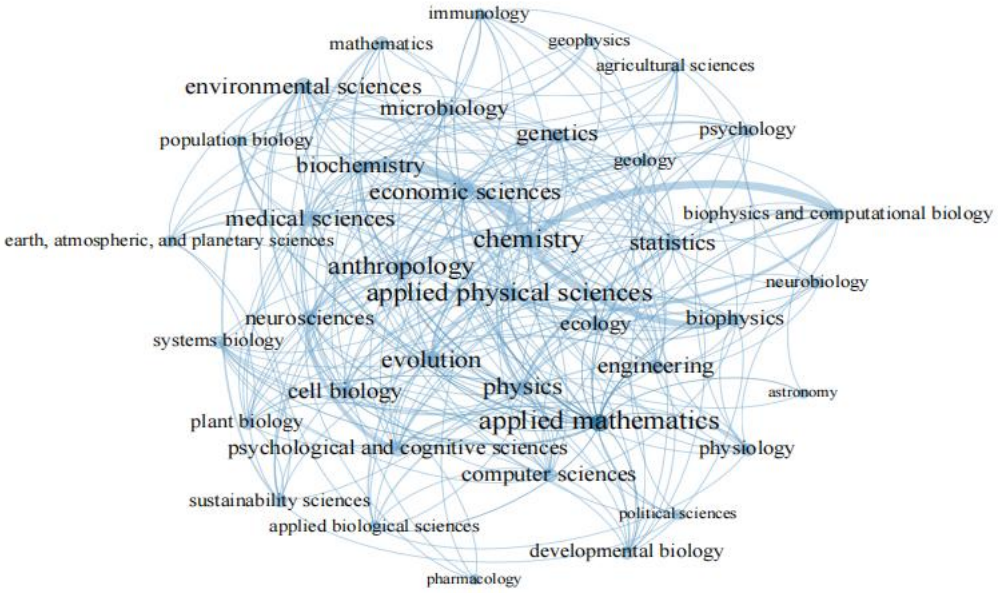

Figure 2 Interactions between the second-class disciplines. The weighted network is built based on the discipline information of interdisciplinary papers in PNAS 1999-2013. Edge width indicates edge weight: the number of interdisciplinary papers between two con-nected disciplines.

analysis. Several methods use the information of the provided names on papers (e. g. initial based methods [43]). The dominant misidenti cation of initial based methods is caused by merging two or more di erent authors as one. Hence, it de ates the number of unique authors, and in ates the size of the ground-truth giant component. Requiring additional information (e. g. email address) helps to reduce merging errors, but brings the di culty of collecting information.

In PNAS 1999-2013, 93:1% authors provide full rst name. So the provided names on papers are directly used to identify authors. However, utilizing sur-name and the initial of the rst given name will generate a lot of merging errors of name disambiguation [44]. The proportion of these authors in the data is 2:9%, and the proportion of these authors further conditioned on publishing more than one paper is 0:3%. Meanwhile, even utilizing full names still pro-duces merging errors, if some authors provide exactly the same name. Chinese names were found to account for name repetition [44]. We calculated the pro-portion of the names with a given name less than six characters and a surname among major 100 Chinese surnames.[b] The proportion of these authors in the data is 2:7%, and that of these authors further conditioned on publishing more than one paper is 1:1%. The small values of these four proportions show that the impact of name repetition is limited. These proportions for speci c subsets of the data are listed in Table 2.

The method adopted here will split one author as two or more, if the au-thor does not provide his name consistently. Splitting underestimates the giant component size, and the indexes used as evidences for universality of interdis-ciplinary research. Hence the results in Subsection 3.5, 3.6 could be regarded as

Table 1 Speci c indexes of the second-class disciplines in PNAS 1999-2013.

Disciplinary	m	n	k_1	k_2	b
Agricultural science	22	226	9	20	3.19
Anthropology	114	556	24	110	40.02
Applied biological science	135	767	9	134	1.79
Applied mathematics	191	380	27	182	49.39
Applied physical science	309	816	26	299	29.14
Astronomy	3	50	3	3	0.13
Biochemistry	333	6,303	19	327	16.96
Biophysics	359	3,957	16	359	7.91
Biophysics and computational biology	468	1,532	11	467	7.95
Cell biology	135	3,717	18	130	12.71
Chemistry	1,003	8,645	26	1,003	49.73
Computer science	77	101	17	70	9.50
Developmental biology	33	1,525	12	30	1.66
Earth, atmospheric, and planetary sciences	78	243	9	77	1.58
Ecology	162	1,084	15	162	10.00
Economic science	94	171	21	94	20.88
Engineering	217	392	19	217	13.85
Environmental science	184	695	20	183	25.44
Evolution	233	2,274	22	216	25.81
Genetics	103	2,664	20	97	12.68
Geology	137	285	10	136	2.79
Geophysics	23	175	7	23	1.51
Immunology	43	3,070	10	38	1.45
Mathematics	18	561	11	17	3.36
Medical science	181	4784	20	170	14.01
Microbiology	92	2,812	17	89	11.85
Neurobiology	16	1,003	9	16	0.87
Neuroscience	290	4,398	16	280	12.00
Pharmacology	26	594	4	26	0.08
Physics	229	4,818	22	227	18.24
Physiology	33	1,116	12	32	5.82
Plant biology	27	1,700	12	27	4.62
Political science	7	17	5	7	0.54
Population biology	27	166	11	26	4.04
Psychological and cognitive science	160	487	16	159	5.09
Psychology	83	449	12	83	3.62
Statistics	90	146	20	85	19.34
Sustainability science	123	399	11	120	7.66
Systems biology	36	159	11	36	1.80

The number of papers n and that of interdisciplinary papers m of a discipline are counted based on the discipline information provided by PNAS. The degree k_1, weighted degree k_2, and betweenness centrality b of a discipline are calculated based on the weighted network in Fig. 2.

conservative ones. In addition, the inaccuracy caused by the adopted method does not change the ground truth distribution type of collaborators per author and that of papers per author [44].

Table 2 Speci c statistical indexes of the analyzed networks.

Data	a	b	c	d
PNAS 1999-2013	2.9%	0.3%	2.7%	1.1%
Biological sciences	2.7%	0.2%	2.7%	1.1%
Physical sciences	4.8%	0.4%	4.4%	0.9%
Social sciences	2.3%	0.1%	2.2%	0.3%
Biophysics	4.1%	0.3%	4.0%	1.0%
Interdiscipline	2.6%	0.1%	3.6%	0.6%

Indexes a and b are the proportion of the authors only providing the initial of their rst given name and their surname, and that of these authors further conditioned on publishing more than one paper respectively. Indexes c and d are the proportion of the authors with a surname among the major 100 Chinese surnames and a given name less than six characters, and that of these authors further conditioned on publishing more than one paper respectively.

3 Data analysis

3.1 Network properties

Coauthorship is a n-ary relation, n 2 Z^{+}, hence it can be expressed by a hyper-graph, a generalization of a graph in which an edge (termed as hyperedge) can join any number of nodes. Represent authors as nodes, and the author group of each paper (paper team) as a hyperedge. Then we can extract a coauthor-ship network from a hypergraph as a simple graph, where edges are formed between every two nodes in each hyperedge, and the multiple edges are treated as one. The terms \degree" and \hyperdegree" for nodes are used to express the number of collaborators and that of papers for authors respectively.

The data show that the average paper team size of biological sciences (6.624) and that of physical sciences (5.254) are larger than that of social sci-ences (4.634). The size relation ts the reality that the sizes of research teams are usually larger in natural sciences, and smaller in social sciences [41]. Now let us consider the coauthorship networks of the considered papers in speci c disciplines or science categories. All of these networks are highly clustered, as-sortative, and their average shortest path length scale as the logarithms of their number of nodes (logNN AP in Table 3). These properties do not mean all of the networks are small-world. The network of social sciences is an exception, which even has no component containing more than 10% authors. However, it does not mean that the research in social sciences is carried out in isolation. In fact, 71.5% authors in social sciences belong to the giant component of the coauthorship network generated by the whole data. Therefore, analyzing the collaborations of authors restricting in single discipline has limitations. So we proceeded the analysis in the environment of all disciplines.

Table 3 Speci c statistical indexes of the analyzed networks.

Network	NN	NE	GCC	AC	AP	PG
PNAS 1999-2013	202,664	1,225,176	0.881	0.230	6.422	0.868
Biological sciences	184,872	1,150,362	0.881	0.232	6.364	0.880
Physical sciences	24,766	101,166	0.933	0.452	10.89	0.455
Social sciences	5,121	18,786	0.946	0.683	6.574	0.087
Biophysics	13,480	48,012	0.905	0.177	7.665	0.636
Interdiscipline	13,680	53,588	0.951	0.558	9.397	0.093

The indexes are the number of nodes (NN), the number edges (NE), global clustering coe cient (GCC), degree assortativity coe cient (AC), average shortest path length (AP), the node proportion of the giant component (PG). The AP of the rst two networks are approximately calculated by sampling 400,000 pairs of nodes.

3.2 Degree and hyperdegree

Aggregate degree and hyperdegree on the data (not restricted in single science category), and observe the degree distributions and hyperdegree distributions of three author sets (which come from the three science categories respec-tively). We found that although collaboration level di ers from one science category to another, all of the distributions emerge a hook head, a fat tail, and a cross-over between them, which could be viewed as a common feature of coauthorship networks (Fig. 3). The head and tail can be tted by log-normal distribution and power-law distribution respectively [45].

These distributions can also be tted, as a whole, by a mixture of a gener-alized Poisson distribution and a power-law distribution. The tting param-eters are listed in Table 4. We performed a two-sample Kolmogorov-Smirnov (KS) test to compare the distributions of two data vectors: node indexes (i. e. degrees, hyperdegrees), the samples drawn from the corresponding tting dis-tribution. The null hypothesis is that the two data vectors are from the same distribution. The p-value of each tting shows the test cannot reject the null hypothesis at 5% signi cance level. Note that ^{2} goodness-of- t test is not suitable here, due to the small number of large degree authors.

Regarding authors as samples, a mixture distribution means those sam-ples come from di erent populations, namely the collaboration patterns of the authors with few collaborators and papers di er from those with many collab-orators and papers. In Reference [46], a possible explanation (which is free of disciplines) is given for the emerged mixture type of empirical degree distribu-tions. With the same general ideas, a similar explanation can be adopted for hyperdegree distributions as follows.

The event whether a researcher collaborates with one another to publish a paper can be regarded as a \yes/no" decision. So the hyperdegree of a re-searcher is equal to the number of successes in a sequence of decisions made by the candidates who want to coauthor with that researcher. Suppose the num-ber of those candidates to be n. Suppose the collaboration probability of each candidate to be p. Then, the hyperdegrees will follow a binomial distribution B(n; p), and so a Poisson distribution with expected value np approxima-

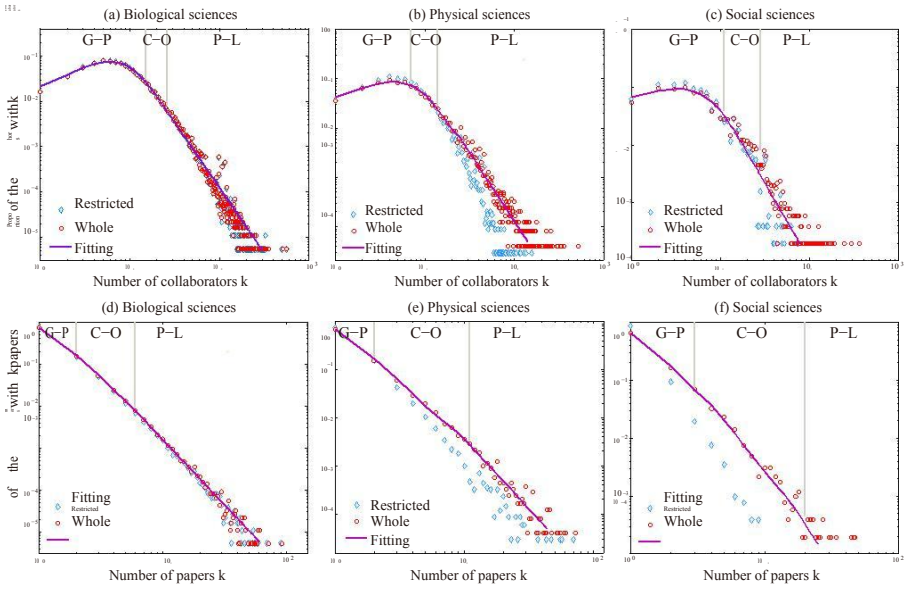

Figure 3 Distributions of collaborators/papers per author. The panels show the distributions counted in PNAS 1999-2013 (red plots), and those counted in the papers of each science category (blue diamonds). Fitting distributions (purple curves) are mixtures of a generalized Poisson distribution and a power-law distribution. Fitting parameters are listed in Table 4. The regions \G-P", \C-O", \P-L" stand for generalized Poisson, cross-over and power-law respectively.

tively (Poisson limit theorem). The value of np varies from author to author, due to the diversity of authors' ability to attract collaborators.

Decisions of authors could be dependent. For example, collaborating with the researchers who have publishing experience helps to publish a paper. Hence we could regard hyperdegree as a random variable following a generalized Pois-son distribution (which allows the occurrence probability of an event to involve memory [47]). In empirical data, most hyperdegrees are around their mode. Hence we could think of that they follow some generalized Poisson distribu-tions with an expected value around their mode, and so form the generalized Poisson part of a hyperdegree distribution. A few authors experience a cumu-lative process of papers, which makes a hyperdegree distribution skew to the right and form a fat tail.

3.3 Transitivity of coauthorship

Transitivity in society is that \the friend of my friend is also my friend", which is a typical feature of social a liation networks. In academic society, collaborators of an author likely acquaint and so coauthor with each other. For example, organizational and institutional contexts drive the formation of

Table 4 The parameters of tting functions.

Degree distribution	a	b	c	d	s	B	E	G	P	p-value
Biological sciences	4.843	0.464	74.27	2.889	1.049	15	26	20	50	0.203
Physical sciences	3.958	0.477	49.31	2.798	1.037	7	14	20	53	0.178
Social sciences	3.292	0.513	20.78	2.657	1.046	11	28	20	35	0.111
Hyperdegree distribution	a	b	c	d	s	B	E	G	P	p-value
Biological sciences	0.028	0.269	1.968	3.099	35.57	2	6	10	13	0.979
Physical sciences	0.021	0.320	2.977	2.916	47.15	2	11	10	10	0.625
Social sciences	0.022	0.375	19.48	3.665	46.24	3	20	10	11	0.206

The ranges of generalized Poisson $f_1(x)$, cross-over, and power-law $f_2(x)$ are [1; E], [B; E], and [B; max(x)] respectively. The tting function is $f(x) = q(x)sf_1(x) + (1\ q(x))f_2(x)$, where $q(x) = e^{(x\ B\ )=(E\ x)}$. The tting processes are: obverse proper G and P ; calculate parameters of $sf_1(x)$ (i.e. a, b, s) and $f_2(x)$ (i.e. c, d) through regressing the empirical distribution in [1; G] and [P; max(x)] respectively; nd B and E through exhaustion to make f(x) pass KS test (p-value> 0:05). The sum of each f(x) over [1; max(x)] is near unity, which means that f(x) can be regarded as a probability density function.

transitive coauthorship, and so contribute to clustering structures emerging in coauthorship networks.

The transitivity of a network can be quanti ed by two indexes in graph theory, namely global clustering coe cient (the fraction of connected triples of nodes which also form \triangles") and local clustering coe cient (the prob-ability of a node's two neighbors connecting). High transitivity is a common feature of coauthorship networks [15].

To what extent the transitivity is due to the activity of authors in academic society? The activity can be partly re ected through the number of collabora-tors, namely degree. Hence, the extent can be sketched through the correlation coe cients between degree and local clustering coe cient. Note that the cor-relation coe cients indicate the extent of a linear relationship between two variables or their ranks. The coe cients of variables X and Y generally do not completely characterize correlation, unless the conditional expected value of Y given X, denoted by E(Y jX), is linear or approximate linear function in X. The conditional expected value of local clustering coe cient given degree is the average local clustering coe cient of k-degree nodes, denoted by CC(k). The approximatively linear trend of CC(k) shown in Fig. 4 guarantees the ef-fectiveness of correlation analysis in Table 5. The decreasing trend cannot be deduced out from degree information. The denominator of the local clustering coe cient of a node grows quadratically with its degree, but the numerator cannot be calculated from degree information.

Does the decreasing trend of CC(k) mean activity depresses transitivity? A positive answer to it is against common sense. In PNAS 1999-2013, 74.62% authors only publish one paper in the data, and the paper team sizes of 99.9% papers follow a generalized Poisson part, namely are around the average pa-per team size 6.028. The boundary of generalized Poisson part is detected by the boundary point detection algorithm for probability density functions in Reference [46] (listed in Appendix). Hence the local clustering coe cients of

Table 5 Correlation coe cients between degree and transitivity/clustering in-dexes.

Discipline	Indicator	Mean	Std	SCC	PCC
Biological sciences	LCC	0.860		-0.398	-0.401
	LTC	0.001	0.005	0.275	0.077
	DN	21.09		0.543	0.400
	HN	3.015	15.47	0.070	-0.046
Physical sciences	LCC	0.806		-0.336	-0.382
	LTC	0.001	0.005	0.306	0.074
	DN	15.48		0.625	0.346
	HN	2.682	12.44	0.169	0.015
Social sciences	LCC	0.784		-0.177	-0.263
	LTC	0.001	0.006	0.292	0.050
	DN	12.87		0.723	0.482
	HN	2.268	10.89	0.175	0.030

The indexes are local clustering coe cient (LCC), the local transitivity of collaboration (LTC), the average degree of node neighbors (DN), the average hyperdegree of node neighbors (HN). We calculated the mean of these indexes over authors, the Spearman rank correlation coe cient (SCC) and Pearson product-moment correlation coe cient (PCC) between each index and degree. For the two indexes with small PCC, we calculated their standard deviation (Std).

most small degree authors are close to 1 (Fig. 4). A few authors experience a long period of collaborations, whose degree is obtained by accumulated over papers. For these authors, their collaborators in di erent papers could not col-laborate, which decreases their local clustering coe cient. Hence the puzzling thing does not contradict with common sense, but is due to insu ciency of measuring transitivity such a dynamical property by counting \triangles" on a static network.

To design a more reasonable index measuring transitivity, let us come back to the original meaning of transitivity on coauthorship: the probability of two collaborators (who do not coauthor yet) of a researcher coauthoring in future. The probability can be calculated for dynamic hypergraphs of collaborations through time information. Averaging the probability over authors measures the global transitivity, the value of which is quite low in each science cate-gory (Table 5). Note that the calculation is limited in PNAS 1999-2013, and transitivity may happen in other journals or in other time period. So the val-ues of transitivity here may be underestimated. The increasing trend of the transitivity probability of k-degree authors (TC(k) in Fig. 4) means the activ-ity contributes to transitivity. It ts common sense: a researcher with many collaborators is likely to introduce his collaborators to cooperate.

3.4 Homophily of coauthorship

Coauthorship is based on speci c features of researchers in common, including interest and geography. The homophily phenomenon appears in many social relations, and is called assortative mixing in network science [18]. Do authors

of each science category prefer to coauthor with others that are similar in so-cial activity or productivity? The social activity and productivity of authors can be quanti ed by two indexes, namely degree and hyperdegree respectively. Then the preference of an index could be sketched through the correlation co-e cient between two variables, namely the index of a author and the average index of the author's neighbors. Positive correlation means assortative, nega-tive disassortative, and zero no preference.

Degree assortativity is a feature of coauthorship networks [18]. Does it mean sociable researchers (with many collaborators) will preferentially coau-thor with other sociable researchers, and unsociable to unsociable? In a pre-vious study [48], we showed that the proportion of top 5.99% most sociable authors (measured according to degree) having coauthored with another such author is 99.5%. The proportion may even be underestimated, because these authors probably coauthored before 1999 or in other situations. Note that the splitting and merging errors of the used name disambiguation method a ect the proportion at certain levels. Even so, the proportion is still remarkable.

However, if sociable researchers only coauthor with sociable ones, then there will exist many sociable researchers, which is against empirical degree distributions. Now let us analyze the in uence of the social activity of authors on degree assortativity. For the authors with k-degree, denote the average degree of their neighbors by DN(k). There exists a trend change in DN(k) of each empirical dataset: the head part has a clear increasing trend, but the tail part does not (Fig. 4). It means that degree assortativity are mainly contributed by small degree authors.

The tipping point of the trend of DN(k) is detected by the boundary point detection algorithm for general functions in Reference [46] (listed in Appendix). Inputs of the algorithm are DN(k), $g(\ ) = \log(\ )$ and $h(x) = a_1x^3 + a_2x^2 + a_3x + a_4$ ($x, a_i \in R, i = 1; :::; 4$). Using those inputs is based on the observation of DN(k). Degrees of most authors are around their mode 5, and only a few authors have a large degree. Hence the neighbors of an author are likely to be small degree authors. Therefore, for small degree authors, the degree di erences between those authors and their neighbors are small, and large for large degree authors, which leads to the trend change of DN(k).

The correlation coe cient between hyperdegree and the average hyperde-gree of neighbors is around zero in each science category (Table 5). For the authors with k-hyperdegree, denote the average hyperdegree of their neighbors by HN(k). It means choosing collaborators is free of the factor of productivity. In reality, members of a research team may have various scienti c ages (new-comers, incumbents), so di erent hyperdegrees. Since collaborations mainly happen in a research team, collaborators of an author could have various hy-perdegrees, which appears as the stable trend of HN(k).

Based on the average value of HN(k) larger than 2, and 74.62% authors only having one paper in the data, we can derive that a large fraction of authors collaborate with at least one author who has published a paper in PNAS 1999-2013 to publish their rst paper in the data. The proportions of these authors are 79.22%, 71.17% and 65.12% in biological, physical and social

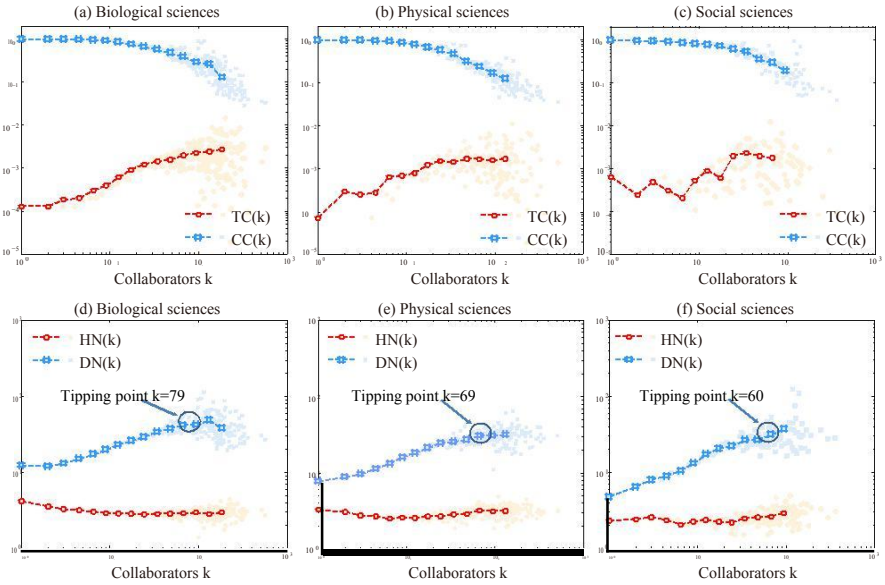

Figure 4 Conditional expected values of speci c indexes given degree. From k = 1 to max(degree), we average over k-degree nodes for local clustering coe cient (CC(k)), the local transitivity of collaborations (TC(k)), the average degree of node neighbors (DN(k)), and the average hyperdegree of node neighbors (HN(k)). The data are binned on abscissa axes to extract the trends hiding in noise.

sciences respectively. The proportions may be overestimated, because some of these authors may publish papers in PNAS before 1999.

3.5 Interdisciplinarity at discipline level

The co-category proportion measures the activities of interdisciplinary re-search. There are 49.2%, 46.0% and 7.3% authors of social, physical and bi-ological sciences who published interdisciplinary papers. The common sense suggests that social scientists engage in research solitary. The proportion of social sciences shows that the common sense does not hold in PNAS. Refer-ene [49] also shows, there has been a move towards increased interdisciplinarity in recent decades in social sciences.

Above analysis process could be implemented to the second-class disciplines to obtain a high-resolution result. However some disciplines only have a few papers, e. g. only 17 papers of political science. So the analysis for those disci-plines loses statistical meaning. Hence we took another perspective to analyze the interactions among the second-class disciplines by visualizing them as the network in Fig. 2. The network is connected, i. e. no discipline is isolated. Top three nodes of this network in terms of degrees and those in terms of between-ness centralities are Applied mathematics, Chemistry and Anthropology (Ta-ble 1). It means the theories, methods and problems of those disciplines are

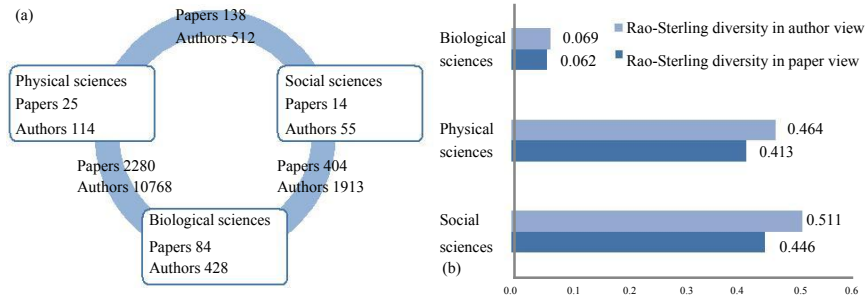

Figure 5 The interdisciplinary research of PNAS 1999-2013 between and within biological, physical and social sciences. Panel (a) shows the proportions of papers and those of authors involved in dyadic interactions between the three science categories, and those proportions involved in interactions within each science category. Panel (b) shows the Rao-Sterling diversity in paper/author view of each science category, which measures the discipline diversity of interdisciplinary research.

directly or indirectly used or studied by many disciplines. For each rst-class discipline, we contracted its second-class disciplines as one node, and calcu-lated the betweenness centrality of the contracted node. Their betweenness centrality (Biological sciences 47.51, Physical sciences 163.81, Social sciences 161.72) support the above analysis.

The co-category proportion only describes interdisciplinary activities. Now let us measure the discipline diversity of interdisciplinary research in each science category through Rao-Sterling index [38] $\Delta = \Sigma_{i,j(i\neq j)} d^{a}_{ij} (p_i p_j)^{\beta}$, where p_i and p_j are proportional representations of the papers/authors in science category i and j and d_{ij} is the level of dierence attributed to categories i and j. Discipline information is used to classify authors into science categories: if one of his papers belongs to a discipline, an author can be classi ed into the discipline, so into the corresponding sciences. Note that an author can be classi ed into several science categories, if his papers belong to more than one discipline. Here we let $= = d_{ij} = 1$ for all i and j, hence the calculated Rao-Sterling index measures the balance-weighted variety of interdisciplinary research in the level of science categories. The index in author view and that in paper view show that the discipline diversity of interdisciplinary research in social sciences and that in physical sciences are much higher than that in biological sciences (Fig. 5).

3.6 Interdisciplinarity at author level

We analyzed the relationship between author degree/hyperdegree and the probability of doing interdisciplinary research, and the relationship between paper team size and the probability of being an interdisciplinary paper. Fig. 6 shows that in each science category, interdisciplinary research is not just car-ried out by authors with a large degree or those with a large hyperdegree.

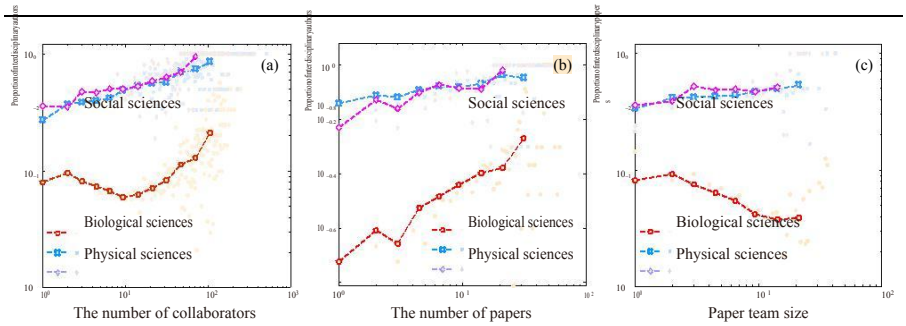

Figure 6 The relationship between authors' and papers' speci c indexes and their interdisciplinarity. Panels (a,b) show the relationship between author de-gree/hyperdegree and the probability of doing interdisciplinary research. Panel (c) shows the relationship between paper team size and the probability of being an interdisciplinary paper.

Fig. 6 also shows that large degree or hyperdegree authors are likely to en-gage in interdisciplinary research, and a paper with a large team size is likely to be an interdisciplinary one. It seems these phenomena can be expected at random. Take a set of elements (collaborators, papers) of several classes, and select a subset randomly. Then a larger subset more likely contains elements from more than one class. This reasoning, though plausible, is incorrect, be-cause scientists do not randomly select topic and collaborators. Research costs (investments of time and e ort) make scientists tend to work within their fa-miliar elds. In addition, the reasoning is based on that the selection scope of collaborators is limited to empirical data, which does not hold in reality.

We analyzed the giant component of coauthorship network PNAS 1999-2013, which contains more than 86.8% authors. There are 71.5%, 76.7% and 88.9% authors of social, physical and biological sciences in the giant compo-nent (Fig. 7e). Note that the author misidenti cation caused by initial-based methods increases the size of the ground-truth giant component [44]. Hence we identi ed authors by their provided names on papers (which likely split one author into two) to obtain a conservative result.

Interdisciplinary research and multidisciplinary research contribute to the giant component containing most authors of each science category. We ana-lyzed the relationship between the author proportion of the giant component and author activity/productivity. Remove authors from high degree and hy-perdegree to low respectively, and calculate the proportion of the giant com-ponent. From the relation curve between the proportion of removed authors and that of the giant component, we can nd that the formation of giant com-ponent is contributed by a considerable number of authors, e. g. the top 10% authors ranked by degree (Fig. 8). Consider the relationship in three time pe-riods, viz. 1999-2003, 2004-2008 and 2009-2013. The relation curve shifts to the left over time, which means author activity and productivity are playing increasingly important roles in the formation of the giant component.

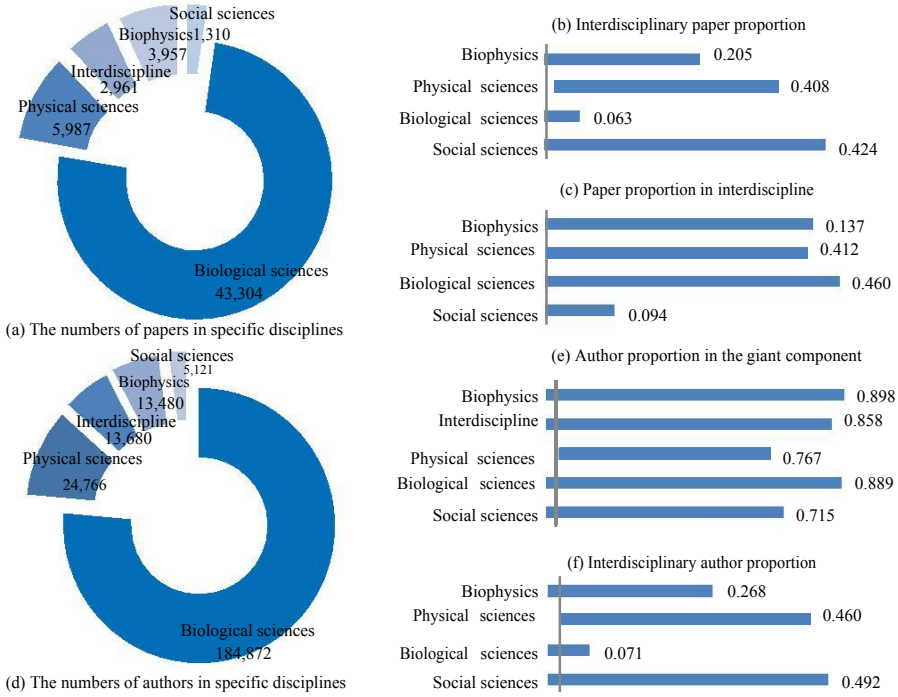

Figure 7 Interdisciplinary extents of speci c disciplines. For each considered disci-pline i, we denote its authors, its authors involved in interdisciplinary research, its papers, and its interdisciplinary papers by sets A_i, A^I_i , P_i and P_i^I respectively. Denote the giant component of coauthorship network PNAS 1999-2013 by S. The indexes are jP_ij in Panel (a), $jP_i \setminus I_ij=jI_ij$ in Panel (b), $jP_i \setminus P_i^I j=jP_i^I j$ in Panel (c), jA_ij in Panel (d), $jA_i \setminus Sj=jA_ij$ in Panel (e), and $jA_i \setminus A^I_i j=jA_ij$ in Panel (f).

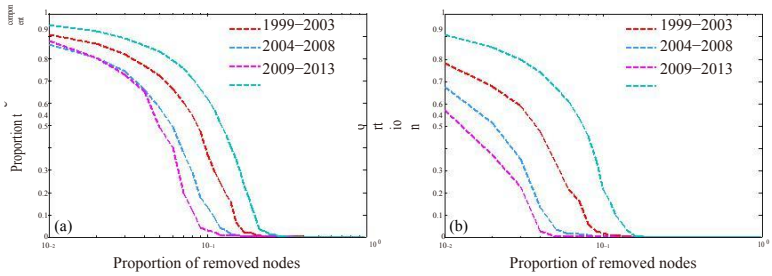

Figure 8 The relationship between giant component size and de-gree/hyperdegree. Nodes are removed from high degree/hyperdegree to low respectively. For degree and hyperdegree respectively, the relation curves between the proportion of removed nodes and that of the giant component show that a considerable number of authors contribute to the formation of giant component. The left-shifting trend of the relation curves in three time periods (1999-2003, 2004-2008 and 2009-2013) over time shows the increasing contributions of author activity and productivity to the formation of the giant component.

4 Discussion and conclusions

Our case study on PNAS 1999-2013 veri es the similar transitivity and assorta-tivity of collaboration patterns in biological, physical and social sciences. The data demonstrate that the degree distribution types of the three science cate-gories are identical, which are a mixture of a generalized Poisson distribution and a power-law. This also holds for hyperdegree. We provided an explanation for the emergence of this distribution type through authors' \yes/no" decisions and their di erent abilities to attract collaborations.

The data show that a considerable number of authors pursue interdisci-plinary research, and the giant component of coauthorship network PNAS 1999-2013 contains most authors of each science category. We took network perspective to analyze the interactions among the second-class disciplines, and quantify their interdisciplinarity by network indexes such as degree and be-tweenness centrality. We found that speci c second-class disciplines (such as Applied mathematics and Anthropology) play an important role in interdisci-plinary research.

The case study contributes to understanding multidisciplinary and inter-disciplinarity collaboration patterns, due to the importance of PNAS and to the accurate discipline information of its papers. The selection of data might a ect the details of our ndings about interdisciplinarity. Our results may not be interpreted as the patterns of general researchers. For example, we cannot expect to observe a high extent of interdisciplinarity by analyzing a domain speci c journal. We nished the case study by asking a question: What are the grounds of interdisciplinary research? While a thorough discussion of this question is beyond the scope of this paper, the following provides a simple discussion.

There is a tendency of fragmentation for disciplines in the development of sciences: going to split into sub-disciplines and speci c topics. Although the research objects are di erent, their research paradigms are in common, which can be grouped into four categories, namely theoretical research, experiment, simulation, and data-driven [50]. Meanwhile, many scienti c problems are too complex to be understood through the methodology of single discipline. In-tegrating theoretical and methodological perspectives drawn from di erent disciplines creates a uni ed methodology for research problems and even vo-cabulary used to present concepts in speci c disciplines [51], which drives the formation of transdisciplinary disciplines [52].

Systems science, as a typical transdisciplinary discipline, studies systems from simple to complex, from natural to social sciences. The parts of a sys-tem and the relations between parts can be abstracted as networks. The rapid development of research on networks (model, algorithm,...) breeds a new dis-cipline, namely network science. Some researchers from biological, physical and social elds investigate their respective problems under network frame-work [53], e. g. our case study.

To follow up the above, one would think that common research paradigms and methodology, especially those integrated as transdisciplinary disciplines,

give grounds for the interactions between science categories and for the forma-tion of giant components in coauthorship networks. It seems promising that analyzing paper content helps to validate the universality of those paradigms and methodologies. Over half the papers of PNAS 1999-2013 contain the topic words \system" and \control" [42]. The high proportion of the papers contain-ing a topic word at certain levels re ects the typicality of the topic. However, it is not easy to say which is the relation between a paper containing the word \system" and a paper applying research results of systems science. Hence val-idating the universality at semantic level is a subject for further study.

Availability of data and materials

The data are freely available from the their website http://www.pnas.org. Feel free to get in contact with the corresponding author in case you need more information.

Competing interests

The authors declare that they have no competing interests.

Funding

ZX acknowledges support from National Science Foundation of China (NSFC) Grant No. 61773020.

Authors' contributions

All authors conceived and designed the research. ZX and ML wrote the pa-per. ZX and JPL analyzed the data. OYZZ acquired the data. ZX and XJD wrote the discussion. All authors discussed the research and approved the nal version of the manuscript.

Acknowledgments

We thank the anonymous reviewers for their valuable suggestions and great help.

Endnotes

[a]http://www.pnas.org.
[b]Wikipedia shows that people with major 100 Chinese surnames account for 84.77% of the total Chinese population.

References

1. Weingart P (2012) A short history of knowledge formations. In R. Frodeman, J. Thomp-son Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 3-14). Oxford, England: Oxford University Press.
2. National Academies (U.S.). Committee on Facilitating Interdisciplinary Research (2004). Facilitating interdisciplinary research. Washington: National Academy Press. Retrieved from http://www.nap.edu/books/0309094356/html/
3. Hurd JM (1992) Interdisciplinary research in the sciences: Implications for library orga-nizations. Coll Res Liber 53(4), 283-297.
4. Cooper G (2013) A disciplinary matter: Critical sociology, academic governance and interdisciplinarity. Sociology 47(1), 74-89.
5. Hadorn GH, Pohl C, Bammer G (2012) Solving problems through transdisciplinary re-search. In R. Frodeman, J. Thompson Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 431-452). Oxford, England: Oxford University Press.
6. Siedlok F, Hibbert P (2014) The organization of interdisciplinary research: Modes, drivers and barriers. Int J Manage Rev 16(2), 194-210.
7. Liu Y, Rafols I, Rousseau R (2012) A framework for knowledge integration and di usion. J Doc 68(1), 31-44.
8. Gooch D, Vasalou A, Benton L (2017) Impact in interdisciplinary and cross-sector re-search: Opportunities and challenges. J Assoc Inf Sci Technol 68(2), 378-391.
9. Lariviere V, Gingras Y, Archambault E (2006) Canadian collaboration networks: A com-parative analysis of the natural sciences, social sciences and the humanities. Scientometrics 68(3): 519-533.
10. Moody J (2004) The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. Am Sociol Rev 69(2), 213-238.
11. Gl•anzel W, Schoep in U (1999) A bibliometric study of reference literature in the sciences and social sciences. Inform Process Manag 35(1): 31-44.
12. Hicks D (1999) The di culty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics 44(2): 193-215.
13. Sarig•ol E, P tzner R, Scholtes I, Garas A, Schweitzer F (2014) Predicting scienti c success based on coauthorship networks. EPJ Data Science 2014:9.
14. Barabasi AL, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scienti c collaborations. Physica A 311: 590-614.
15. Newman M (2001) The structure of scienti c collaboration networks. Proc Natl Acad Sci USA 98: 404-409.
16. Newman M (2004) Coauthorship networks and patterns of scienti c collaboration. Proc Natl Acad Sci USA 101: 5200-5205.
17. Xie Z, Ouyang ZZ, Li JP (2016) A geometric graph model for coauthorship networks. J Informetr 10: 299-311.
18. Newman M (2002) Assortative mixing in networks. Phys Rev Lett 89: 208701.
19. Tomasello MV, Vaccario G, Schweitzer F (2017) Data-driven modeling of collaboration networks: A cross-domain analysis. EPJ Data Science 6: 22.
20. Braun T, Schubert A (2003) A quantitative view on the coming of age of interdisci-plinarity in the sciences, 1980-1999. Scientometrics 58(1), 183-189.
21. Levitt JM, Thelwall M, Oppenheim C (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Assoc Inf Sci Technol 62(6), 1118-1129.
22. Porter AL, Roessner JD, Cohenm AS, Perreault M (2006). Interdisciplinary research: Meaning, metrics and nurture. Res Eval 15(3), 187-195.
23. Porter AL, Rafols I (2009) Is science becoming more interdisciplinary? Measuring and mapping six research elds over time. Scientometrics 81(3), 719-745.
24. Rafols I, Meyer M (2010) Diversity and network coherence as indicators of interdisci-plinarity: Case studies in bionanoscience. Scientometrics 82(2), 263-287.
25. Abramo G, D'Angelo CA, Costa F (2012) Identifying interdisciplinarity through the disciplinary classi cation of coauthors of scienti c publications. J Assoc Inf Sci Technol 63(11): 2206-2222.

26. Chen S, Arsenault C, Gingras Y, Lariviere V (2015) Exploring the interdisciplinary evolution of a discipline: The case of Biochemistry and Molecular Biology. Scientometrics 102(2), 1307-1323.
27. Bordons M, Zulueta MA, Romero F, Barrigon S (1999) Measuring interdisciplinary collaboration within a university: The e ects of the multidisciplinary research programme. Scientometrics 46(3), 383-398.
28. Leydesdor L, Goldstone RL (2014) Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal Cognitive Science. J Assoc Inf Sci Technol 65(1), 164-177.
29. Zhang L, Rousseau R, Gl•anzel W (2015) Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject elds into account. J Assoc Inf Sci Technol 67(5), 1257-1265.
30. Lungeanu A, Huang Y, Contractor NS (2014) Understanding the assembly of interdis-ciplinary teams and its impact on performance. J Informetr 8(1), 59-70.
31. Lariviere V, Gingras, Y (2010) On the relationship between interdisciplinarity and sci-enti c impact. J Assoc Inf Sci Technol 61(1), 126-131.
32. Lariviere V, Haustein S, B•orner K (2015) Long-distance interdisciplinarity leads to higher scienti c impact. Plos One 10(3), e0122565.
33. Rinia EJ, van Leeuwen TN, van Raan AFJ (2002) Impact measures of interdisciplinary research in physics. Scientometrics 53(2), 241-248.
34. Wan J, Thijs B, Gl•anzel W (2015) Interdisciplinarity and impact: Distinct e ects of variety, balance, and disparity. Plos One 10(5), e0127298.
35. Levitt JM, Thelwall M (2009) The most highly cited library and information science articles: interdisciplinarity, rst authors and citation patterns. Scientometrics 78(1), 45-67.
36. Levitt JM, Thelwall M (2008) Is multidisciplinary research more highly cited?: A macrolevel study. J Assoc Inf Sci Technol 59(12), 1973-1984.
37. Chen S, Arsenault C, Lariviere V (2015) Are top-cited papers more interdisciplinary? J Informetr 9(4): 1034-1046.
38. Stirling A (2007) A general framework for analyzing diversity in science, technology and society. J Roy Soc Interf 4(5), 707-719.
39. Leydesdor L (2007) Betweenness centrality as an indicator of the interdisciplinarity of scienti c journals. J Assoc Inf Sci Technol 58(9), 1303-1319.
40. Van den Besselaar P, Heimeriks G (2001, July). Disciplinary, multidisciplinary, inter-disciplinary: Concepts and indicators. In ISSI (pp. 705-716).
41. Kagan J. The three cultures: Natural sciences, social sciences, and the humanities in the 21st century. Cambridge University Press, 2009.
42. Xie Z, Duan XJ, Ouyang ZZ, Zhang PY (2015) Quantitative analysis of the interdisci-plinarity of applied mathematics. Plos One 10(9): e0137424.
43. Milojevic S. (2013). Accuracy of simple, initials-based methods for author name disam-biguation. J Informetr 7(4): 767-773.
44. Kim J, Diesner J (2016) Distortive e ects of initial-based name disambiguation on measurements of large-scale coauthorship networks. J Assoc Inf Sci Technol 67(6):1446-1461.
45. Milojevic S (2010) Modes of Collaboration in Modern Science: Beyond Power Laws and Preferential Attachment. J Assoc Inf Sci Technol 61(7): 1410-1423.
46. Xie Z, Ouyang ZZ, Li JP, Dong EM, Yi DY (2018) Modelling transition phenomena of scienti c coauthorship networks. J Assoc Inf Sci Technol 69(2): 305-317.
47. Consul PC, Jain GC (1973) A generalization of the Poisson distribution. Technometrics 15(4): 791-799.
48. Xie Z, Xie ZL, Li M, Li JP, Yi DY (2017) Modeling the coevolution between citations and coauthorship of scienti c papers. Scientometrics 112, 483-507.
49. Levitt JM, Thelwall M, Oppenheim C (2011) Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Assoc Inf Sci Technol 62(6), 1118-1129.
50. Hey T, Tansley S, Tolle KM (2009) The fourth paradigm: data-intensive scienti c dis-covery, Microsoft research, Redmond, Washington.
51. Haythornthwaite C (2006). Learning and knowledge networks in interdisciplinary col-laborations. J Assoc Inf Sci Technol 57(8), 1079-1092.

52. Grauwin S, Beslon G, Eric Fleury, Franceschelli S, Robardet C, Rouquier JB, Jensen P (2012) Complex systems science: dreams of universality, interdisciplinarity reality. J Assoc Inf Sci Technol 63(7), 1327-1338.

53. Brier S (2013) Cybersemiotics: a new foundation for transdisciplinary theory of infor-mation, cognition, meaningful communication and the interaction between nature and culture. Integr Rev 9: 222-263.

5 Appendix

The following boundary detection algorithms come from Reference [46].

Table 6 A boundary detection algorithm for probability density functions.

Input: Observations D_s (s = 1; :::; n), rescaling function g( ), and tting model h( ).
For k from 1 to max(D_1; :::; D_n) do:
Fit h( ) to the PDF h_0( ) of fD_s; s = 1; :::; njD_s kg by maximum-likelihood estimation;
Do KS test for two data g(h(t)) and g(h_0(t)), t = 1; :::; k
with the null hypothesis they coming from the same distribution; Break if the test rejects the null hypothesis at signi cance level 5%.
Output: The current k as the boundary point.

Table 7 Boundary point detection algorithm for general functions.

Input: Data vector h_0(s) (s = 1; :::; K), rescaling funtion g( ), and tting model h( ).
For k from 1 to K do:
Fit h( ) to h_0(s), s = 1; :::; k by regression;
Do KS test for two data vectors g(h(s)) and g(h_0(s)), s = 1; :::; k with the null hypothesis they coming from the same distribution;
Break if the test rejects the null hypothesis at signi cance level 5%.
Output: The current k as the boundary point.

天津大学

本科生毕业设计（论文）中文译文

外文原文题目：Feature analysis of multidisciplinary scientific collaboration patterns based on PNAS

中文译文题目：基于 PNAS 的多学科科学合作模式的特色分析

毕业设计（论文）题目：美国科学研究系统建模及合作模式挖掘

学　　院　管理与经济学部

专　　业　信息管理与信息系统

年　　级　2019 级

姓　　名　蒋世华

学　　号　3019209018

指导教师　王文俊 教授

摘　要

合作模式的特色通常被看作不同学科之间的差异。同时，在孵化一些交叉学科的现代科学研究也出现了不同学科合作的明显特色。基于 1999 年 ~ 2013 年的多学科杂志 PNAS 发表的 52803 篇论文数据分析了在生物、物理和社会科学间的合作特色。从这些数据中，我们发现合作模式的传递性和同配性与那些作者合作者和作者论文的相同分布一样，都满足广义泊松分布和幂律分布的混合。而且，我们发现有相当一部分作者从事交叉学科研究，而不仅仅是具有许多合作者或具有很多论文的作者。这个事实提供了一个了解多学科和交叉学科合作模式方面的一个途径。

关键词： 合作模式；交叉学科；超图；复杂网络

1、介绍

自然和社会科学提供了很多方法分别去学习、预测和解释自然现象和社会(人类行为和精神状况)[i]，这些专业化的科学知识形成了各种学科。同时，为了解决那些解决方案超出单个学科范围的问题，研究人员需要整合多个学科的数据、技术、概念和理论[iiiiiivv]。学科之间的相互作用孕育了很多的跨学科，模糊了自然科学和社会科学的边界，并产生了许多重要的科学突破[viviiviii]。

研究跨学科或学科间的合作模式有助于理解合作行为和知识融合方式的多样性。因为自然科学和社会科学主要依靠论文，多学科期刊的论文为这项研究提供了信息和可靠的平台[ixxxixii]。这里我们调查了 1999-2013 年发表在《美国国家科学院院刊》（PNAS）上的 52803 篇论文。数据集的内容涵盖三个科学类别：社会科学和自然科学中的两个主要子科学，即生物科学和物理科学。

合作关系可以被表示为合作网络。因此可以从网络的角度来研究合作模式。来自不同科学领域的合作网络出现了特定的相似性，例如合作者的部分传递性、合作者数量的同质性、每个作者的合作者分布的右偏分布[xiiixivxvxvixviixviiixix]。这些共性也出现在三个作者的合作网络中（分别来自 PNAS 的三个科学类别）。我们深入了解这些共性的规律和原因。我们发现每个作者的合作者分布和每个作者的论文分布遵循相同的分布类型：广义泊松分布和幂律的混合。我们通过作者吸引合作的能力的多样性为分布类型和这些共性提供了可能的解释。

之前的一系列工作讨论了科学[xxxxixxii]、学科[xxiiixxivxxvxxvi]、大学[xxvii]、期刊[xxviiixxix]和研究团队[xxx]的跨学科定量指标。一些工作探讨了跨学科性和科学影响之间的相关性[xxxixxxiixxxiiixxxiv]（例如，引文捕获能力[xxxvxxxvixxxvii]）。基于这些参考文献的特定总体思路，我们通过交叉学科的论文共现，以及基于复现的一些指标计算如 Rao-Sterling 多样性[xxxviii]，和介数中心性[xxxix]，研究 PNAS 的交叉学科活动。

我们进一步通过多学科研究了跨学科的合作模式，并发现相当一部分物理和社会科学的作者和论文参与了跨学科研究。从巨大组成部分的数据中提取的多学科合作者关系网络分别包含了超过 88%的生物、超过 80%的物理和超过 71%的社会科学的作者。相当多的作者对巨大成分的形成做出了贡献。作者活动和生产力对巨大组成部分的贡献随着时间的推移而增加。案例研究所显示的高度跨学科性可能不能代表一般的合作模式，因为作者可以向多学科期刊提交比特定领域期刊更多的跨学科成果。

本论文的结构如下：在第 2 节中描述了数据处理过程；在第 3 节分析了相似性和相互作用；并在第 4 节中得出结论。

2、数据集

2.1 使用该数据集的原因

案例研究涉及两个概念，即多学科性（来自不同学科的研究人员在其学科内进行研究）和跨学科性（超越学科界限的研究）[xl]。多学科可以被看作是多学科交叉地出现在其中的学科的结合。一本涵盖自然科学和社会科学的多学科期刊可以用来分析科学类别之间的相互作用。这本期刊也可以用来比较多学科的合作模式并找到相似之处。PNAS 发表了高质量的研究论

文，并且提供这些论文的可靠的学科信息。该期刊还为分析全球合作模式提供了一个高质量的数据平台，因为其近一半的论文来自美国以外的作者。

包括 SCIENCE、Nature 和 Nature Communications 不提供论文的学科信息。《英国皇家学会学报》专注于物理科学和生命科学之间的交叉学科研究，但不涉及社会科学。我们的分析仅限于 PNAS，这给我们的发现带来了局限性。例如，社会科学媒体不仅依赖论文，还依赖书籍[xixii]。因此，所获得的结果必须仔细解释为在所选期刊上发表论文的研究人员的模式。然而，由于 PNAS 的影响力和代表性，案例研究可能有助于理解多学科和跨学科合作模式的各个方面。

2.2 学科信息

数据集的大多数论文被分为三个一级学科（生物、物理和社会科学）和 39 个二级学科（表 1）。跨学科论文被分为几个学科。数据包含 43304 篇生物学论文（包括 3957 篇生物物理学论文），占总数的 82.01%。该数据还包含 5987 篇物理论文和 1310 篇社会论文。二级学科以上的跨学科论文 2961 篇，占总数的 5.61%，交叉学科的重要差异并不意味更特别受到 PNAS 的偏好。

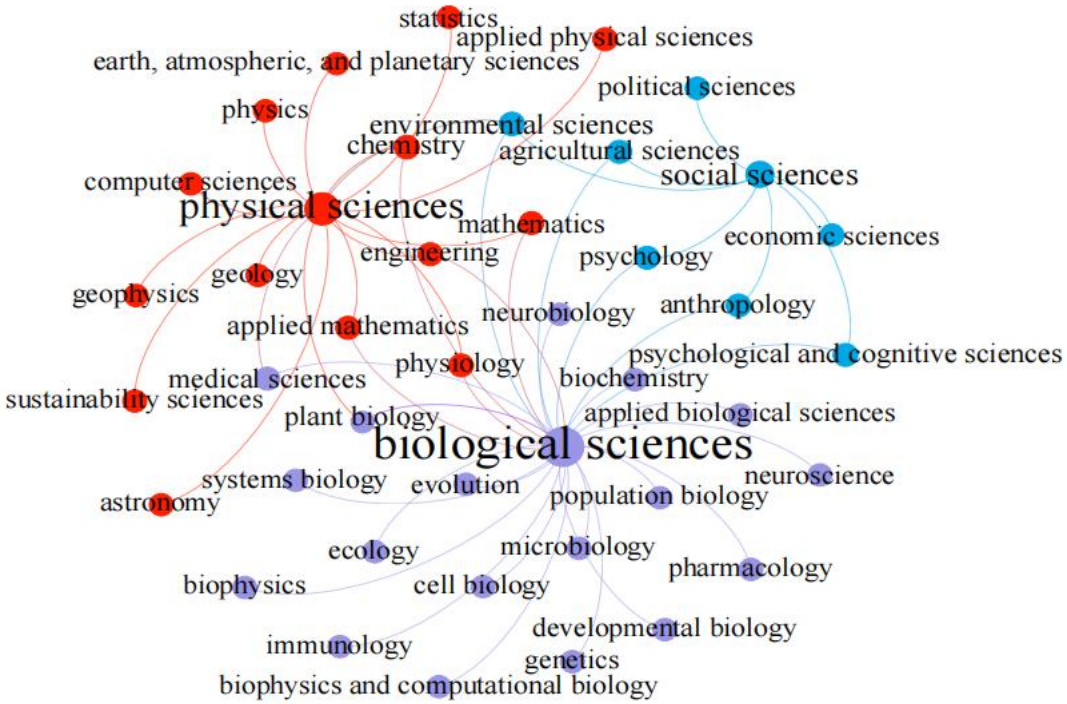

图 1 一级学科和二级学科的关系这个网络基于 PNAS 1999～2013 年的多学科论文的信息构建，如果两个学科是同一篇论文的一级学科和二级学科它们就会相连。节点大小代表了节点的度。

事实上，参与自然科学（尤其是生物科学）的研究人员数量远远超过参与社会科学的研究人员[xli]。仅有 1842 篇论文被归类为一级学科。对于这些论文，它们的二级学科被认为是缺失的，但在我们以前的工作中，它们被认为与一级学科相同[xlii]。因此表 1 中的数据与参考文献[xliii]中的数据不同。

基于论文的学科信息，我们构建了一个网络来表达一级学科和二级学科之间的关系（图 1），其中如果两个学科是同一篇论文的一级学科和二级学科，它们之间相互连接。我们还可以构建一个网络来表达二级学科之间的相互作用（图 2），其中每个节点都是一个学科，如果有一篇论文同时属于它们，则两个节点是连接的。这些网络可能会随着新发表论文的形成而发展。因此，使用最新数据，人们可能会有一个更全面的观点。

2.3 合作者网络

识别真实的作者，即消除作者姓名的歧义，是一个重要、耗时但必要的合作网络分析。

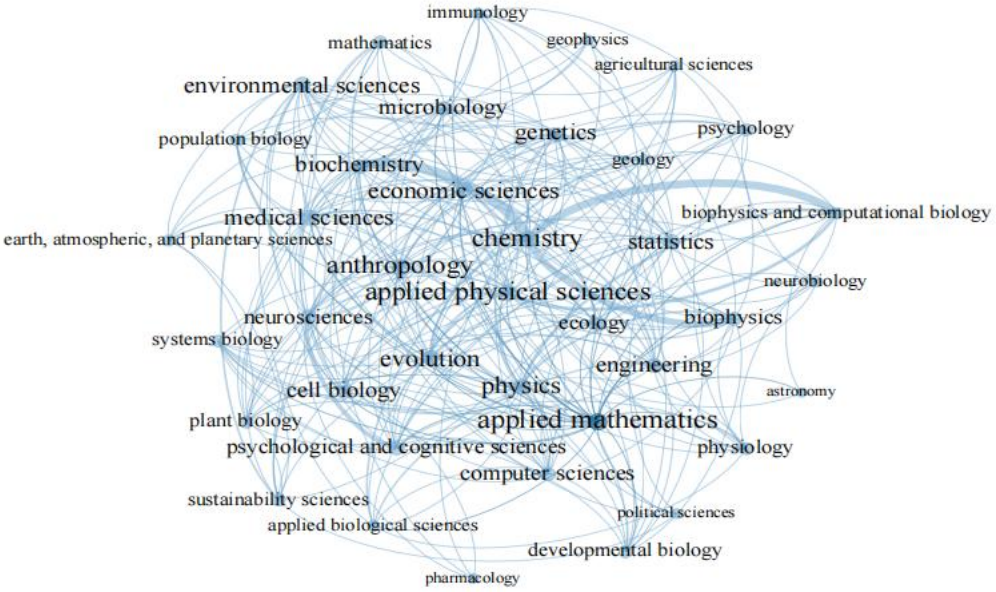

图 2 二级学科间的交互作用 带有权重的网络基于 PNAS 1999～2013 年交叉学科的论文信息构建，边的宽度表示了边的权重：在两个连接的学科间的交叉学科论文数量

有几种方法使用论文中提供的名称信息（例如，基于初始值的方法[xliii]）。基于初始值的方法的主要错误识别是由两个或多个不同的作者合并为一个引起的。因此，它减少了唯一作者的数量，并扩大了巨大组成成分的大小。获得更多的信息（如电子邮件地址）有助于减少合并错误，但也带来了收集信息的困难。

在 PNAS 1999-2013 中，93.1%的作者提供了全部的姓氏。因此，论文上提供的名字被直接用来识别作者。然而，使用姓和第一个给定名称的首字母将产生许多名称消歧的合并错误[xliv]。这些作者在数据中的比例为 2.9%，这部分作者以发表一篇以上论文为条件的比例为 0.3%。同时，如果一些作者提供了完全相同的名字，即使使用全名也会产生合并错误。中文名字被发现是名字重复的原因[xliv]。我们计算了在 100 个主要中文姓氏中，名字小于 6 个字符和姓氏的比例。这些作者在数据中的比例为 2.7%，而在这些作者中以发表一篇以上论文为条件的比例为 1.1%。这四个比例的值表明，名字重复的影响是有限的。表 2 中列出了特定子数据的这些比例。

如果作者没有全程提供自己的名字，这里采用的方法将把一位作者分成两位或两位以上。这样会分割巨大组成部分的大小，这些指数被用作跨学科研究普遍性的证据。因此 3.5、3.6 小节的结论是有争议的。

表 1 在 PNAS 1999～2013 年二级学科的特定标签

Disciplinary	m	n	k_1	k_2	b
Agricultural science	22	226	9	20	3.19
Anthropology	114	556	24	110	40.02
Applied biological science	135	767	9	134	1.79
Applied mathematics	191	380	27	182	49.39
Applied physical science	309	816	26	299	29.14
Astronomy	3	50	3	3	0.13
Biochemistry	333	6,303	19	327	16.96
Biophysics	359	3,957	16	359	7.91
Biophysics and computational biology	468	1,532	11	467	7.95
Cell biology	135	3,717	18	130	12.71
Chemistry	1,003	8,645	26	1,003	49.73
Computer science	77	101	17	70	9.50
Developmental biology	33	1,525	12	30	1.66
Earth, atmospheric, and planetary sciences	78	243	9	77	1.58
Ecology	162	1,084	15	162	10.00
Economic science	94	171	21	94	20.88
Engineering	217	392	19	217	13.85
Environmental science	184	695	20	183	25.44
Evolution	233	2,274	22	216	25.81
Genetics	103	2,664	20	97	12.68
Geology	137	285	10	136	2.79
Geophysics	23	175	7	23	1.51
Immunology	43	3,070	10	38	1.45
Mathematics	18	561	11	17	3.36
Medical science	181	4784	20	170	14.01
Microbiology	92	2,812	17	89	11.85
Neurobiology	16	1,003	9	16	0.87
Neuroscience	290	4,398	16	280	12.00
Pharmacology	26	594	4	26	0.08
Physics	229	4,818	22	227	18.24
Physiology	33	1,116	12	32	5.82
Plant biology	27	1,700	12	27	4.62
Political science	7	17	5	7	0.54
Population biology	27	166	11	26	4.04
Psychological and cognitive science	160	487	16	159	5.09
Psychology	83	449	12	83	3.62
Statistics	90	146	20	85	19.34
Sustainability science	123	399	11	120	7.66
Systems biology	36	159	11	36	1.80

表中 n 为论文的数量，m 为交叉学科的数量，k_1 为度，k_2 为度的权重，b 为在图 2 的带权重的网络中计算的介数中心性

此外，所采用的方法造成的不准确并没有改变每个作者合作者的基本事实分布类型和每个作者的论文分布类型[xliv]。

表 2 分析网络的特定统计标签

Data	a	b	c	d
PNAS 1999-2013	2.9%	0.3%	2.7%	1.1%
Biological sciences	2.7%	0.2%	2.7%	1.1%
Physical sciences	4.8%	0.4%	4.4%	0.9%
Social sciences	2.3%	0.1%	2.2%	0.3%
Biophysics	4.1%	0.3%	4.0%	1.0%
Interdiscipline	2.6%	0.1%	3.6%	0.6%

标签 a 和 b 分别表示了提供首字母和姓氏的比例、这些作者中以发表一篇以上论文为条件的比例。标签 c 和 d 分别表示了作者姓氏在 100 个主要的中国姓氏并且小于 6 个字符的比例、这些作者中以发表一篇以上论文为条件的比例。

3 数据分析

3.1 网络性质

合作是一种多元关系，$n \in Z+$，因此它可以用超图来表示，超图是图的推广，其中一条边（称为超边）可以连接任意数量的节点。将作者表示为节点，将每篇论文的作者组（论文团队）表示为超边。然后，我们可以从超图中提取一个合作者关系网络作为一个简单图，其中在每个超边的每两个节点之间形成边，并且将多个边作为一个边处理。节点的“度”和“超度”分别用来表示合作者的数量和作者的论文数量。

数据显示，生物科学（6.624）和物理科学（5.254）的平均论文团队规模大于社会科学的平均团队规模（4.634）。这种规模关系符合研究团队规模为通常在自然科学中较大，在社会科学中较小 [xli]。现在让我们考虑特定学科或科学类别中被考虑的论文的合作网络。所有这些网络都是高度集群的、具有度相似性的，其平均最短路径长度以节点数的对数表示（表 3 中的 $\log NN \approx AP$）。这些性质并不意味着所有的网络都是小世界。社会科学网络是个例外，它甚至没有超过 10%的作者。然而，这并不意味着社会科学的研究是孤立进行的。事实上，71.5%的社会科学作者属于由整个数据生成的作者网络的巨大组成部分。因此，分析作者在单个学科中的合作是有局限性的。因此，我们在所有学科的环境中进行分析。

表 3 分析网络的特定统计标签

Network	NN	NE	GCC	AC	AP	PG
PNAS 1999-2013	202,664	1,225,176	0.881	0.230	6.422	0.868
Biological sciences	184,872	1,150,362	0.881	0.232	6.364	0.880
Physical sciences	24,766	101,166	0.933	0.452	10.89	0.455
Social sciences	5,121	18,786	0.946	0.683	6.574	0.087
Biophysics	13,480	48,012	0.905	0.177	7.665	0.636
Interdiscipline	13,680	53,588	0.951	0.558	9.397	0.093

标签分别为节点的数量（NN），边的数量（NE），全局聚类系数（GCC），度相似性（AC），平均最短路径（AP），巨大组成部分的节点比例（PG）。前两个网络的平均最短路径大约计算了 400,000 对节点的样本。

3.2 度和超度

在数据上统计度和超度（不限于单个科学类别），并观察三个作者集（分别来自三个科学类别）的度分布和超度分布。我们发现尽管合作级别因科

学类别而异，但所有分布都出现了钩头、肥尾和它们之间的交叉，这可以被视为合作网络的一个共同特征（图 3）。头部和尾部可以分别通过对数正态分布和幂律分布来拟合[xlv]。

这些分布也可以作为一个整体，通过广义泊松分布和幂定律分布的混合来拟合。拟合参数如表 4 所示。我们进行了两个样本的 Kolmogorov-Smirnov（KS）检验，以比较两个数据向量的分布：节点索引（即度，超度），从相应的拟合分布中提取的样本。零假设是两个数据向量来自相同的分布。每个拟合的 p 值表明，在 5%的显著性水平上，检验不能拒绝零假设。注意到由于大量作者数量较少，$\chi 2$ 拟合度检验在这里不适用。

把作者当作样本，混合分布意味着这些样本来自不同的群体，即合作者和论文较少的作者的合作模式与合作者和论文较多的作者的合作模式不同。在参考文献[xlvi]中，对出现的混合型经验度分布给出了一种可能的解释（不含学科）。对于相同的一般思想，可以对超度分布采用类似的解释，如下所示。

研究人员是否合作发表论文可以被视为“是/否”的决定。因此，研究者的超度等同于希望与该研究人员合作的候选人在一系列决定中成功的次数。假设这些候选人的人数是 n。假设每个候选人的合作概率为 p。那么超度数将遵循二项式分布 B(n，p)，并且泊松分布将满足 np 的期望值。由于作者吸引合作者的能力的多样性，不同作者的期望值 np 各不相同。

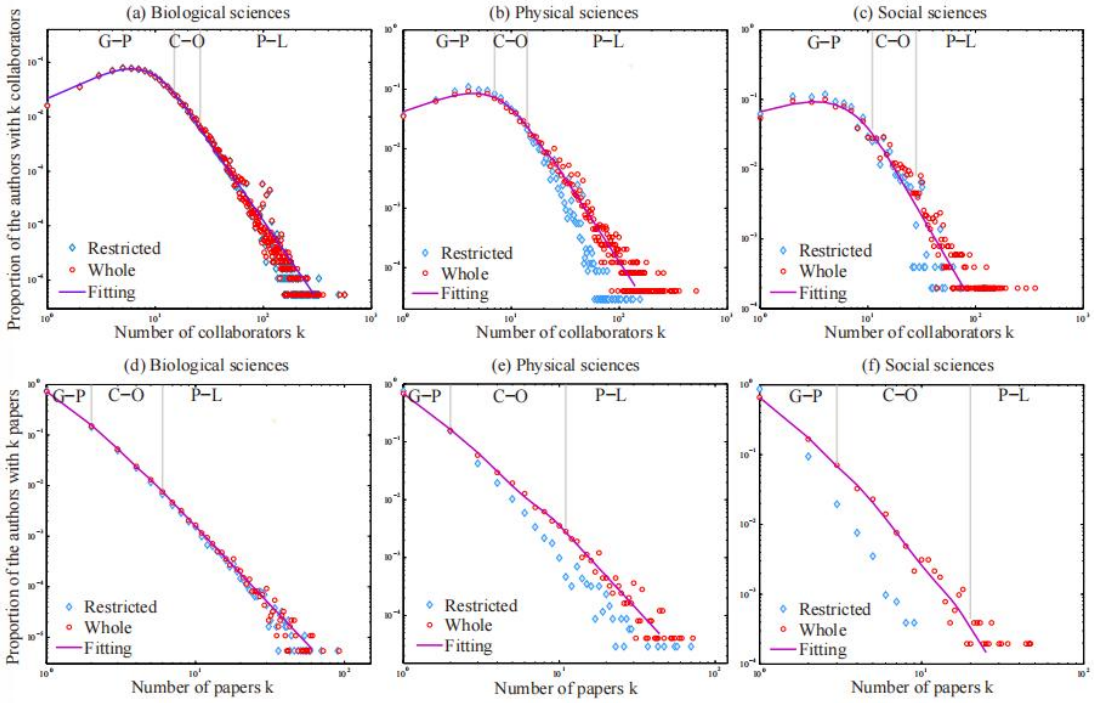

图 3 每个作者的合作者/论文分布 页面展示了 PNAS 1999～2013 年的分布（红色部分），和每个学科分类的论文（蓝色部分）。拟合分布是广义泊松分布和幂律分布的混合，拟合参数在表 4“G-P”、“C-O”、“P-L”分布表示了广义泊松分布、混合和幂律分布。

作者的决定可能具有依赖性。例如，与有出版经验的研究人员合作有助于发表论文。因此我们可以将超度视为遵循广义泊松分布的随机变量（允许事件的发生概率具有记忆性[xlvii]）。在经验数据中，大多数超度都围绕着

它们的模式。因此，我们可以认为它们遵循一些广义泊松分布，在它们的模式周围有一个期望值，因此形成了超度分布的广义泊松部分。一些作者经历了一个复杂的论文过程，这使得超度分布向右倾斜并形成一个肥尾。

3.3 合作网络的传递性

社会中的传递性是“我朋友的朋友也是我的朋友”，这是社会关系网络的典型特征。在学术界中，作者的合作者很可能彼此熟悉，因此也是合作者。例如，组织和机构背景可能会导致合作者关系的传递性，从而有助于出现合作者网络的集群结构。

网络的传递性可以用图论中的两个指标来量化，即全局聚类系数(在“三角形”中连成三元组的部分）和局部聚类系数（节点的两个邻居连接的可能性）。高传递性是合作网络的一个共同特征[xv]。

但是传递性在多大程度上是由于作者在学术社会中的活动造成的呢？这种活动可以通过合作者的数量，即度来部分地反映。因此，可以通过度和局部聚类系数之间的相关系数来描绘传递性。注意到相关系数表示两个变量或它们的秩之间的线性关系的程度。除非对于每个期望条件的 Y 给出 X 来表示 E(Y|X)作为 X 中的线性或近似线性函数，不然变量 X 和 Y 的系数通常不能完全描述相关性。对于给定的度的局部聚类系数的条件期望值是用 CC（k）表示度为 k 的节点的平均局部聚类系数。图 4 所示的 CC（k）的近似线性趋势保证了表 5 中相关性分析的有效性。减少的趋势不能从度的信息中得出。一个节点的局部聚类系数的分母随着其度的二次增长，但分子不能根据度的信息计算得到。

CC（k）均值活动的下降趋势是否会抑制传递性？肯定的答案是违背常识的。在 PNAS 1999-2013 中，74.62%的作者在数据中只发表了一篇论文，99.9%的论文团队规模遵循广义泊松部分，即每个团队规模的平均论文约为 6.028。广义泊松部分的边界由参考文献[xlvi]中概率密度函数的边界点检测算法检测（见附录）。因此大部分小度的作者的局部聚类系数接近于 1（图 4）。一些作者经历了一段时间的合作，他们的度是通过计算论文得到。对于这些作者来说，他们在不同论文中的合作者无法合作，这降低了他们的局部聚类系数。因此令人困惑的事情和常识并不矛盾，而是由于通过计算静态网络中的“三角形”来测量传递性这一动态性质的不足。

表 5 度与传递性/聚类系数指标的相关性

Discipline	Indicator	Mean	Std	SCC	PCC
Biological sciences	LCC	0.860		-0.398	-0.401
	LTC	0.001	0.005	0.275	0.077
	DN	21.09		0.543	0.400
	HN	3.015	15.47	0.070	-0.046
Physical sciences	LCC	0.806		-0.336	-0.382
	LTC	0.001	0.005	0.306	0.074
	DN	15.48		0.625	0.346
	HN	2.682	12.44	0.169	0.015
Social sciences	LCC	0.784		-0.177	-0.263
	LTC	0.001	0.006	0.292	0.050
	DN	12.87		0.723	0.482
	HN	2.268	10.89	0.175	0.030

图中指标为局部聚类系数（LCC），合作者的局部传递性（LTC），节点邻居的平均度（DN），节点邻居的平均超度（HN）。我们计算了这些作者的平均指标，这些指标和度的斯皮尔曼等级相关系数（SCC）和皮尔森积矩相关系数（PCC）。对于两个较小的皮尔森积矩相关系数（PCC），我们计算了它们的标准差。

为了设计一个更合理的测量传递性的指标，从最开始对合作的最初定义来考虑：一次研究合作中的两个合作者（他们还没有合作过）在未来合作的概率。这个概率可以通过合作的动态超图的时间信息来计算。基于作者的平均概率测量全局传递性，其值在每个科学类别中都很低（表 5）。注意到 PNAS 1999-2013 中的计算是有限的，在其他期刊或其他时间段可能会发生变化。因此，这里传递性的值可能会被低估。度为 k 的作者的传递性概率（图 4 中的 TC（k））的增加趋势意味着合作有助于传递性。这符合常识：一个拥有许多合作者的研究人员很可能会介绍他的合作者进行合作。

3.4 合作关系的同质性

合作是基于特定研究人员的共同特征，包括兴趣和地理位置。同质性现象出现在许多社交关系中，在网络科学中被称为同质混合[xviii]。每种学科的作者更倾向于与社会活动和生产力相似的合作者合作吗？作者的社会活动和生产力可以分别用度和超度这两个指标来量化，然后通过作者的指标和作者邻居的平均指标这两个变量之间的相关系数来描述指标的偏好。正相关意味着同质，负相关意味着不同质，零则无偏好。

度的同质性是合作网络的一个特征[xviii]。这是否意味着善于交际的研究人员（与许多合作者）会更倾向于与其他善于交际的研究员合作，不善于社交的与不善于社交的合作呢？在之前的一项研究[xlviii]中，我们发现排名前 5.99%的大多数善于社交的作者（根据度来衡量）与另一位此类作者合作的比例为 99.5%。这一比例甚至可能被低估，因为这些作者可能在 1999 年之前或其他情况下合作。注意到名称消歧方法的拆分和合并错误在一定程度上影响了比例。即便如此，这一比例仍然很高。

然而，如果善于社交的研究人员只与善于社交的研究人员合作，那么就会有许多善于社交的研究人员，这违背了经验度分布。现在让我们分析作者的社会活动对度的同质性的影响。对于具有 k 度的作者，用 DN（k）表示其邻居的平均度。每个经验数据集的 DN（k）都存在趋势的变化：头部有明显的增加趋势，而尾部则没有（图 4）。这意味着度同质性主要由小度的作者控制。

DN（k）的趋势的临界点由参考文献 [xlvi]（附录中列出）中常用函数的边界点检测算法检测得出。该算法的输入是 DN（k）、$g(\cdot)=\log(\cdot)$ 和 $h(x)=a_1x^3+a_2x^2+a_3x+a_4(x,a_i\in R,i=1,...,4)$。使用这些输入是基于对 DN（k）的观察。大多数作者的度都在他们的众数 5 左右，只有少数作者拥有大度。因此，作者的邻居很可能是小度的作者。因此，对于小度作者来说，这些作者与其邻居之间的度差距很小，而对于大度作者来说则很大，这导致了 DN（k）的趋势变化。

在每个科学类别中，超度和邻居的平均超度之间的相关系数约为零（表 5）。对于具有 k 超度的作者，用 HN（k）表示其邻居的平均超度。这意味着选择合作者不受生产力因素的影响。事实上，研究团队的成员可能有各种各样的科学年龄（新加入者、现任者），因此有不同的超度。由于合作主要发生在一个研究团队中，一个作者的合作者可能具有不同的超度，这导致了 HN（k）的稳定趋势。

基于 HN（k）的平均值大于 2，以及 74.62%的作者在数据中仅有一篇论文，我们可以得出，在数据中很大一部分作者与至少一位在 PNAS 1999-2003 发表过第一篇论文的作者合作。这些作者在生物学、物理学、社会科学上的比例分别为 79.22%、71.17%和 65.12%。这一比例可能被高估了，因为其中一些作者可能在 1999 年之前在 PNAS 上发表论文。

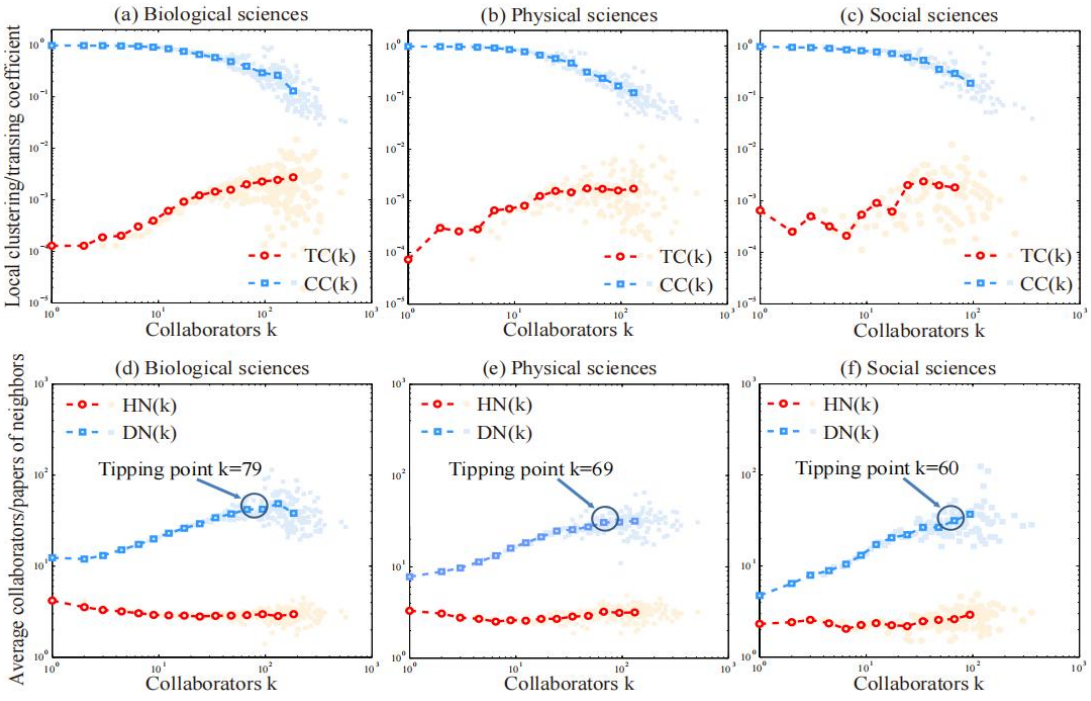

图 4 给定特定指标的条件期望值 从度为 1 到最大值，我们计算了在 k 度的节点上的平均局部聚类系数（CC（k））、合作者的局部传递性（TC（k））、节点邻居的平均度（DN（k））、节点邻居的平均超度（HN（k））。数据按横坐标归类来提取藏在噪声中的趋势。

3.5 在学科层面的跨学科性

同类别比例衡量跨学科研究的活动。发表跨学科论文的社会、物理和生物科学作者分别占 49.2%、46.0%和 7.3%。常识表明，社会科学家单独从事科学研究。社会科学的比例表明常识在 PNAS 中并不成立。参考文献[xlix]还表明，近几十年来，社会科学的跨学科性有所增加。

上述分析过程可以应用于二级学科，以获得更准确的结果。然而，有些学科只有少量的论文，例如只有 17 篇政治学论文。因此，对这些学科的分析失去了统计学意义。因此，我们从另一个角度分析了二级学科之间的相互作用，将它们可视化为图 2 中的网络。网络是连接的，也就是说没有任何学科是孤立的。该网络的度和介数中心性前三的节点是应用数学、化学和人类学（表 1）。它意味着这些学科的理论、方法和问题直接或间接地使用或研究在许多学科中。对于每个一类学科，我们将其二类学科收缩为一个节点，并计算收缩节点的介数中心性。它们的介数中心性（生物科学 47.51，物理科学 163.81，社会科学 161.72）支持上述的分析。

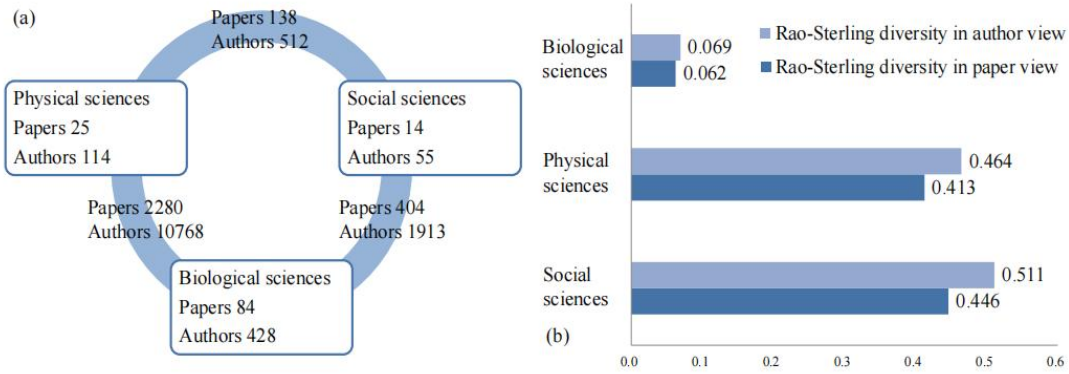

图 5 PNAS 1999～2013 年在生物、物理、社会科学的交叉学科研究 面板 a 中展示了这三个科学学科的论文和这些作者参与动态交互的部分，并且这些参与交互的部分不是单学科。面板 b 中展示了每个科学学科的作者/论文视角的 Rao-Sterling 多样性，其测量了交叉学科研究的学科多样性。

同类别比例仅描述跨学科活动。现在，让我们通过 Rao-Sterling 指数[xxxviii] $\Delta=\sum_{i,j(i\neq j)} d_{ij}^{\alpha}(p_i p_j)^{\beta}$ 来衡量每个科学类别中跨学科研究的学科多样性，其中，p_i 和 p_j 是论文/作者在 i 和 j 科学类别中的比例，d_{ij} 是在 i 和 j 类别的不同贡献水平。学科信息用于将作者分为科学类别：如果他的一篇论文属于某个学科，则作者可以被分为该学科，从而被分为相应的科学。注意到如果一位作者的论文属于多个学科，那么他可以被分为几个科学类别。这里，我们让所有 i、j 满足 $\alpha=\beta=d_{ij}=1$，因此计算的 Rao-Sterling 指标衡量了科学类别水平上跨学科研究的平衡加权变化。作者观点和论文观点的指标表明，社会科学和物理科学跨学科研究的学科多样性远高于生物科学（图 5）。

3.6 在作者层面的跨学科性

我们分析了作者度/超度与进行跨学科研究的可能性之间的关系，以及论文团队规模与成为跨学科论文的概率之间的关系。图 6 显示，在每个科学类别中，跨学科研究不仅仅是由拥有大度或拥有大超度的作者进行。

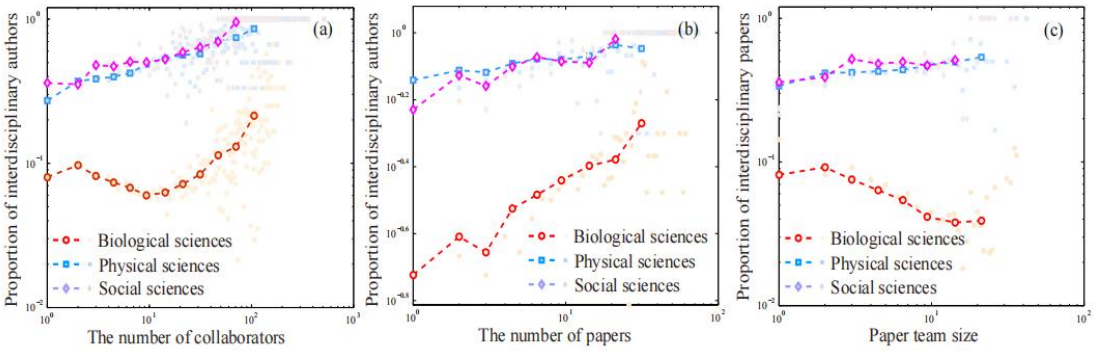

图 6 在作者和论文的特定指标以及他们的交叉学科性之间的关系 面板 a、b 展示了作者的度/超度和进行交叉学科研究的可能性的关系。面板 c 展示了论文团队规模和成为交叉学科论文的可能性的关系。

图 6 还表明，大度或超度作者很可能参与跨学科研究，团队规模大的论文很可能是跨学科的。这些现象似乎被预期是随机的。取几类的一组元素（合作者、论文），随机选择一个子集。那么，一个较大的子集更有可能包含来自多个类的元素。这种推理虽然合理，但并不正确，因为研究人员并没有随机选择主题和合作者。研究成本（时间和精力的投资）使研究人员倾向于在自己熟悉的领域内工作。此外，这个推理基于合作伙伴的选择范围仅限于经验数据，这在现实中是不成立的。

我们分析了合作网络 PNAS 1999-2013 的庞大组成部分，该网络包含超过 86.8%的作者。社会、物理和生物科学的作者分别占 71.5%、76.7%和 88.9%（图 7e）。注意到由初始基本方法引起的作者错误识别增加了巨大组成部分的大小[xliv]。因此我们通过论文中提供的作者姓名来识别作者（这可能会将一位作者分成两位），以获得保守的结果。

跨学科研究和多学科研究为包含每个科学类别的大多数作者的巨大组成部分做出了贡献。我们分析了巨大组成部分的作者比例与作者活动/生产力之间的关系。将作者从高度和超度分别移除到低度，并计算出巨大成分的比例。从删除作者的比例与巨大成分的比例之间的关系曲线可以发现，巨大成分的形成是由相当多的作者贡献的，例如按度排名的前 10%的作者（图 8）。考虑三个时期的关系，即 1999-2003 年、2004-2008 年和 2009-2013 年。随着时间的推移，关系曲线向左移动，这意味着作者的活动和生产力在这个巨大组成部分的形成中发挥着越来越重要的作用。

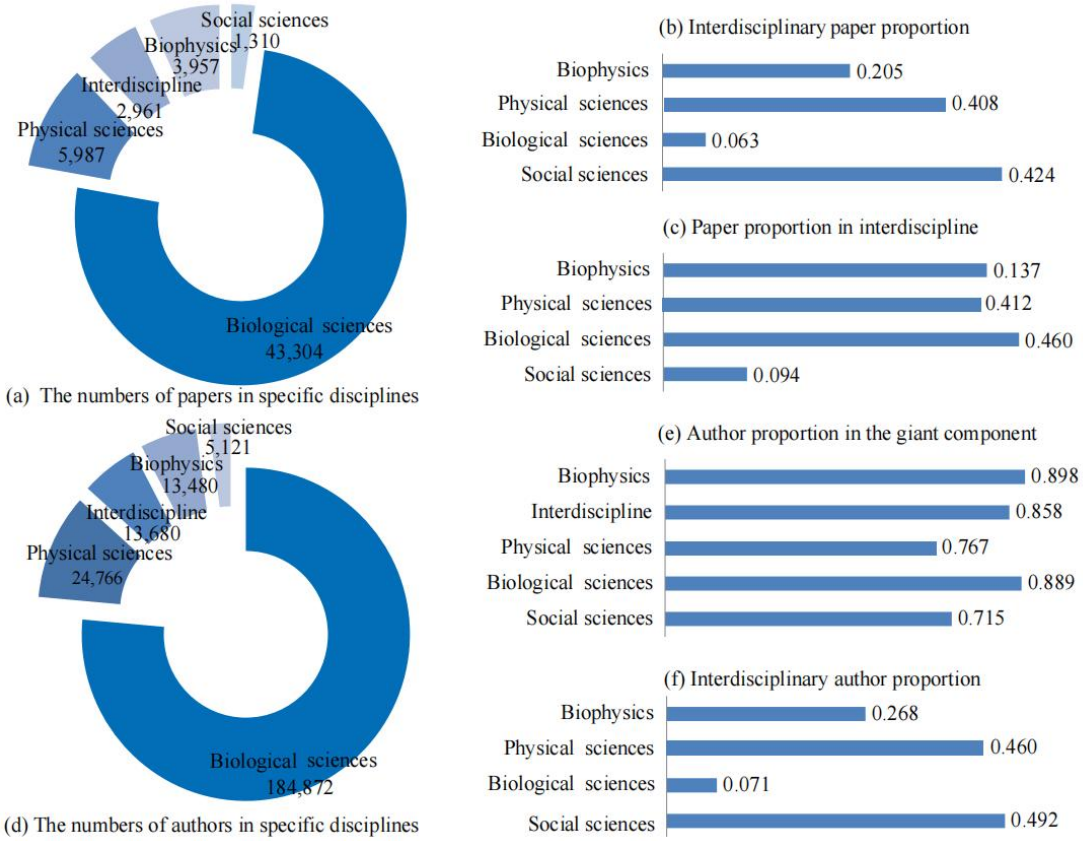

图 7 特定学科的交叉程度 对于每一个考虑到的学科 i，我们分别用集合 A_i、A_i^I、P_i、和 P_i^I 表示了它的作者、它参与交叉学科研究的作者、它的论文、它的交叉学科论文。用 S 表示 PNAS 1999～2013 合作网络的巨大组成部分。面板 a 中指标为 $|P_i|$，面板 b 中为 $|P_i \cap I_i| / |I_i|$，面板 c 中为 $|P_i \cap P_i^I| / |P_i^I|$，面板 d 中为 $|A_i|$，面板 e 中为 $|A_i \cap S| / |A_i|$，面板 f 中为 $|A_i \cap A_i^I| / |A_i|$。

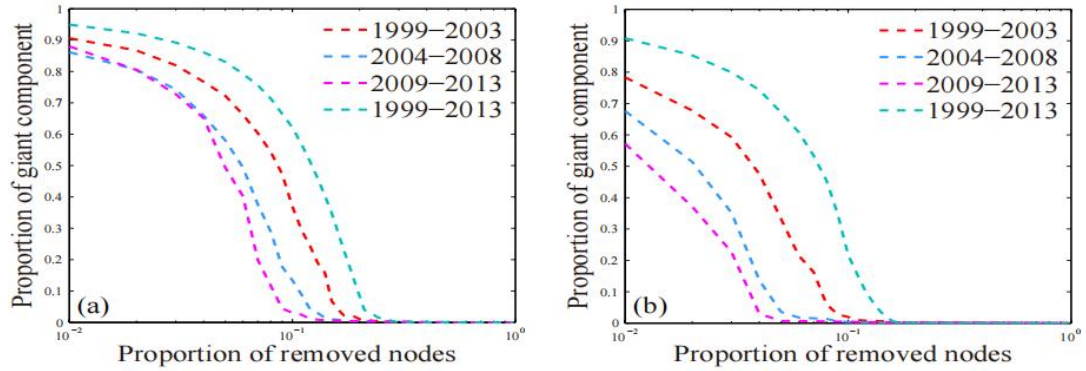

图 8 巨大组成部分和度/超度的关系 分别将节点从高度/超度移到低度/超度，对于不同的度和超度来说，移除节点的部分和巨大组成部分的关系曲线展示了大部分作者为巨大组成部分作出了贡献，在三个时间周期（1999-2003、2004-2008、2009-2013）的关系曲线左移趋向展示了作者活动和生产力对巨大组成部分的构成的贡献在上升。

4. 结论和展望

我们对 PNAS 1999-2013 的案例研究验证了生物、物理和社会科学中合作模式的传递性和同质性。数据表明，这三个科学类别的度分布类型是相同的，它们是广义泊松分布和幂律的混合。这也适用于超度。我们通过作

者的“是/否”决定以及他们吸引合作的不同能力，为这种分布类型的出现提供了解释。

数据显示，相当多的作者从事跨学科研究，合作网络 PNAS 1999-2013 的巨大组成部分包含了每个科学类别的大多数作者。我们采用网络视角分析了二级学科之间的相互作用，并通过度和介数中心性等网络指标量化了它们的跨学科性。我们发现特定的二级学科（如应用数学和人类学）在跨学科研究中发挥着重要作用。

由于 PNAS 的重要性及其论文的准确学科信息，该案例研究有助于理解多学科和跨学科的合作模式。数据的选择可能会影响我们关于跨学科研究结果的细节。我们的研究结果可能不能被解释为一般研究人员的模式。例如，我们无法期望通过分析特定领域的期刊来观察到高度的跨学科性。我们在完成案例研究时提出了一个问题：跨学科研究的基础依据是什么？虽然对这个问题的探究的讨论超出了本文的范围，但下面提供了一个简单的讨论。

在科学的发展过程中，学科有分裂的趋势：分为子学科和特定主题。尽管研究对象不同，但它们的研究范式是共同的，可以分为四类，即理论研究、实验、模拟和数据驱动[l]。同时，许多科学问题过于复杂，无法通过单一学科的方法论来理解。将来自不同学科的理论和方法论观点整合在一起，为研究问题创造了统一的方法论，甚至为特定学科中的概念提供了工具[li]，这推动了跨学科学科的形成[lii]。

系统科学作为一门典型的跨学科学科，研究系统从简单到复杂，从自然到社会科学。系统的各部分以及各部分之间的关系可以抽象为网络。网络（模型、算法等）研究的迅速发展孕育了一个新的学科，即网络科学。一些来自生物、物理和社会领域的研究人员在网络框架下调查了他们各自的问题[liii]，例如我们的案例研究。

为了跟进上述内容，人们会认为常见的研究范式和方法论，特别是那些整合为跨学科学科的研究范式，为科学类别之间的相互作用以及合作网络中巨大组成部分的形成提供了依据。分析论文内容有助于验证这些范式和方法的普遍性，这似乎很有希望。超过一半的 PNAS 1999-2013 论文包含“系统”和“控制”这两个主题词[xlii]。论文在一定程度上包含主题词的比例很高，这反映了主题的典型性。然而，这并不意味着一篇包含“系统”一词的论文和一篇应用系统科学研究成果的论文之间有很大的关系。因此，在语义层面上评估普遍性是一个有待进一步研究的课题。

参考文献

[i] Weingart P (2012) A short history of knowledge formations. In R. Frodeman, J. Thompson Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 3–14). Oxford, England: Oxford University Press.

[ii] National Academies (U.S.). Committee on Facilitating Interdisciplinary Research (2004).Facilitating interdisciplinary research. Washington: National Academy Press. Retrieved from http://www.nap.edu/books/0309094356/html/

[iii] Hurd JM (1992) Interdisciplinary research in the sciences: Implications for library organizations. Coll Res Liber 53(4), 283–297.

[iv] Cooper G (2013) A disciplinary matter: Critical sociology, academic governance and interdisciplinarity. Sociology 47(1), 74–89.

[v] Hadorn GH, Pohl C, Bammer G (2012) Solving problems through transdisciplinary research. In R. Frodeman, J. Thompson Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 431–452). Oxford, England: Oxford University Press.

[vi] Siedlok F, Hibbert P (2014) The organization of interdisciplinary research: Modes, drivers and barriers. Int J Manage Rev 16(2), 194–210.

[vii] Liu Y, Rafols I, Rousseau R (2012) A framework for knowledge integration and diffffusion.J Doc 68(1), 31–44.

[viii] Gooch D, Vasalou A, Benton L (2017) Impact in interdisciplinary and cross–sector research: Opportunities and challenges. J Assoc Inf Sci Technol 68(2), 378–391.

[ix] Larivi´ere V, Gingras Y, Archambault E (2006) Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. Scientometrics 68(3): 519–533.

[x] Moody J (2004) The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. Am Sociol Rev 69(2), 213–238.

[xi] Gl¨anzel W, Schoepflflin U (1999) A bibliometric study of reference literature in the sciences and social sciences. Inform Process Manag 35(1): 31–44.

[xii] Hicks D (1999) The diffiffifficulty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics 44(2): 193–215.

[xiii] Sarig¨ol E, Pfifitzner R, Scholtes I, Garas A, Schweitzer F (2014) Predicting scientifific success based on coauthorship networks. EPJ Data Science 2014:9.

[xiv] Barab´asi AL, Jeong H, N´eda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientifific collaborations. Physica A 311: 590–614.

[xv] Newman M (2001) The structure of scientifific collaboration networks. Proc Natl Acad Sci USA 98: 404–409.

[xvi] Newman M (2004) Coauthorship networks and patterns of scientifific collaboration. Proc Natl Acad Sci USA 101: 5200–5205.

[xvii] Xie Z, Ouyang ZZ, Li JP (2016) A geometric graph model for coauthorship networks. J Informetr 10: 299–311.

[xviii] Newman M (2002) Assortative mixing in networks. Phys Rev Lett 89: 208701.

[xix] Tomasello MV, Vaccario G, Schweitzer F (2017) Data–driven modeling of collaboration networks: A cross–domain analysis. EPJ Data Science 6: 22.

[xx] Braun T, Schubert A (2003) A quantitative view on the coming of age of interdisciplinarity in the sciences, 1980–1999. Scientometrics 58(1), 183–189.

[xxi] Levitt JM, Thelwall M, Oppenheim C (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Assoc Inf Sci Technol 62(6), 1118–1129.

[xxii] Porter AL, Roessner JD, Cohenm AS, Perreault M (2006). Interdisciplinary research: Meaning, metrics and nurture. Res Eval 15(3), 187–195.

[xxiii] Porter AL, Rafols I (2009) Is science becoming more interdisciplinary? Measuring and mapping six research fifields over time. Scientometrics 81(3), 719–745.

[xxiv] Rafols I, Meyer M (2010) Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. Scientometrics 82(2), 263–287.

[xxv] Abramo G, D'Angelo CA, Costa F (2012) Identifying interdisciplinarity through the disciplinary classifification of coauthors of scientifific publications. J Assoc Inf Sci Technol 63(11): 2206–2222.

[xxvi] Chen S, Arsenault C, Gingras Y, Larivi´ere V (2015) Exploring the interdisciplinary evolution of a discipline: The case of Biochemistry and Molecular Biology. Scientometrics 102(2), 1307–1323.

[xxvii] Bordons M, Zulueta MA, Romero F, Barrig´on S (1999) Measuring interdisciplinary collaboration within a university: The effffects of the multidisciplinary research programme.Scientometrics 46(3), 383–398.

[xxviii] Leydesdorffff L, Goldstone RL (2014) Interdisciplinarity at the journal and specialty level:The changing knowledge bases of the journal Cognitive Science. J Assoc Inf Sci Technol 65(1), 164–177.

[xxix] Zhang L, Rousseau R, Gl¨anzel W (2015) Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject fifields into account. J Assoc Inf Sci Technol 67(5), 1257–1265.

[xxx] Lungeanu A, Huang Y, Contractor NS (2014) Understanding the assembly of interdisciplinary teams and its impact on performance. J Informetr 8(1), 59–70.

[xxxi] Larivi´ere V, Gingras, Y (2010) On the relationship between interdisciplinarity and scientifific impact. J Assoc Inf Sci Technol 61(1), 126–131.

[xxxii] Larivi´ere V, Haustein S, B¨orner K (2015) Long–distance interdisciplinarity leads to higher scientifific impact. Plos One 10(3), e0122565.

[xxxiii] Rinia EJ, van Leeuwen TN, van Raan AFJ (2002) Impact measures of interdisciplinary research in physics. Scientometrics 53(2), 241–248.

[xxxiv] Wan J, Thijs B, Gl¨anzel W (2015) Interdisciplinarity and impact: Distinct effffects of variety, balance, and disparity. Plos One 10(5), e0127298.

[xxxv] Levitt JM, Thelwall M (2009) The most highly cited library and information science articles: interdisciplinarity, fifirst authors and citation patterns. Scientometrics 78(1), 45–67.

[xxxvi] Levitt JM, Thelwall M (2008) Is multidisciplinary research more highly cited?: A macrolevel study. J Assoc Inf Sci Technol 59(12), 1973–1984.

[xxxvii] Chen S, Arsenault C, Larivi´ere V (2015) Are top–cited papers more interdisciplinary?J Informetr 9(4): 1034–1046.

[xxxviii] Stirling A (2007) A general framework for analyzing diversity in science, technology and society. J Roy Soc Interf 4(5), 707–719.

[xxxix] Leydesdorffff L (2007) Betweenness centrality as an indicator of the interdisciplinarity of scientifific journals. J Assoc Inf Sci Technol 58(9), 1303–1319.

[xl] Van den Besselaar P, Heimeriks G (2001, July). Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In ISSI (pp. 705–716).

[xli] Kagan J. The three cultures: Natural sciences, social sciences, and the humanities in the 21st century. Cambridge University Press, 2009.

[xlii] Xie Z, Duan XJ, Ouyang ZZ, Zhang PY (2015) Quantitative analysis of the interdisciplinarity of applied mathematics. Plos One 10(9): e0137424.

[xliii] Milojevi´c S. (2013). Accuracy of simple, initials–based methods for author name disambiguation. J Informetr 7(4): 767–773.

[xliv] Kim J, Diesner J (2016) Distortive effffects of initial–based name disambiguation on measurements of large–scale coauthorship networks. J Assoc Inf Sci Technol 67(6):1446– 1461.

[xlv] Milojevi´c S (2010) Modes of Collaboration in Modern Science: Beyond Power Laws and Preferential Attachment. J Assoc Inf Sci Technol 61(7): 1410–1423.

[xlvi] Xie Z, Ouyang ZZ, Li JP, Dong EM, Yi DY (2018) Modelling transition phenomena of scientifific coauthorship networks. J Assoc Inf Sci Technol 69(2): 305–317.

[xlvii] Consul PC, Jain GC (1973) A generalization of the Poisson distribution. Technometrics 15(4): 791–799.

[xlviii] Xie Z, Xie ZL, Li M, Li JP, Yi DY (2017) Modeling the coevolution between citations and coauthorship of scientifific papers. Scientometrics 112, 483–507.

[xlix] Levitt JM, Thelwall M, Oppenheim C (2011) Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Assoc Inf Sci Technol 62(6), 1118–1129.

[l] Hey T, Tansley S, Tolle KM (2009) The fourth paradigm: data–intensive scientifific discovery, Microsoft research, Redmond, Washington.

[li] Haythornthwaite C (2006). Learning and knowledge networks in interdisciplinary collaborations. J Assoc Inf Sci Technol 57(8), 1079–1092.

[lii] Grauwin S, Beslon G, Eric Fleury, Franceschelli S, Robardet C, Rouquier JB, Jensen P (2012) Complex systems science: dreams of universality, interdisciplinarity reality. J Assoc Inf Sci Technol 63(7), 1327–1338.

[liii] Brier S (2013) Cybersemiotics: a new foundation for transdisciplinary theory of information, cognition, meaningful communication and the interaction between nature and culture. Integr Rev 9: 222–263.