Feature analysis of multidisciplinary scientific collaboration behaviors: A case study on PNAS

Zheng Xie · Zonglin Xie · Miao Li · Zhenzheng Ouyang · Xiaojun Duan · Jianping Li

Received: date / Accepted: date

Abstract The features of collaboration behaviors are often considered to be different from discipline to discipline. Meanwhile, collaborating among disciplines is an obvious feature emerged in modern scientific research, which incubates several interdisciplines, such as sustainability science. 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. In the aspect of similarities, the data emerge the similar transitivity and assortativity of collaboration behaviors, the identical distribution type of collaborators per author and that of papers per author. In the aspect of interactions, the data show a considerable proportion of authors engaging in interdisciplinary research, and the more collaborators and papers an author has, the more likely the author pursues interdisciplinary research. The analysis of the paper contents illustrates that the development of each science category has an equilibrium relationship in the long-run with the developments of typical research paradigms and transdisciplinary disciplines. Hence, those unified methodologies can be viewed as grounds for the interactions.

Keywords Coauthorship behavior · Interdiscipline · Hypergraph · Complex network

Z. Xie*, Z. Ouyang, J. Li
College of Science, National University of Defense Technology, Changsha, 410073, China
*E-mail: xiezhe81@nudt.edu.cn
M. Li
School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, 200240, China
The first four authors have contributed equally to this work.
1 Introduction

According to their purposes, sciences are divided into three divisions by Aristotle, namely theoretical sciences (mathematics, physics,...), practical sciences (ethics, politics,...) and productive sciences (poetry, rhetoric,...) [1]. It is the initial version for the classification of natural sciences, social sciences and humanities. Natural and social sciences provide methodical approaches to study, predict and explain the natural phenomena and sociality (human behaviors and psychological states) respectively. Humanities present a methodology to study human culture, including history, literature, music and arts with an emphasis on understanding particular individuals, events, or eras [2]. Differences among the three methodical types exist on several dimensions, ranging from research problems, the evidences on which inferences are based, the vocabularies presenting concepts and theories, etc. J. Kagan even described them as three cultures [3].

Collaboration level is one of the differences among the three cultures. In common sense, the collaboration activities of researchers in the social sciences are more comparable to those of researchers in the natural sciences than those in the humanities, which can be verified by bibliometrical analysis on papers [4,5]. In reality, the media mainly count on scientific papers for the natural sciences, but diversify for social sciences and humanities, e.g. books [6,7]. However, the collaboration level in books is usually lower than that in papers. Hence, measuring collaboration level based on papers is an informative and reliable approach [8]. Bibliometrical studies based on papers also show that the collaboration level in each of the three cultures grows steadily over years in the world, especially the level of collaborations among nations and institutions [4].

Widely and frequently interacting between disciplines is a new feature emerged in modern research [9–16]. Exploring the hidden mechanism under the complex phenomena in nature and society, researchers often need to integrate data, techniques, concepts, and theories from several disciplines to solve problems whose solutions are beyond the scope of a single discipline [17,18]. Therefore, collaborations occur not only in one discipline solitary, also across disciplines [19–21]. Consequently, interactions between disciplines and cross-sector projects incubate several interdisciplines, and fuzz the boundary of natural and social sciences [22–25]. A range of important scientific discoveries and breakthroughs are produced by the interactions, which have positive influences on academic, economic and social fields [26–32].

Quantitatively analyzing the interactions needs a fine dataset, e.g. papers of an influential journal of multidisciplinary studies. Discipline information of papers also should be provided. The Proceedings of the National Academy of Sciences (PNAS) is one such journal, the content of which spans social sciences and two principal sub-sciences in natural sciences, viz. biological and physical sciences. Moreover, the journal provides a fine data platform for analyzing the interactions at the level of global worldwide, because nearly half of its papers come from authors outside the United States. The data considered here consist of 52,803 papers published in PNAS over the years 1999-2013. Based on the
data, we analyze the similarities of collaboration behaviors in the three science categories and the interactions between each two of them.

Collaboration relationships can be expressed by graphs (called coauthorship networks), where nodes represent authors, and edges coauthor relationships [33][35]. In the language of social networks, the data show specific similarities of the collaboration behaviors in the three categories of sciences, viz. partial transitivity of coauthorship, homophily on the number of collaborators and the distribution-type of collaborators/papers per author. The type is a mixture of generalized Poisson and power-law distributions. A presumed explanation is given to show the type can be deduced from a range of “yes/no” decisions of authors.

The data show a considerable proportion of authors and papers in physical and social sciences involving interdisciplinary research. Meanwhile, the authors having more collaborators and papers are probably to carry interdisciplinary research, and to introduce their collaborators coauthor in future. In a network view, the interactions make the coauthorship network extracted from the data have a giant component and small-world property. More than 88%, 80% and 71% authors in biological, physical and social sciences belong to the giant component respectively. Note that the author misidentification caused by initial-based methods increases the size of the ground-truth giant components [36]. Hence we identify authors by their provided names on papers (which likely split one author into two) to obtain a conservative result.

Research paradigms of sciences can be classified into four categories, namely theoretical research, experiment, simulation and data-driven. Meanwhile, transdisciplinary disciplines (e. g. systems science) integrate the theoretical and methodological perspectives drawn from all disciplines to build a unified methodology. The universality of research paradigms and transdisciplinary disciplines gives grounds for the interactions and the emergence of giant components in coauthorship networks. To validate the universality and transdisciplinarity, we analyze the contents of papers, and find that in each science category, the papers containing one of the seven selected topic words of research paradigms and transdisciplinary disciplines account a considerable fraction. Moreover, the quarterly number of papers in each science category and that containing each of the selected words is cointegrated and positively correlated.

This report is structured as follows: the data processing is described in Section 2; the similarities and interactions are analyzed in Sections 3-5 respectively; the content analysis is shown in Section 6; and the conclusion is drawn in Section 7.

2 The Data

2.1 Reason for using the data

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 behaviors of multidisciplines and find similarities. PNAS publishes high quality research papers, commentaries, reviews etc. Most importantly, it provides reliable discipline information of papers.

Multidiscipline journals: Science, Nature and Nature communication do not provide discipline information of papers. Journal of the royal society interface focuses on the cross-disciplinary research at the interface between the physical and life sciences, but does not involve social sciences. Our study is restricted into one journal. The restriction brings limitation of our study. However, knowledge media of social sciences are not limited to research papers. Hence the results obtained must be carefully interpreted as being the behaviors of researchers who publish papers in the chosen journal. However, due to the influence and representability of PNAS, the case study potentially contributes to understanding aspects of multidisciplinary collaboration behaviors.

2.2 Discipline information

Almost papers of the dataset have been classified into three first-class disciplines (biological, physical, and social sciences) and 39 second-class disciplines, e. g. mathematics (Fig. 1). A small fraction of papers are only classified into the first-class disciplines. For those papers, we regard their second-class discipline to be the same as their first-class discipline. The data have 43,304 biological papers (including 3,957 papers of biophysics), which account for 77% of the total. The data also contain 5,987 physical papers and 1,310 social papers. There are 3,007 interdisciplinary papers belonging to more than one of the second-class disciplines, which count for 5.7% of the total. The significant difference of discipline proportion does not mean the preference of PNAS. In reality, the number of researchers involving natural sciences (especially, biological sciences) is far more than that of researchers involving social sciences.

2.3 Coauthorship

Analyzing coauthorship needs to identify ground-truth authors. Current methods of author disambiguation can be classified into two classes: the methods only using the information of provided names (e. g. initial based methods), and the methods requiring additional information (e. g. email address). The second-class of methods are often hard to implement, due to the difficulty of collecting additional information. The dominant misidentification of initial based methods is caused by merging two or more different authors as one. Hence, it deflates the number of unique authors, and inflates the size of the ground-truth giant component. However, it does not much affect the distribution type of collaborators per author and that of papers per author.

The result in Section 5 is based on giant components, will be inflated if using initial based methods. Hence we identify authors by their provided
names on papers, the effectiveness analysis of which is given in Appendix. The splitting risk (treat the author using different names as different authors) of the adopted method is higher than those of initial based methods. Hence the result obtained here is conservative. and count the quarterly numbers of the papers (for each of the three science categories) containing specific words, which are listed in Appendix.

2.4 Topic words

The contents of papers are analyzed to explore the deep reasons for interactions between disciplines, and those for the formation of giant components in coauthorship networks. The package of Python: Natural Language Toolkit (NLTK, www.nltk.org) is used to extract specific words (nouns and the words whose synsets contain nouns) from the data. The words are called topic words, if they can express topics. The high proportion of the papers containing a topic word at certain levels reflects the typicality of the topic.

Over half the papers contain the topic words “model”, “experiment”, “data”, “system” and “control”, which inspires us to think the research paradigms in common and transdisciplinary sciences (e.g. systems science) reflected by those words can be regarded as the grounds for the interactions. To confirm the universality and transdisciplinarity, we use some statistical technologies (e.g. cointegration test of time series) to analyze the relationships between the de-
velopmental trend of each science category and that of each considered research paradigm and transdiscipline. Hence, we extract the papers’ publication time, and count the quarterly numbers of the papers (for each of the three science categories) containing specific words, which are listed in Appendix.

3 Common features of multidisciplinary collaboration behaviors

3.1 Statistical properties of multidisciplinary coauthorship networks

Collaboration relationships can be expressed by a hypergraph, where nodes represent authors, and the authors of a paper (paper-team) form a hyperedge. 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 sciences (4.634). The size relation fits the reality that the sizes of research teams are usually larger in natural sciences, and smaller in social sciences [3].

A coauthorship network is extracted from a hypergraph as a simple graph, in which edges are formed between every two nodes in each hyperedge, and the multiple edges are treated as one. The technical terms “degree” and “hyperdegree” of nodes in theory of hypergraph are used to express the number of authors’ collaborators and papers respectively.

We consider the coauthorship networks of the considered papers in specific disciplines or science categories. All of the networks are highly clustered, assortative, and their average shortest path length scale as the logarithms of their number of nodes (Table 1). Those 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 goes solitary. In fact, 71.5% authors in social sciences belong to the giant component of the network generated by the whole data. Therefore, analyzing the collaboration of authors restricting in single discipline has limitations. So we proceed the analysis in the environment of all disciplines together.

Table 1 Specific statistical indicators of the analyzed networks.

| Network            | NN    | NE       | GCC   | AC   | AP   | MO  | PG |
|--------------------|-------|----------|-------|------|------|-----|----|
| PNAS 1999-2013     | 201,748 | 1,225,176 | 0.881 | 0.230 | 5.736 | 0.884 | 0.868 |
| Biological sciences| 184,872 | 1,150,362 | 0.881 | 0.232 | 6.385 | 0.883 | 0.880 |
| Physical sciences  | 24,766  | 101,166  | 0.933 | 0.452 | 10.89 | 0.986 | 0.455 |
| Social sciences    | 5,121   | 18,786   | 0.946 | 0.683 | 6.574 | 0.980 | 0.097 |
| Biophysics         | 13,480  | 48,012   | 0.905 | 0.177 | 7.665 | 0.957 | 0.636 |
| Interdiscipline    | 17,968  | 54,986   | 0.952 | 0.560 | 9.400 | 0.993 | 0.073 |

The indicators are the numbers of nodes (NN) and edges (NE), global clustering coefficient (GCC), assortativity coefficient for the number of collaborators (AC), average shortest path length (AP), modularity (MO), the node proportion of the giant component (PG). The values of AP for the first two networks are calculated by sampling 15,000 pairs of nodes.
3.2 Distribution type of degrees and hyperdegrees

Counting in the data (not restricted in each science category), the shapes of degree distributions for the authors in the three first-class disciplines are quite similar, so do the shapes of hyperdegree distributions (Fig. 2). Although collaboration level differs from one discipline 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 [39]. In the language of statistics, those distributions belong to one type: a mixture of generalized Poisson and power-law distribution. Regarding authors as samples, such a mixture distribution means those samples come from different populations, namely the collaboration behaviors of authors with few collaborators and papers differ from those with many collaborators and papers. In reality, authors mainly are teachers and students in institutes and universities, who can be viewed as two populations. Students on average only write few papers with few collaborators, but teachers are on the contrary.

An illustration free of disciplines in Reference [40] is given to explain the emergence of such type for degree distribution. With the same general ideas, a similar illustration 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 researcher is equal to the number of successes in a sequence of decisions made by the

Fig. 2 The distributions of collaborators/papers per author. The red circles show the distributions counted in the whole data, and the blue squares show those counted in papers of each science category. The purple lines are mixtures of generalized Poisson and power-law function. The parameters of fittings are shown in Table 2. The regions “G-P”, “C-O”, “P-L” stand for generalized Poisson, cross-over and power-law respectively.
candidates who want to coauthor with that researcher. Denote the number of those candidates to be \( n \). Suppose the collaboration probability of each candidate to be \( p \), and those “yes/no” decisions to be independent. Then, the hyperdegrees will follow a binomial distribution \( B(n, p) \). Poisson limit theorem shows when \( n \) is large and \( p \) is small, \( B(n, p) \) can be approximated by a Poisson distribution with expected value \( np \). The value of \( np \) varies from author to author, due to the diversity of authors’ ability to attract collaborators.

In reality, the decisions of authors could be affected by previous decisions, e.g. collaborating with the researchers who have publishing experience contributes to publish another paper. Hence, it is reasonable to regard small hyperdegrees as random variables drawn from a generalized Poisson distribution, which allows the occurrence probability of an event to affect by previous events \([41]\). For the researchers with large hyperdegrees, the numbers of their candidates are large enough that the “yes/no” decisions of candidates can be regarded to be independent. So their hyperdegrees could be regarded as random variables drawn from a range of Poisson distributions with sufficiently large expected values. The diversity guarantees the relative commonness for the existence of those authors with large hyperdegree, which appears as the fat tails of hyperdegree distributions.

| Table 2 The parameters of fitting 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 | 3.999 | 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 fitting function is \( f(x) = q(x)s f_1(x) + (1 - q(x)) f_2(x) \), where \( q(x) = e^{-(x-B)/(E-x)} \). The fitting 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; find \( B \) and \( E \) through exhaustion to make \( f(x) \) pass KS test (\( p \)-value > 0.01). 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.

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 affiliation networks. In academic society, collaborators of an author likely acquaint and so coauthor with each other. Organizational and institutional contexts drive the formation of transitive coauthorship, and so contribute the emergence of clusters and communities of authors.
The transitivity of a network can be quantified by two indexes in graph theory, namely global clustering coefficient (the fraction of connected triples of nodes which also form “triangles”) and local clustering coefficient (the probability of a node’s two neighbors connecting). High transitivity is a common feature of scientific collaboration networks [42], which can be reflected by the high values of global clustering coefficients in Table 3.

To what extent the transitivity is due to the activity of authors in academic society? The activity can be partly reflected through the number of collaborators, namely degree. Hence, the extent can be sketched through the correlation coefficients between degree and local clustering coefficient. Note that the correlation coefficients indicate the extent of a linear relationship between two variables or their ranks. The coefficients of variables $X$ and $Y$ generally do not completely characterize correlation, unless the conditional expected value of $Y$ given $X$, denoted by $E(Y|X)$, is linear or approximate linear function in $X$. The conditional expected value of local clustering coefficient given degree is the average local clustering coefficient of $k$-degree nodes, denoted by $CC(k)$. The approximatively linear trends of $CC(k)$ shown in Fig. 3 guarantee the effectiveness of correlation analysis in Table 3.

| Discipline          | Indicator | Mean   | Std   | SCC     | PCC      |
|---------------------|-----------|--------|-------|---------|----------|
| Biological sciences | LCC       | 0.8604 | -0.3977 | -0.4007 |
|                     | LTC       | 0.0005 | 0.0047 | 0.2753  | 0.0766   |
|                     | DN        | 21.086 | 0.5428 | 0.4003  |
|                     | HN        | 3.0146 | 15.468 | 0.0697  | -0.0459  |
| Physical sciences   | LCC       | 0.8058 | -0.3357 | -0.3819 |
|                     | LTC       | 0.0005 | 0.0045 | 0.3058  | 0.0735   |
|                     | DN        | 15.478 | 0.6252 | 0.3463  |
|                     | HN        | 2.6818 | 12.442 | 0.1689  | 0.0146   |
| Social sciences     | LCC       | 0.7839 | -0.1766 | -0.2625 |
|                     | LTC       | 0.0005 | 0.0058 | 0.2924  | 0.0498   |
|                     | DN        | 12.870 | 0.7228 | 0.4818  |
|                     | HN        | 2.2676 | 10.888 | 0.1750  | 0.0303   |

The indicators are local clustering coefficient (LCC), the local transitivity of collaboration (LTC), the average degree of node neighbors (DN), the average hyperdegree of node neighbors (HN). We calculate the mean of those indicators over authors, the Spearman rank correlation coefficient (SCC) and Pearson product-moment correlation coefficient (PCC) between each indicator and degrees. For the two indicators with small PCC, we calculate their standard deviation (Std).

Do the negative correlation coefficients, or equivalently, the decreasing trends of $CC(k)$ mean activity depresses transitivity? A positive answer to it is against common sense. In fact, those reflect the reality that the local clustering coefficients of students are larger than those of teachers on average. Many students leave their research teams after graduations, so the students studying in different periods of time are unlikely to collaborate. Leaving makes...
the neighbors of students probably are in the same time period and even in the
same paper team. So students have high local clustering coefficient on average.
Hence the puzzling thing does not contradict with common sense, but is due to
insufficiency of measuring transitivity such a dynamical property by counting
“triangles” on a static network.

To design a more reasonable index measuring transitivity, we come back
to the original meaning of transitivity on coauthorship: the probability of two
uncoauthored collaborators of a researcher coauthoring in future. The prob-
ability 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 category (Table 3).
Note that the calculation is imitated in the dataset, and transitivity may hap-
pen in other journals. In fact, 74.62% authors only have one paper in PNAS
1999-2013, whose local transitivity is zero. So the transitivity values here are
underestimated.

The increasing trends of the transitivity probability of $k$-degree nodes (TC$(k)$ in Fig. 3) mean the neighbors of the authors with many collabora-
tors tend to coauthor in future on average. It means the activity contributes
to transitivity. It fits common sense: a teacher (usually with a large degree)
is more likely to introduce two students (who haven’t worked together) to
collaborate.

3.4 Homophily in coauthorship

Coauthorship is based on specific features of researchers in common, including
interest, geography, academic reputation, etc. The homophily phenomenon
appears in many social relations, and is called assortative mixing in network
science [43]. Whether authors prefer to coauthor with others that are similar
in social activity or productivity in each science category? The social activity
and productivity of authors can be quantified by two scalar indexes of nodes,
namely degree and hyperdegree respectively. Then the preference of an index
could be sketched through the correlation coefficient between two variables,
namely the index of nodes and the average index of each node’s neighbors.
Positive correlation means assortative, negative disassortative, and zero no
preference.

Degree assortativity is feature of scientific collaboration networks [43]. It
means sociable researchers (with many collaborators) will preferentially coau-
thor with other sociable researchers, and unsociable to unsociable. Hence one
can expect to get a core of sociable authors tending to stick together and sur-
rounded by a less dense periphery of less sociable authors. It has emerged in
the data. The proportion of top 5.99% most sociable authors (measured ac-
cording to degree) having coauthored with another such author is 99.5% [44].
The proportion may even be underestimated, because those authors proba-
bly coauthored before 1999 or in other situations. Note that the splitting and
Fig. 3 The relation between degree and specific indicators. From $k = 1$ to max(degree), average over authors with $k$ degree for specific indicators, viz. local clustering coefficient (CC($k$)), local transitivity of collaboration (TC($k$)), the average degree of node neighbors (DN($k$)), the average hyperdegree of node neighbors (HN($k$)). The data are binned on abscissa axes to extract the trends hiding in noise.

merging errors of the used name disambiguation method affect the proportion at certain levels. Even so, the proportion is still remarkable.

If sociable researchers preferentially coauthor with sociable ones, then there will exist many sociable researchers, which is against the empirical data. Now we analyze the influence of the social activity of authors on degree assortativity. For the nodes with $k$-degree, denote the average degree of their neighbors by DN($k$). There exists a transition in DN($k$) of each empirical dataset: the head part has a clear increasing trend, but the tail part does not (Fig. 3). It means that degree assortativity are mainly contributed by the authors with small degree.

Actually, the transition fits common sense. Consider the research team of a teacher. Members of small research teams are likely to write papers together on average. Hence the authors in a small research team may have similar degrees. As the cumulative size of the research team increases over time, the degrees of students, on average, do not increase with the degree of their teacher, which leads to the non-positive and positive slopes of DN($k$) in large and small $k$ regions respectively.

The correlation coefficient between hyperdegree and the average hyperdegree of neighbors is around zero in each of the three science category (Table 3). For the nodes with $k$-hyperdegree, denote the average hyperdegree of their neighbors by HN($k$). It means choosing collaborators is free of the productiv-
ity factor. In reality, members of a research team may have various scientific ages (newcomers, incumbents), so different hyperdegrees. Since collaborations mainly happen in a research team, collaborators of an author could have various hyperdegrees, 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 derive that most new appearing authors of the data (it does not mean the authors have not published papers on PNAS before 1999) collaborate with at least one author who has published a paper in the data. The proportions of those authors are 79.22\%, 71.17\% and 65.12\% in biological, physical and social sciences respectively.

4 Features of interdisciplinary collaboration behaviors

4.1 Interdisciplinarity of disciplines

The data have 58.1\% and 48.2\% papers of social and physical sciences belonging to interdisciplinary research (Fig. 4). Only 7.4\% papers of biological sciences belong to interdisciplinary research, but occupy 40.8\% interdisciplinary papers. Papers of biophysics (which belong to biological sciences) occupy 37.1\% interdisciplinary papers. So 90.9\% interdisciplinary papers of biological sciences belong to biophysics. Those interactions are between biological and physical sciences, namely happens within natural sciences (Fig. 5). There exist 22.5\% interdisciplinary papers between natural and social sciences.

The discipline information can be used to classify authors into science categories: if one of his/her papers belongs to a discipline, an author can be classified into the discipline, so into the corresponding sciences. The giant component of coauthorship network PNAS 1999-2013 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 component. Note that 85.9\% authors of interdisciplinary papers (less than that of the physical sciences) belong to the giant component, which is underestimated due to data boundary. There are 49.2\%, 46.0\% and 7.3\% authors of social, physical and biological sciences who published interdisciplinary papers. The indicator is statistically significant high in social sciences, which undermines the common sense that socialists engage in research solitary. In fact, there has been a move towards increased interdisciplinarity in recent decades in social sciences, especially in information science and library science [45].

It seems above analysis process could be implemented to second-class disciplines to obtain a high-resolution result. Since some disciplines only have a few papers, e. g. 17 papers in political science, the analysis would lose statistical meaning. Here we provide a bird view of the interactions between second-class discipline (Fig. 6), where two disciplines are connected if there is a paper belonging to them simultaneously. As the bird view shows, no discipline is isolated. More details of the view are presented in our previous work [38].
Fig. 4 Interdisciplinary extent of first-class disciplines in the view of authors and papers. The giant component is that one of the coauthorship network extracted from PNAS 1999-2013, in which authors are identified by their names on papers. The discipline information is provided by the journal.

Fig. 5 Interdisciplinary research between and within biological, physical and social sciences. The panel shows how many papers and authors of PNAS 1999-2013 are involved in the interdisciplinary research between each two of the three science categories and within each individual of them.
Fig. 6 The big picture of interaction between the second-class disciplines. The graph is built based on the discipline information of interdisciplinary papers of PNAS 1999-2013. The communities of networks (distinguished by colors) are detected by the Louvain method [37].

4.2 Which authors and papers tend to involve interdisciplinary research?

The data show that in each science category, an author with more collaborators is more likely to engage in interdisciplinary research (Fig. 7a). Paper productivity of authors also has a significantly positive effect on the formation of interdisciplinary collaborations (Fig. 7b). The probability of a paper to be an interdisciplinary one is an increasing function of paper team size in physical and social sciences, but not in biological sciences (Fig. 7c). However, in all of the science categories, papers with a very large paper team size almost involve in interdisciplinary research.

The hub authors of the data neither concentrate on high degree nodes, nor distribute uniformly according to degree. This phenomenon is different from those of ER graph and the network generated by BA model (Fig. 8a). In this sense, the model in Reference [40] fits this phenomenon well. Authors with a comparable large hyperdegree are also responsible for the existence of giant components (Fig. 8b). Hence a considerable number of authors make coauthorship networks have the small-world property, not just the authors with numerous collaborators and papers. It means that interdisciplinary research is widely performed, and can be viewed as a reason for the emerged giant component of authors in all disciplines.
Fig. 7 Which authors and papers tend to involve interdisciplinary research. Panels (a,b) show that the more collaborators and papers an author has, the more likely the author does interdisciplinary research. Panel (c) shows the relationship between the probability of a paper to be an interdisciplinary one and the number of the paper’s authors.

Fig. 8 Which nodes are responsible for the existence of a giant component. In Panels (a) and (b), nodes are removed from high degree and hyperdegree to low respectively. In BA model, each new node connects to one old node. In ER graph, the connecting probability is 0.3. In both models, the number of nodes is 10,000. The parameter of the geometric model is that of the modeled network 1 in Reference [40].

The above view cannot be verified in the level of papers. Because removing the edges generated by interdisciplinary papers, the coauthorship network PNAS 1999-2013 still has a giant component. The reason is that some interdisciplinary authors acting as hubs also publish papers in a single discipline, so those hubs cannot be deleted from the network by removing interdisciplinary papers.

4.3 Exploring the grounds of interdisciplinary research

With the development of sciences, there is a tendency of fragmentation for disciplines: going to split into sub-disciplines and specific topics. Although the research objects are different, their research paradigms are in common, which can be grouped into four categories, namely theoretical research (e.g., modelling), experiment, simulation, and data-driven [46].

Many scientific problems are too complex to be understood through the methodology of single discipline. Integrating theoretical and methodological perspectives drawn from different disciplines creates a unified methodology
for research problems and even vocabulary used to present concept in specific disciplines [47]. The integration drives the formation of transdisciplinary disciplines [48]. For example, systems science, as a methodology, studies systems from simple to complex, from natural to social sciences. Complexity science risen in 1980s is a new stage in the development of systems science.

Common research paradigms and methodology, especially those integrated as transdisciplinary disciplines, give grounds for the interactions between science categories and for the formation of giant components in coauthorship networks. To validate the universality of those paradigms and methodologies in the three science categories, we analyze the contents of papers in the level of words. A paper containing a topic word means the corresponding topic is utilized or discussed by that paper. We choose “model”, “experiment”, “simulation” and “data” to represent the four basic research paradigms, and choose “system”, “network” and “control” to represent three typical topics of systems science.

We extract the quarterly numbers of papers containing those words in the three science categories respectively, which illustrate the developmental trends of the topics expressed by the selected words in each category respectively (Fig. 9). Those time series are non-stationary, but first-order integrated (Table 4). Restricting in each science category, we test the cointegrations between the quarterly number of papers containing each word and that of papers. The Johansen test shows that all tested pairs are cointegrated (Table 4). The results demonstrate that the development of each science category and that of each research paradigm or transdisciplinary topic obey an equilibrium relationship in the long-run.

Table 4 The P-values of Johansen test on specific time-series pairs.

|                     | system | network | control | model | experiment | simulation | data |
|---------------------|--------|---------|---------|-------|------------|------------|------|
| Biological papers   | 0.091  | 0.901   | 0.094   | 0.526 | 0.094      | 0.334      | 0.133|
| Physical papers     | 0.681  | 0.788   | 0.700   | 0.759 | 0.652      | 0.713      | 0.535|
| Social papers       | 0.775  | 0.848   | 0.737   | 0.783 | 0.856      | 0.772      | 0.876|

The series are listed in Appendix (Tables 7, 8, 9). Let the scalars of nominal significance levels of the following tests be 0.05. The augmented Dickey-Fuller test (maxlags = 3) shows that all of the involved time series are first order integrated. In Johansen test, we choose lags 2, and assume that there are intercepts and linear trends in the cointegration relations and there are quadratic trends in the data. The values larger than 0.05 indicate cointegration.

Over half the papers contain “system” and “control”. The proportions of papers containing “network” in all categories of sciences emerge an increasing trend (Table 5). The high proportions also verify the transdisciplinarity of systems science. To understand the deep reason of the transdisciplinarity, we must fall back on the definitions and functions of systems science. A system is composed of several parts, and has specific structure and functions. Systems broadly exist in nature and society, such as aerospace systems, ecosystem,
The core view of studying systems is “the whole is greater than the sum of the parts” (an immortal line of Aristotle).

Table 5  The means and variances of quarterly proportions of the papers in the three science categories containing specific topic words.

|                  | Biological papers | Physical papers | Social papers |
|------------------|-------------------|-----------------|--------------|
| system           | 0.759 (0.023)     | 0.800 (0.067)   | 0.676 (0.149) |
| network          | 0.189 (0.058)     | 0.222 (0.063)   | 0.277 (0.157) |
| control          | 0.766 (0.022)     | 0.451 (0.074)   | 0.533 (0.190) |
| model            | 0.104 (0.025)     | 0.350 (0.086)   | 0.173 (0.110) |
| experiment       | 0.727 (0.059)     | 0.735 (0.085)   | 0.722 (0.156) |
| simulation       | 0.768 (0.023)     | 0.682 (0.060)   | 0.441 (0.194) |
| data             | 0.933 (0.036)     | 0.802 (0.062)   | 0.854 (0.128) |

In each table cell, the value outsider and insider of the bracket is the mean and variance respectively.

The aim of studying systems is to reveal their operation rules, and to understand the macro behavior, etc. To achieve this goal, the relations between parts are needed to be analyzed deeply. Without regard for the functionalities and features of parts, systems can be abstracted as networks. Therefore, after entering in 21 century, the rapid development of research on networks (model, algorithm, etc.) breeds a new discipline, namely network science. Some researchers from biological, physical and social fields investigate their respective
problems under network framework. For example, social network analysis has become an important research technique in social sciences. Since understanding of natural and social systems is reflected in our ability to control them. Control theory (cybernetics) has a distinctly transdisciplinary mission to provide theories and approaches for comprehending the complex systems 49.

5 Conclusion

The case study on PNAS 1999-2013 verifies the similar transitivity and assortativity of collaboration behaviors in biological, physical and social sciences. The data demonstrate that the degree distribution types of the three science categories are identical, which are a mixture of generalized Poisson and power law distributions. The property also holds for hyperdegree. An illustration shows the type can be generated through authors’ “yes/no” decisions for collaborations and the diversity of authors’ ability to attract collaborations.

The data show that a considerable number of authors pursue interdisciplinary research. The interactions between disciplines are often regarded as a reason for the formation of giant components in coauthorship networks. Meanwhile, the data show that the formation goes beyond the authors with many collaborators and papers, but counts on a considerable number of authors. Therefore, the interactions can be considered to have been developed widely.

The data show that a substantial fraction of papers involve four typical research paradigms and three transdisciplinary topics in systems science. The development of each science category and that of each of the considered research paradigms and transdisciplinary topics obeys an equilibrium relationship in the long-run. Based on those statistical analyses, we could regard the universal research paradigms and transdisciplinary disciplines as grounds for the interactions.

The case study potentially provides a window for understanding aspects of multidisciplinary collaboration modes, due to the importance of PNAS in multidisciplinary research and to the provided accurate discipline information of papers. The selection of data might affect the details of our research findings: our results may not be interpreted as the behaviors of all researchers. Some aspects are indicative of the need for further research: Analysis of the interaction with the information of citation between sciences; Relationships of sciences or disciplines in the level of topics; Prediction of interdisciplinary fields via the trends of collaboration strength.

Acknowledgments

This work is supported by the open fund from the national university of defense technology teacher training project (No. 201602-06).
References

1. 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.
2. Snow CP. The two cultures. Cambridge University Press, 2012.
3. Kagan J. The three cultures: Natural sciences, social sciences, and the humanities in the 21st century. Cambridge University Press, 2009.
4. Larivière V, Gingras Y, Archambault É (2006) Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. Scientometrics 68(3): 519-533.
5. Abramo G, D’Angelo CA, Costa F (2012) Identifying interdisciplinarity through the disciplinary classification of coauthors of scientific publications. J Am Soc Inf Sci Technol 63(11): 2206-2222.
6. Glänzel W, Schoepflin U (1999) A bibliometric study of reference literature in the sciences and social sciences. Inform Process Manag 35(1): 31-44.
7. Hicks D (1999) The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics 44(2): 193-215.
8. Moody J (2004) The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. Am Sociol Rev. 69(2), 213-238.
9. 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/.
10. Bordons M, Zulueta MA; Romero F, Barrigón S (1999) Measuring interdisciplinary collaboration within a university: The effects of the multidisciplinary research programme. Scientometrics, 46(3), 383-398.
11. 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.
12. Hurd JM (1992) Interdisciplinary research in the sciences: Implications for library organizations. Coll Res Libr 53(4), 283-297.
13. Cooper G (2013) A disciplinary matter: Critical sociology, academic governance and interdisciplinarity. Sociology, 47(1), 74-89.
14. Porter AL, Rafols I (2009) Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. Scientometrics, 81(3), 719-745.
15. Chen S, Arsenault C, Gingras Y, Larivière V (2015) Exploring the interdisciplinary evolution of a discipline: The case of Biochemistry and Molecular Biology. Scientometrics, 102(2), 1307-1323.
16. Leydesdorff 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.
17. Hirsch Hadorn C, Pohl C, Bamer 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.
18. Siedlok F, Hibbert P. (2014) The organization of interdisciplinary research: Modes, drivers and barriers. Int J Manage Rev 16(2), 194-210.
19. Liu Y, Rafols I, Rousseau R (2012) A framework for knowledge integration and diffusion. J Doc 68(1), 31-44.
20. Lungeanu A, Huang Y, Contractor NS (2014) Understanding the assembly of interdisciplinary teams and its impact on performance. J Informetr 8(1), 59-70.
21. Levitt JM, Thelwall M, Oppenheim C (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Am Soc Inf Sci Technol 62(6), 1118-1129.
22. Porter AL, Roessner JD, Cohen AS, Perreault M (2006). Interdisciplinary research: Meaning, metrics and nurture. Res Eval 15(3), 187-195.
23. Rafols I, Meyer M (2010) Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. Scientometrics, 82(2), 263-287.
24. Zhang L, Rousseau R, Glänzel W (2015) Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject fields into account. J Am Soc Inform Sci Technol 67(3), 1257-1265.
25. 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.

26. Levitt JM, Thelwall M (2009) The most highly cited Library and information science articles: interdisciplinarity, first authors and citation patterns. Scientometrics, 78(1), 45-67.

27. Levitt JM, Thelwall M (2008) Is multidisciplinary research more highly cited?: A macrolevel study. J Am Soc Inform Sci Technol 59(12), 1973-1984.

28. Chen S, Arsenault C, Larivière V (2015) Are top-cited papers more interdisciplinary? J Informetr, 9(4): 1034-1046.

29. Larivière V, Gingras, Y (2010) On the relationship between interdisciplinarity and scientific impact. J Am Soc Inform Sci Tec, 61(1), 126-131.

30. Larivière V, Haustein S, Börner K (2015) Long-distance interdisciplinarity leads to higher scientific impact. Plos One, 10(3), e0122565.

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

32. Wan J, Thijs B, Glänzel W (2015) Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. Plos One, 10(5), e0127298.

33. Barabási AL, Jeong H, Neda Z, Ravasz E, Schubert A, Vízöek T. (2002) Evolution of the social network of scientific collaborations. Physica A 311: 590-614.

34. Newman M (2004) Coauthorship networks and patterns of scientific collaboration. Proc Natl Acad Sci USA 101: 5200-5205.

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

36. Kim J, Diesner J (2016) Distortive effects of initial-based name disambiguation on measurements of large-scale coauthorship networks. J Am Soc Inf Sci Tec. 67(6):1446-1461.

37. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech 10: P10008.

38. Xie Z, Duan XJ, Ouyang ZZ, Zhang PY (2015) Quantitative analysis of the interdisciplinarity of applied mathematics. Plos One 10(9): e0137244.

39. Milojević S (2010) Modes of Collaboration in Modern Science: Beyond Power Laws and Preferential Attachment. J Am Inf Sci Technol 61(7): 1410-1423.

40. Xie Z, Ouyang ZZ, Dong EM, Yi DY, Li JP (2016) Modelling transition phenomena of scientific coauthorship networks. [arXiv:1604.08891].

41. Consul PC, Jain GC (1973) A generalization of the Poisson distribution. Technometrics 15(4): 791-799.

42. Newman M (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci USA 98: 404-409.

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

44. Xie Z, Xie ZL, Li M, Li JP, Yi DY (2017) Modeling the coevolution between citations and coauthorship of scientific papers. Scientometrics 112, 483-507.

45. Levitt JM, Thelwall M, Oppenheim C (2011) Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Am Soc Inf Sci Technol 62(6), 1118-1129.

46. Hey T, Tansley S, Tolle KM (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft research, Redmond, Washington.

47. Haythornthwaite C (2006). Learning and knowledge networks in interdisciplinary collaborations. J Am Soc Inf Sci Technol 57(8), 1079-1092.

48. Grauwin S, Beslon G, Éric Fleury, Franceschelli S, Robardet C, Rouquier JB, Jensen P (2012) Complex systems science: dreams of universality, interdisciplinary reality. J Am Soc Inf Sci Technol 63(7), 1327-1338.

49. 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.

50. Milojević S. (2013). Accuracy of simple, initials-based methods for author name disambiguation. J Informetr 7(4): 767-773.
6 Appendix

6.1 Author name disambiguation

Current methods of author disambiguation can be classified into two classes: the methods only using the information of provided names, and the methods requiring additional information. There are two traditional methods in the first class, namely using surname and the initial of the first name token, using surname and initials of all name token. S. Milojević provided a hybrid method, which uses surname and the first name token’s initial. If a name with a first name token’s initial has two or more match names with different second name token’s initials, then all of the names are treated as different authors [50].

The misidentification caused by initial based methods is mainly through merging different authors as one [56]. There are 93.1% authors in PNAS 1999-2013 providing full first name. So we use name tokens per se, rather than their initial to identify authors. However, this can still produce errors by merging, if the authors provide exactly the same name. Chinese names were found to account for the repetition of names. We count the number and appearance frequency of names with a given name less than six characters and a surname among major 100 Chinese surnames. Note that people with those surnames account for 84.77% of the total population, where the statistical data come from Wikipedia. The small proportions of such kind authors in the data (2.7%), especially that of such authors publishing more than one paper (1.1%), make us feel quite confident that the impact of name repetition is limited. The detail of the analysis for each science category and specific disciplines is listed in Table 6.

Table 6 Specific statistical indicators of the analyzed networks.

| Data                  | FI   | HI   | AI   | PN   | a    | b    | c    | d    | e    | f    |
|-----------------------|------|------|------|------|------|------|------|------|------|------|
| PNAS 1999-2013        | 138,320 | 137,385 | 164,714 | 202,664 | 2.9% | 0.4% | 95.1% | 39.1% | 2.7% | 1.1% |
| Biological sciences   | 127,896 | 144,398 | 151,071 | 184,872 | 2.7% | 0.2% | 93.7% | 38.7% | 2.7% | 1.1% |
| Physical sciences     | 20,661  | 21,539  | 22,111  | 24,766  | 4.8% | 0.4% | 88.9% | 43.5% | 4.4% | 0.9% |
| Social sciences       | 4,845   | 4,919   | 5,001   | 5,121   | 2.3% | 0.1% | 93.6% | 42.6% | 2.2% | 0.3% |
| Biophysics            | 11,852  | 11,852  | 12,403  | 13,480  | 4.1% | 0.3% | 91.5% | 40.1% | 4.0% | 1.0% |
| Interdiscipline       | 12,545  | 12,904  | 13,162  | 14,012  | 2.6% | 0.1% | 93.2% | 45.5% | 3.7% | 0.5% |

The indicators are the numbers of authors identified by first initial (FI), hybrid method (HI), all initials (AI), provided name (PN), the percent of authors only providing first initial and surname and that of those authors with more than one papers (a, b), the percent of authors providing full first name (c), providing middle names (d), and the percent of authors with major 100 Chinese surnames and one given name less than 6 characters, and that of those authors with more than one papers (e, f).

The method adopted here will split one author as two or more, if the author does not provide his/her name consistently. Splitting underestimates the indexes used as evidences for universality of interdisciplinary research. Meanwhile, the misidentification caused by initial based methods is mainly through merging different authors as one, which makes the initial based methods unsuitable for analyzing the most productive authors and the size of largest
component. Wrong merging will overestimate the indexes in Fig. 4. We prefer to obtain a conservative result rather than a wrong result, so use the provided names to identify authors. In addition, the inaccurate caused by the adopted method does not much affect the distribution type of collaborators per author and that of papers per author, which we will show elsewhere.

6.2 The geometric hypergraph model of collaborations in Reference [40]

Denote the probability density functions of generalized Poisson and power-law by \( f_1(x) = a(a + bx)^{x-1}e^{-a-bx}/x! \) and \( f_2(x) = cx^{-d} \) respectively, where \( a, b, c, d \in \mathbb{R}^+ \), and \( x \) belongs to some subsets of \( \mathbb{Z}^+ \). Generate random variables of an \( f_1(x) \) with head \( f_1(x) \) and tail \( f_2(x) \) by sampling random variables of \( f_1(x) \) and \( f_2(x) \) with probability \( q \) and \( 1-q \) respectively. The modeled hypergraph is built on a cluster of concentric circles \( S_1^t, t = 1, 2, ..., T \) \( (T \in \mathbb{Z}^+) \) as follows:

1. Coordinate and influential zone assignment
   For time \( t = 1, 2, ..., T \) do:
   - Sprinkle \( N_1 \) nodes uniformly and randomly on \( S_1^t \). Identify each node, e.g. \( i \), by its coordinates \((\theta_i, t_i)\), where \( t_i \) is the generating time of \( i \);
   - Select \( N_2 \) new nodes randomly as lead nodes to attach specific zones: the zone of a lead node, e.g. \( j \), is an interval of angular coordinate with center \( \theta_j \) and length \( \alpha t_j^{\beta} \beta^{-1} \), where \( \alpha \in \mathbb{R}^+ \), and \( \beta \in [0.5, 1] \);

2. Connection rules
   For time \( t = 1, 2, ..., T \) do:
   - (a) For each new node \( i \), search the existing lead nodes whose zones cover \( i \). For each such lead node \( j \), generate a hyperedge with size \( m \) by grouping together \( i \), \( j \) and \( m-2 \) neighbors of \( j \) nearest to \( i \), where \( m \) is a random variable drawn from a given \( f(x) \) or the number of \( j \)'s neighbors plus two if the former is larger than the latter.
   - (b) Select \( N_3 \) existing nodes with non-zero degree randomly. For each selected node \( l \), generate a hyperedge by grouping together \( l \) and \( m-1 \) randomly selected nodes with the same degree of \( l \), where \( m \) is a random variable drawn from a given \( f(x) \) or the number of nodes with the same degree of \( l \) if the former is larger than the latter.

The parameters of the modeled geometric network in Fig. 8 are \( q = 0.9625 \) in Rule (a), \( q = 1 \) in Rule (b), \( T = 4,500 \), \( N_1 = 100 \), \( N_2 = 20 \), \( N_3 = 1 \), \( \alpha = 0.19 \) and \( \beta = 0.42 \). In Rules (a-b), \( f_1(x) \) is the Poisson distribution with mean 5.5, and \( f_2(x) \propto x^{-3.7}, x \in [10, 150] \).

6.3 Time series of topic words
Table 7 The quarterly number of the papers in physical sciences containing a certain topic word in PNAS 1999-2003

| Physical Papers | system | network | control | simulation | model | experiment | data |
|------------------|--------|---------|---------|------------|-------|------------|------|
| 24               | 18     | 2       | 6       | 7          | 19    | 18         | 19   |
| 27               | 17     | 4       | 10      | 5          | 18    | 14         | 23   |
| 31               | 23     | 2       | 17      | 12         | 22    | 22         | 29   |
| 34               | 29     | 6       | 12      | 6          | 26    | 24         | 24   |
| 29               | 18     | 3       | 11      | 8          | 14    | 16         | 24   |
| 29               | 24     | 4       | 13      | 6          | 19    | 21         | 18   |
| 38               | 28     | 6       | 13      | 10         | 26    | 23         | 24   |
| 32               | 28     | 8       | 14      | 8          | 23    | 21         | 25   |
| 27               | 18     | 7       | 11      | 8          | 17    | 16         | 21   |
| 32               | 20     | 6       | 14      | 10         | 20    | 19         | 28   |
| 28               | 17     | 4       | 15      | 10         | 20    | 16         | 22   |
| 25               | 17     | 3       | 8       | 11         | 15    | 20         | 20   |
| 123              | 97     | 29      | 59      | 26         | 75    | 81         | 100  |
| 35               | 28     | 5       | 10      | 7          | 22    | 29         | 29   |
| 57               | 40     | 12      | 20      | 15         | 39    | 38         | 42   |
| 38               | 32     | 5       | 14      | 12         | 27    | 25         | 25   |
| 62               | 47     | 12      | 31      | 13         | 47    | 47         | 53   |
| 42               | 34     | 9       | 12      | 16         | 28    | 29         | 28   |
| 50               | 39     | 10      | 23      | 17         | 37    | 35         | 44   |
| 42               | 33     | 12      | 15      | 8          | 27    | 27         | 28   |
| 97               | 79     | 7       | 37      | 24         | 60    | 56         | 80   |
| 85               | 64     | 12      | 38      | 23         | 48    | 52         | 67   |
| 63               | 44     | 20      | 29      | 18         | 43    | 37         | 51   |
| 62               | 46     | 14      | 22      | 16         | 40    | 40         | 52   |
| 82               | 66     | 20      | 34      | 42         | 62    | 54         | 64   |
| 90               | 76     | 24      | 42      | 32         | 66    | 56         | 68   |
| 70               | 60     | 15      | 32      | 21         | 49    | 46         | 57   |
| 73               | 61     | 15      | 31      | 27         | 51    | 56         | 60   |
| 72               | 62     | 18      | 24      | 32         | 59    | 50         | 61   |
| 86               | 64     | 20      | 42      | 28         | 59    | 53         | 72   |
| 101              | 89     | 17      | 50      | 34         | 77    | 77         | 88   |
| 93               | 77     | 27      | 48      | 30         | 71    | 61         | 77   |
| 114              | 98     | 28      | 48      | 47         | 95    | 85         | 101  |
| 86               | 76     | 16      | 38      | 33         | 71    | 64         | 72   |
| 103              | 86     | 32      | 59      | 47         | 78    | 69         | 87   |
| 78               | 60     | 19      | 36      | 28         | 63    | 55         | 73   |
| 79               | 65     | 18      | 37      | 30         | 59    | 56         | 63   |
| 136              | 113    | 33      | 57      | 53         | 107   | 91         | 115  |
| 138              | 117    | 38      | 62      | 53         | 105   | 85         | 116  |
| 111              | 82     | 20      | 51      | 43         | 88    | 79         | 86   |
| 152              | 125    | 46      | 79      | 64         | 122   | 105        | 123  |
| 170              | 144    | 47      | 83      | 74         | 127   | 115        | 138  |
| 127              | 108    | 33      | 71      | 52         | 108   | 75         | 102  |
| 140              | 119    | 37      | 81      | 48         | 102   | 98         | 101  |
| 170              | 143    | 44      | 73      | 67         | 135   | 110        | 137  |
| 125              | 107    | 30      | 60      | 55         | 101   | 90         | 105  |
| 165              | 130    | 37      | 80      | 63         | 134   | 98         | 128  |
| 137              | 123    | 38      | 71      | 57         | 116   | 96         | 104  |
| 177              | 138    | 48      | 81      | 60         | 125   | 120        | 144  |
| 157              | 132    | 32      | 85      | 62         | 125   | 112        | 124  |
| 140              | 117    | 40      | 62      | 57         | 106   | 92         | 123  |
| 152              | 128    | 37      | 79      | 59         | 118   | 105        | 119  |
| 206              | 171    | 52      | 104     | 86         | 159   | 141        | 173  |
| 203              | 168    | 45      | 107     | 85         | 165   | 147        | 165  |
| 194              | 149    | 57      | 87      | 88         | 156   | 135        | 157  |
| 178              | 146    | 44      | 88      | 86         | 158   | 130        | 146  |
| 162              | 134    | 39      | 88      | 66         | 129   | 114        | 130  |
| 179              | 145    | 58      | 89      | 73         | 138   | 119        | 147  |
| 200              | 168    | 57      | 110     | 77         | 162   | 142        | 156  |
Table 8 The quarterly number of the papers in biological sciences containing a certain topic word in PNAS 1999-2003

| Biological papers system network control simulation model experiment data |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 523             | 398            | 47             | 392            | 29             | 306            | 413            | 492            |
| 608             | 441            | 69             | 453            | 46             | 413            | 473            | 570            |
| 596             | 461            | 51             | 451            | 45             | 394            | 466            | 562            |
| 603             | 446            | 65             | 452            | 37             | 392            | 462            | 570            |
| 595             | 457            | 64             | 439            | 42             | 378            | 455            | 560            |
| 540             | 412            | 69             | 407            | 47             | 343            | 431            | 514            |
| 558             | 431            | 56             | 413            | 42             | 347            | 431            | 519            |
| 685             | 526            | 91             | 510            | 50             | 429            | 514            | 645            |
| 508             | 378            | 61             | 392            | 44             | 335            | 399            | 484            |
| 615             | 475            | 77             | 452            | 39             | 405            | 470            | 567            |
| 564             | 441            | 61             | 424            | 46             | 363            | 441            | 528            |
| 650             | 492            | 90             | 490            | 55             | 448            | 483            | 606            |
| 685             | 521            | 86             | 515            | 68             | 468            | 516            | 659            |
| 542             | 397            | 95             | 403            | 41             | 379            | 400            | 514            |
| 713             | 529            | 106            | 522            | 65             | 480            | 536            | 685            |
| 569             | 404            | 74             | 425            | 48             | 389            | 434            | 539            |
| 687             | 487            | 95             | 539            | 49             | 477            | 527            | 655            |
| 580             | 435            | 83             | 428            | 45             | 385            | 435            | 550            |
| 657             | 510            | 102            | 470            | 69             | 457            | 504            | 616            |
| 757             | 593            | 128            | 593            | 60             | 534            | 597            | 728            |
| 768             | 583            | 108            | 629            | 75             | 536            | 613            | 733            |
| 666             | 514            | 82             | 516            | 61             | 479            | 520            | 637            |
| 586             | 457            | 109            | 457            | 75             | 424            | 461            | 557            |
| 766             | 560            | 129            | 592            | 79             | 548            | 614            | 748            |
| 618             | 470            | 109            | 474            | 62             | 426            | 477            | 593            |
| 680             | 536            | 112            | 539            | 68             | 505            | 545            | 665            |
| 768             | 606            | 145            | 611            | 84             | 565            | 609            | 746            |
| 817             | 647            | 161            | 643            | 73             | 621            | 607            | 793            |
| 758             | 596            | 146            | 586            | 86             | 567            | 612            | 739            |
| 676             | 527            | 130            | 536            | 74             | 501            | 524            | 652            |
| 799             | 617            | 158            | 645            | 85             | 583            | 637            | 772            |
| 889             | 701            | 183            | 699            | 93             | 671            | 691            | 853            |
| 788             | 631            | 193            | 619            | 96             | 602            | 616            | 759            |
| 668             | 535            | 158            | 531            | 510            | 531            | 531            | 649            |
| 819             | 632            | 192            | 632            | 118            | 650            | 646            | 795            |
| 782             | 589            | 165            | 607            | 82             | 603            | 612            | 765            |
| 575             | 409            | 122            | 460            | 63             | 441            | 463            | 555            |
| 942             | 718            | 195            | 698            | 119            | 721            | 716            | 911            |
| 871             | 672            | 201            | 658            | 105            | 673            | 658            | 828            |
| 836             | 640            | 194            | 629            | 95             | 646            | 642            | 770            |
| 770             | 585            | 176            | 591            | 111            | 559            | 572            | 690            |
| 891             | 657            | 199            | 659            | 105            | 657            | 668            | 792            |
| 832             | 602            | 212            | 629            | 98             | 653            | 612            | 736            |
| 913             | 675            | 226            | 719            | 101            | 705            | 704            | 788            |
| 841             | 636            | 209            | 621            | 102            | 640            | 608            | 748            |
| 747             | 576            | 174            | 593            | 84             | 566            | 574            | 660            |
| 817             | 600            | 186            | 623            | 107            | 641            | 603            | 707            |
| 798             | 587            | 200            | 623            | 89             | 621            | 575            | 676            |
| 758             | 555            | 171            | 573            | 105            | 588            | 556            | 674            |
| 830             | 616            | 208            | 620            | 106            | 673            | 621            | 726            |
| 744             | 551            | 177            | 581            | 82             | 566            | 561            | 665            |
| 796             | 598            | 201            | 608            | 96             | 624            | 586            | 695            |
| 850             | 653            | 201            | 659            | 127            | 692            | 621            | 765            |
| 769             | 558            | 186            | 590            | 85             | 612            | 570            | 670            |
| 831             | 644            | 235            | 629            | 121            | 667            | 643            | 740            |
| 784             | 591            | 184            | 614            | 117            | 621            | 582            | 702            |
| 726             | 565            | 219            | 375            | 95             | 568            | 557            | 650            |
| 795             | 620            | 237            | 625            | 120            | 636            | 621            | 732            |
| 833             | 666            | 240            | 662            | 111            | 676            | 644            | 759            |
Table 9 The quarterly number of the papers in social sciences containing a certain topic word in PNAS 1999-2003

| Social paper | system | network | control | simulation | model | experiment | data |
|--------------|--------|---------|---------|------------|-------|------------|------|
| 6            | 5      | 2       | 3       | 1          | 5     | 2          | 6    |
| 9            | 5      | 3       | 4       | 1          | 7     | 5          | 7    |
| 3            | 1      | 1       | 0       | 0          | 2     | 0          | 1    |
| 4            | 3      | 0       | 2       | 0          | 4     | 3          | 4    |
| 6            | 4      | 0       | 4       | 0          | 4     | 3          | 6    |
| 3            | 3      | 0       | 0       | 0          | 2     | 2          | 2    |
| 6            | 5      | 3       | 4       | 0          | 3     | 4          | 4    |
| 9            | 5      | 2       | 5       | 2          | 6     | 4          | 6    |
| 8            | 5      | 1       | 2       | 2          | 4     | 3          | 7    |
| 8            | 5      | 1       | 4       | 1          | 3     | 3          | 8    |
| 10           | 7      | 4       | 5       | 2          | 6     | 7          | 8    |
| 7            | 6      | 2       | 4       | 2          | 5     | 4          | 4    |
| 5            | 3      | 2       | 1       | 2          | 2     | 1          | 4    |
| 6            | 3      | 0       | 4       | 0          | 4     | 3          | 5    |
| 10           | 7      | 1       | 6       | 0          | 4     | 7          | 8    |
| 6            | 4      | 1       | 4       | 2          | 6     | 4          | 6    |
| 7            | 5      | 1       | 1       | 0          | 5     | 0          | 6    |
| 19           | 11     | 5       | 8       | 2          | 11    | 8          | 17   |
| 2            | 2      | 2       | 2       | 0          | 1     | 0          | 1    |
| 6            | 4      | 2       | 5       | 0          | 5     | 1          | 6    |
| 8            | 4      | 2       | 6       | 1          | 3     | 2          | 7    |
| 10           | 8      | 2       | 4       | 0          | 8     | 0          | 8    |
| 6            | 3      | 1       | 4       | 1          | 4     | 3          | 6    |
| 11           | 7      | 4       | 6       | 4          | 10    | 8          | 9    |
| 11           | 7      | 0       | 4       | 2          | 9     | 3          | 7    |
| 14           | 8      | 3       | 6       | 3          | 9     | 8          | 11   |
| 16           | 11     | 4       | 9       | 2          | 6     | 5          | 12   |
| 16           | 6      | 6       | 6       | 4          | 11    | 4          | 13   |
| 8            | 5      | 2       | 4       | 1          | 5     | 7          | 7    |
| 12           | 11     | 4       | 6       | 2          | 9     | 7          | 12   |
| 18           | 14     | 4       | 14      | 4          | 15    | 8          | 16   |
| 17           | 9      | 2       | 9       | 2          | 13    | 9          | 17   |
| 19           | 14     | 5       | 10      | 6          | 16    | 8          | 16   |
| 25           | 17     | 5       | 17      | 4          | 19    | 8          | 23   |
| 42           | 28     | 13      | 26      | 11         | 31     | 23         | 33   |
| 19           | 14     | 7       | 7       | 6          | 15    | 9          | 17   |
| 16           | 7      | 3       | 6       | 2          | 6     | 9          | 14   |
| 29           | 19     | 4       | 12      | 8          | 21    | 11         | 26   |
| 22           | 20     | 8       | 9       | 6          | 20    | 5          | 20   |
| 32           | 18     | 8       | 17      | 10         | 25    | 16         | 27   |
| 41           | 29     | 12      | 28      | 6          | 28    | 26         | 35   |
| 35           | 24     | 11      | 15      | 11         | 30    | 12         | 30   |
| 53           | 41     | 16      | 28      | 8          | 43    | 26         | 47   |
| 40           | 32     | 13      | 22      | 7          | 32    | 17         | 35   |
| 41           | 29     | 19      | 29      | 9          | 31    | 19         | 39   |
| 43           | 35     | 15      | 25      | 6          | 36    | 24         | 39   |
| 45           | 35     | 18      | 27      | 12         | 40    | 21         | 40   |
| 34           | 26     | 12      | 21      | 2          | 30    | 20         | 29   |
| 43           | 25     | 12      | 29      | 12         | 38    | 20         | 38   |
| 39           | 31     | 17      | 25      | 7          | 30    | 20         | 36   |
| 33           | 24     | 8       | 17      | 8          | 27    | 15         | 31   |
| 46           | 28     | 19      | 30      | 6          | 34    | 17         | 42   |
| 53           | 36     | 17      | 34      | 17         | 44    | 20         | 52   |
| 43           | 27     | 18      | 29      | 7          | 38    | 21         | 37   |
| 43           | 34     | 18      | 27      | 8          | 34    | 24         | 38   |
| 37           | 30     | 13      | 27      | 11         | 31    | 16         | 36   |
| 50           | 36     | 20      | 27      | 14         | 43    | 23         | 45   |
| 48           | 26     | 11      | 33      | 11         | 36    | 25         | 41   |
| 41           | 30     | 13      | 32      | 9          | 31    | 22         | 40   |