

Network Analysis of Research Project Members in Grant-in-Aid for Scientific Research in Japan

Hozumi Morohosi

National Graduate Institute for Policy Studies, Tokyo, Japan

Email: morohosi@grips.ac.jp

How to cite this paper: Morohosi, H. (2023) Network Analysis of Research Project Members in Grant-in-Aid for Scientific Research in Japan. *American Journal of Operations Research*, 13, 71-92.
<https://doi.org/10.4236/ajor.2023.134005>

Received: June 26, 2023

Accepted: July 24, 2023

Published: July 27, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The researcher network that appeared in research projects funded by the Japanese government was analyzed. Several static and dynamic network analysis methods were applied to the data for 20 years to explore the fine structure of the researcher's network for grants. Our analysis shows that the long-term trend of researchers' group sizes has become smaller, particularly rapidly decreasing in recent years. Some findings on researcher behavior in joining a project have also been reported.

Keywords

Scientific Research Fund, Network Analysis, Power Law

1. Introduction

This study investigates a network of researchers in research projects supported by Grants-in-Aid for Scientific Research (GASR) of the Japanese government [1]. Using longitudinal data on funded projects, we analyzed the researcher's network to determine its characteristics, structure, formation, and historical evolution. The details of the accepted projects are available in an open database accessible online [2]. The dataset provides project details, including the list of researchers involved in the project, their institutions, the total budget, the project period, and the research field. We used data from the last 20 years in our study.

Collaboration among researchers is often of interest in research policy [3] [4]. A plethora of studies have used academic citation networks to identify network structure of researchers [5] [6]. However, our study adopts fund project data, which is rarely used in the literature. The study [7] measures the collaboration of researchers in the context of grant partners; however, it does not use network

analytic tools. An investigation of the Japan GASR has been conducted in a few studies [8] [9], while a network analysis of the data has not yet been conducted. Collaboration between researchers is often argued to be crucial for innovative research development. It is needless to say that funding is an essential resource for research. So investigating the network of researchers seeking research funding would give us a lot of helpful information on research style in Japan.

Each project comprises researchers who can join multiple projects. Thus, we can create a network whose nodes are researchers, and a link between two nodes (researchers) exists when two nodes are in the same project. Because new projects start every year and last for several years, the network is renewed yearly to include new and continuing projects. This study investigates such a sequence of researcher networks for 20 years to obtain insight into the relationship between researchers, which we cannot know by looking at the project data separately.

After a preliminary descriptive analysis of the data, we formulated the networks and analyzed them by computing several network characteristics, such as degree distribution and connected component size. In addition to the entire project network, we constructed sub-networks by choosing some of the projects from four different research fields. Analysis of these subnetworks elucidated the different structures of the four research fields. Moreover, statistical inferences regarding power-law fitting suggest a generation model for the network.

Further investigations were conducted by introducing a directed network. We defined the orientation on the edge (link) by considering role of the members in the projects. Each project nominates one leader, and the other members are organized as collaborators. The edge between the leader and collaborator is oriented with the leader as the tail and the collaborator as the head. Thus, the leader node has an out-degree, whereas the collaborator node has an in-degree. A directed network enables exploration of the microstructure of the network. The difference between the in-degree and out-degree distributions suggests a division of roles in research projects.

We performed an institution-wise analysis of individual in/out links by grouping researchers affiliated with the same institution. We divide the links into two types: those connecting researchers who belong to different institutions and those connecting researchers in the same institution. We refer to the former as external links and the latter as internal links. We observed that the ratio between the two types of links has changed over the past 20 years. Moreover, when applying a classification by the four research fields, we found a striking difference in the internal/external link ratio between the fields.

Finally, we performed Markov analysis of the node degrees. A high degree indicates that the researcher organizes or joins many projects. The time evolution of degrees explores the microscopic behavior of researchers. We estimated the transition probability between degrees for undirected and in/out-degree cases to determine the distinction between the in-degree and out-degree, which would characterize the researcher's behavior.

The contribution of researchers' collaboration to the research outcome is a long-standing question in the study of policy for science and technology [10] [11]. In the present study, we did not address this issue. This will be addressed in future research.

The remainder of this paper is organized as follows. Section 2 provides an overview of the data used in the study. Section 3 is the main body of this study, which explores the network structure of researchers in the GASR fund. Section 4 concludes the paper with remarks and a discussion.

2. Data Overview

This section outlines the data used in this study. We downloaded all the data from the site [2]. The GASR offers various grants depending on the research period, total cost, and the researcher's eligibility. **Table 1** displays the major categories of GASR. The total number of accepted projects in those categories is depicted in **Figure 1(a)**. As shown in the figure, GASR (A), (B), and (C) share the

Table 1. Major categories of grant-in-aid for scientific research.

categories	total budget	period
Specially Promoted Research	200 to 500 million JPY	3 to 5 years
Scientific Research	(S) 50 to 200 million JPY	5 years
	(A) 20 to 50 million JPY	3 to 5 years
	(B) 5 to 20 million JPY	3 to 5 years
	(C) 5 million JPY or less	3 to 5 years
Challenging Research	5 to 20 million JPY	3 to 6 years
Early-Career Scientists	5 million JPY or less	2 to 5 years

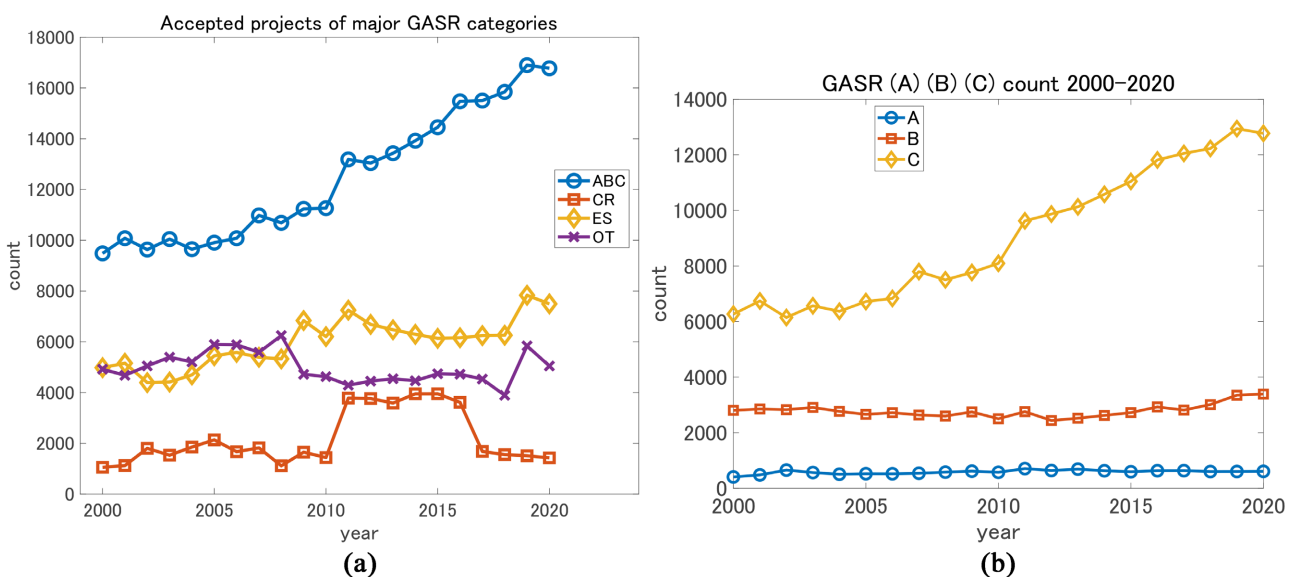


Figure 1. (a) Major categories of accepted GASR projects. ABC is the total of GASR (A), (B), and (C), CR is Challenging Research, ES is Early-Career Scientists, and OT is others. (b) GASR (A), (B), and (C).

central portion of the grant. Other categories, such as Specially Promoted Research, have specific research targets and frequent system changes adjusted according to timely research topics. Challenging research is emerging at an early stage; therefore, there are frequent system changes as well. Therefore, they are not appropriate for our study owing to the lack of continuity in the data. The Early-Career Scientists Fund is for young researchers and supports one-person projects. Therefore, it does not contribute to the researchers' network structure. This study focuses on the longitudinal data of GASR (A), (B), and (C) between 2000 and 2020. Japanese researchers recognize that GASR (A), (B), and (C) are vital research funds in their research works. **Figure 1(b)** shows the number of accepted projects in the GASR (A), (B), and (C). Among them, the GASR (C) shares a large portion and is increasing annually, whereas the others are almost constant.

As a preliminary study, before the network analysis, we checked the group size of the project. Each project organized a group of researchers to conduct the research. As the first observation of the data, **Figure 2** shows the mean group size of the projects for each category. The GASR (A) showed the largest mean group size, (B) the second largest, and (C) the smallest. Their budget size gives rise to the difference. As members share funds, a large budget can be shared among more members. The three categories exhibited similar time variations; therefore, the mean did. Because category (C) occupies the most significant portion, the average appears similar. They exhibited small peaks in 2007 and 2014, except for those that gradually decreased.

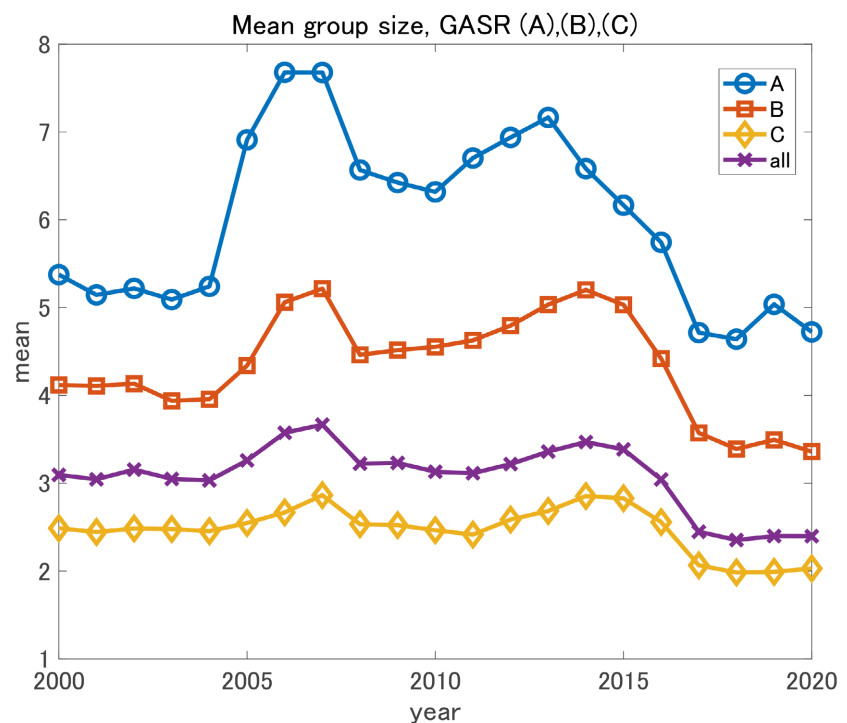


Figure 2. Mean group size of project GASR (A), (B), and (C).

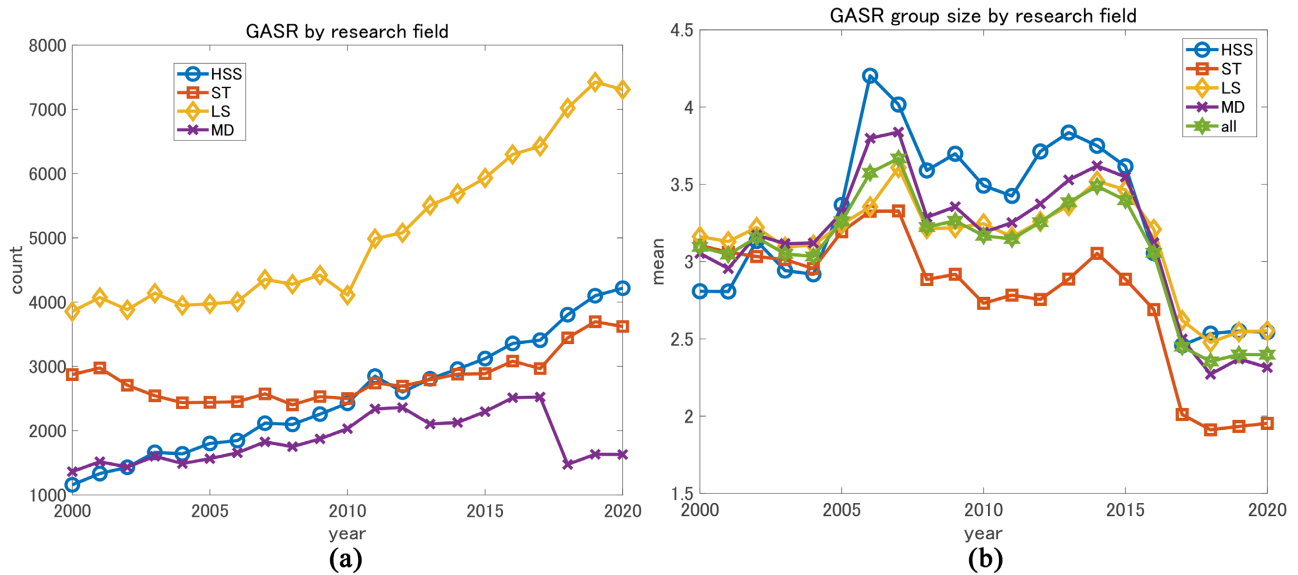


Figure 3. (a) Count of accepted projects for GASR (A), (B), and (C). HSS: Humanities and Social Sciences, ST: Science and Technology, LS: Life Science, MD: Multidisciplinary. (b) Mean group size of the project by field.

Another viewpoint is the differences across research fields. These projects are classified into four main fields: Humanities and Social Sciences (HSS), Science and Technology (ST), Life Science (LS), and Multidisciplinary Studies (MD). We analyzed the data according to their field. **Figure 3(a)** presents the number of projects in each of the four research fields. The large number and rapid growth of LS is apparent. The HSS showed steady growth and recently exceeded the ST. ST does not show an apparent increase until 2017, although it has some growth after 2017. The MD gradually increased until 2017 but dropped in 2018. This year, the GASR changed the classification of research fields for screening. Subsequently, some MD research topics have been classified into other fields. This change may cause decline in the MD and increase in other fields.

The mean group size analysis based on the field is shown in **Figure 3(b)**. All fields exhibited similar time variations. The ST showed the smallest size and specific change over time. Except for the small peaks in 2007 and 2014, their sizes decreased gradually, whereas it decreased rapidly in 2017. The HSS exhibited the smallest size until 2005. However, it showed rapid growth after 2005 and was the largest in size until 2015. After 2015, similar to other fields, it showed a rapid decline while remaining the highest.

Figure 4 shows the number of groups by size to closely examine the group size distribution. Small-size groups ($m = 1, 2, 3$) show an apparent increase, especially for $m = 1$, whereas larger groups do not increase. This contributes to the decrease in the mean group size.

A closer investigation to explain this time variation is yet to be conducted. As mentioned above, a system change in the screening process after 2017 may have contributed to this decrease. However, this has not been fully explained. A small peak around 2007 may have been affected by the administrative reform of na-

tional universities in Japan when university researchers (professors, associate professors, etc.) were strongly encouraged to apply to the GASR to obtain their research funds since the government began cutting the budget for universities. Referring to **Figure 4**, we find that the number of large groups increased rapidly in 2006. It is likely that researchers joined the project in flocks at the time. In the following section, we apply the network analysis methods to the research groups.

3. Network Data Analysis

In this section, several network analysis methods are applied [12]. By its nature, the network changes yearly by adding new project researchers and removing those in finished projects. Firstly we apply static analysis, such as degree and connected component size counting. Then, we follow the dynamic change of degrees to pursue the researcher’s behavior in funding. Both static and dynamic analysis enables us to investigate the researcher’s network structure and dynamical evolution. From the individual project data, we built a network of researchers as follows: A research project is conducted by one or a few researchers who are leaders, called principal investigators, and collaborators, called co-investigators. The key is that a researcher can be the leader of only one project, but can also be a collaborator for multiple projects. Therefore, researchers can participate as collaborators in several projects. To represent the relationship between researchers, we built a network by defining the vertex as a researcher and the edge as the connection between the leader and collaborators in a project. We also considered the project period. The research project continued for a certain period of three to five years. Thus, the entire network at year t should include the projects accepted in year t and those accepted in previous years and continued in year t . Specifically, we introduce some notations. Let r denote a research project in the

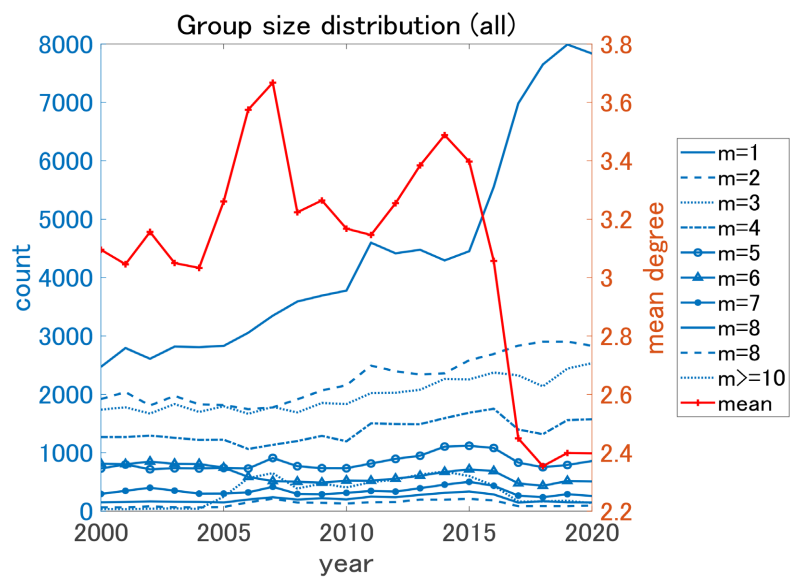


Figure 4. Group size distribution of project GASR (A), (B), and (C).

database used for our investigation. The leader of project r is designated as h_r . The set of collaborators in project r is denoted by M_r . Set M_r can be empty when project r consists of only one researcher, that is, the leader. T_r is the set of years, which denotes the period of the project r . For example, if project r continues from 2005 to 2007, $T_r = \{2005, 2006, 2007\}$.

First, we formulate the undirected graph at year t , $t = 2000, \dots, 2020$, denoted by $G_t = (V_t, E_t)$, where the vertex set V_t corresponds to researchers who are members of project r and the project term includes the year t . V_t is defined as $V_t = \{i : \exists r, \{h_r\} \cup M_r \ni i \text{ and } T_r \ni t\}$. The edge set E_t is the link between a leader and collaborators, both of which are members of project r whose term includes the year t . E_t is defined as $E_t = \{\{i, j\} : \exists r, i = h_r \text{ and } M_r \ni j \text{ and } T_r \ni t\}$.

Figure 5 illustrates how the network was built. In the figure, Projects 1 and 2 are active in year 1; their groups are $\{A, B\}$ and $\{C, D\}$ with leaders A and C , respectively. Therefore, G_1 has two components: A-B and C-D. In year 2, Project 3 becomes active and its group $\{B, D, G\}$ with leader B connects the existing components by linking B and D and adding a new vertex G . Therefore, G_2 has only one component. We construct network G_t in this manner for $t = 2000, \dots, 2020$.

Using the constructed network G_t , we know the real number of researchers, that is, the number of vertices $|V_t|$ and connected components each year in **Figure 6**. The number of researchers in **Figure 6(a)** shows a steady increase in most fields, which is similar to the number of connected components in **Figure 6(b)**. At first glance, they appear linked together to the number of accepted projects in **Figure 3(a)** with a slightly rapid increase. The following section

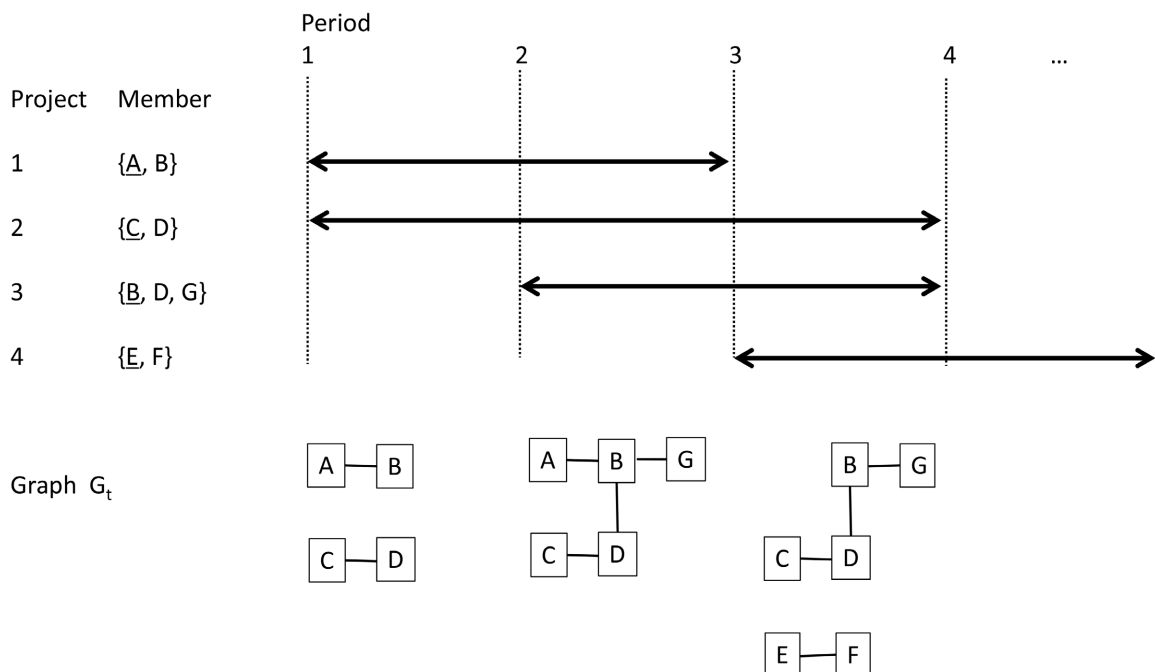


Figure 5. Yearly researcher's network construction. Project member is designated by A, B, ... The letter underlined is the leader. Project 1 has the period $\{1, 2\}$. Project 2 has the period $\{1, 2, 3\}$, etc.

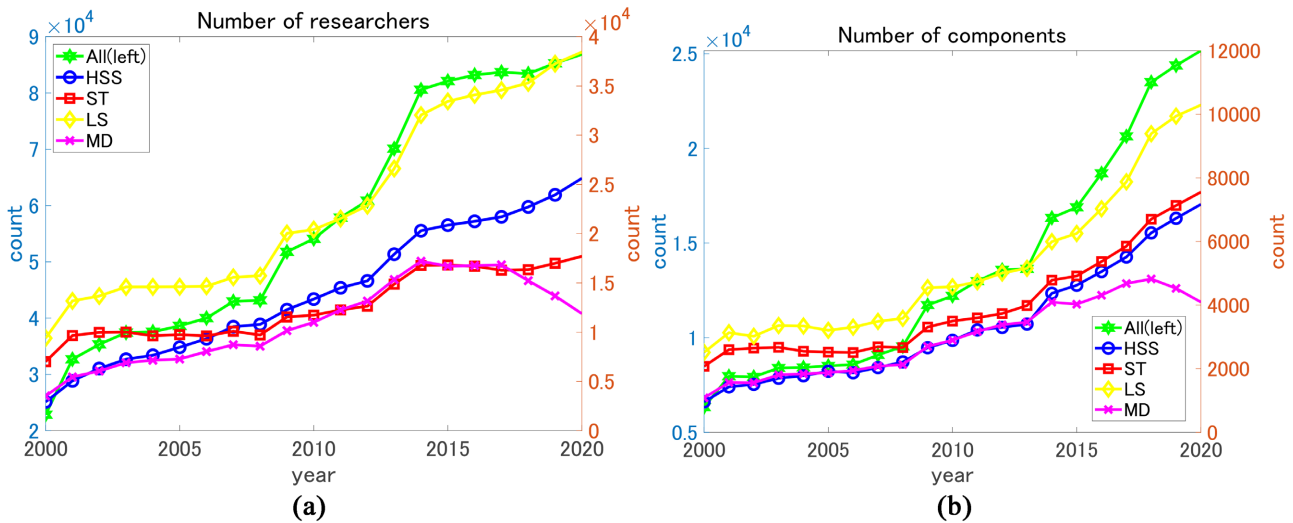


Figure 6. (a) Number of researchers for all and field-wise projects. (b) Number of connected components.

examines the differences in the network degree and component size distribution.

3.1. Degree and Component Size Distribution

Network G_t provides essential information about a researcher’s connections. We focused on the degree and component size distributions, and their time evolution. The degree of vertex i in network G_t is the number of edges connected to it: $d_t(i) = \#\{\{i, j\} \in E_t\}$. Here, we note that because one researcher can join multiple projects, there can be multiple edges to some vertex pairs. The distribution of degrees is shown in 7a. It would be interesting to compare the degree and group-size distributions shown in Figure 4. In the group-size distribution, groups with size $m = 1$ possessed the largest share. A simple thought suggests that these groups consist of a single member, and they would have a degree of zero in network G_t . However, this is not the case, as shown in Figure 7(a), where the vertices with degree one are the largest portion. Vertices with degree zero were third until 2017 and second after 2018. This gap suggests that researchers in one-person projects are not necessarily alone, but often connected to other researchers of different projects. Thus, they have a degree greater than one. Thus, the network analysis revealed the hidden structure of the researcher’s connectivity.

We further examined the network structure by computing the component-size distribution. Network G_t can be decomposed into several connected components. The size of each component, that is, the number of vertices contained in the component, was computed. The distributions are shown in Figure 7(b). The number of components with a size of one apparently increased, which contributed to a decrease in the mean component size. Simultaneously, we observed that the number of components with a size greater than one increased steadily. Nevertheless, these components comprise a relatively small percentage and do not increase the mean component size.

We computed the degree and component size distribution based on the filed-wise data. **Figure 8(a)** shows the mean degree, whereas **Figure 8(b)** shows the mean component size. The four fields exhibit similar time variations. The mean degree of each field was similar during the duration, although it differed in size. Among the four, a decrease in the ST was observed. It is interesting to compare **Figure 3(b)** and **Figure 8(b)**. In **Figure 3(b)**, the HSS, LS, and MD behave similarly; however, the MD is separated in **Figure 8(b)**. Although the HSS clearly shows the highest level of group size between 2005 and 2015 (**Figure 3(b)**), its component size does not show the highest level for this period. However, the peak of the LS component size around 2015 in **Figure 8(b)** was less than that in **Figure 3(b)**. The difference between the project group size and the network's actual component size suggests that HSS projects are less connected to each other than are LS projects. The degree size distribution seems to support this

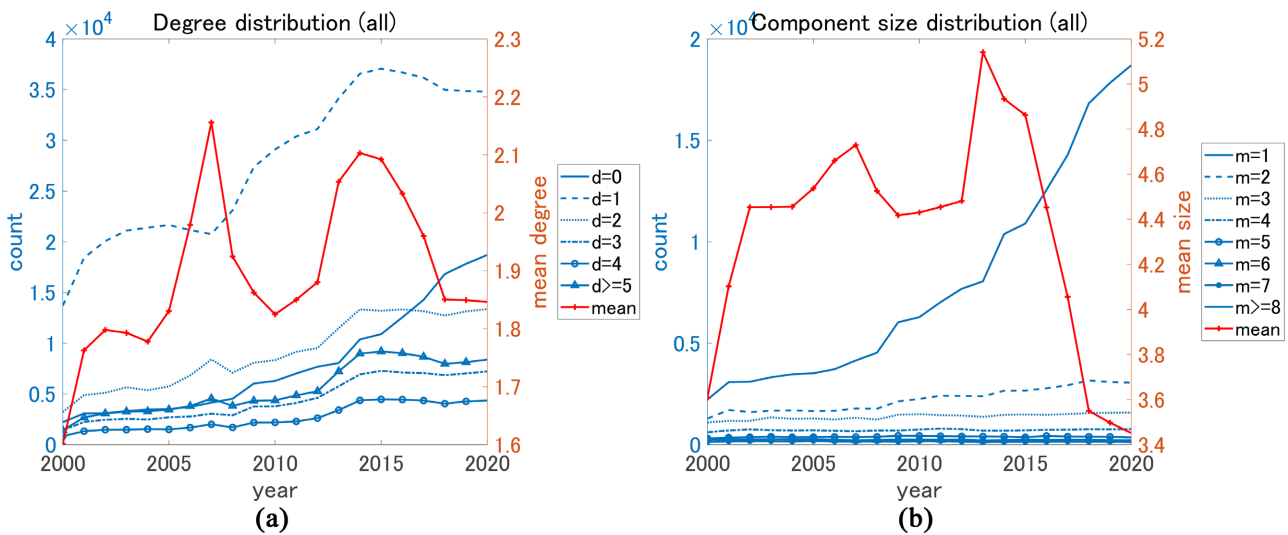


Figure 7. (a) Degree distribution of GASR researcher's network. (b) Component size distribution.

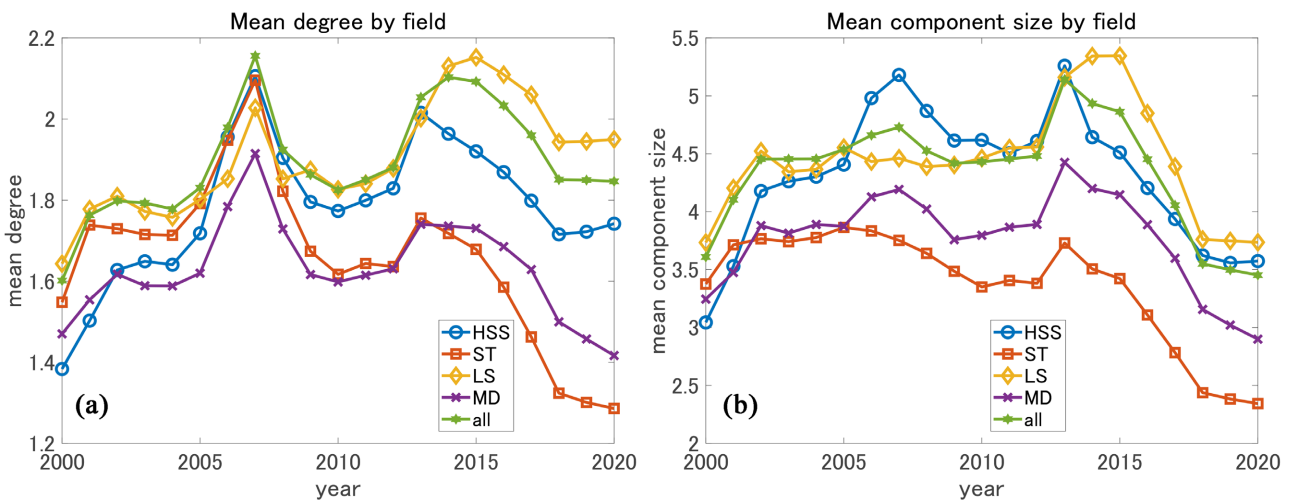


Figure 8. (a) Mean degree of GASR researcher's network by field. HSS: Humanities and Social Sciences, ST: Science and Technology, LS: Life Science, MD: Multidisciplinary. (b) Mean component size by field.

inference, because the degree of the LS vertex is higher than that of the HSS. The styles of HSS and LS researchers' connections differ.

Similar to other findings, MD and ST showed similar degree size time variations, but their component sizes appeared different. The MD maintains a larger component size than the ST, and a similar argument applies to the LS and HSS. MD researchers may be more likely to connect than ST researchers but less likely than LS researchers. The discrimination between the fields shown by the network analysis is displayed by plotting the mean group size and mean component size, as shown in **Figure 9**. ST and MD were consistently below the total mean, whereas HSS and LS were often above the total mean. Although the mean group sizes of the HSS, LS, and MD, as shown in **Figure 3(b)**, were similar, introducing the mean component size can be separated into two groups: MD versus HSS and LS.

3.2. Power-Law Fitting

In many studies of network science, it is well recognized that the degree and component size distributions follow a power law. We attempted to fit a power-law distribution to the researcher's network to determine how it agrees with the data by adopting the statistical method in [13].

Let X be the observed value and take a non-negative integer. A power-law distribution is described as a probability distribution, such that

$$p(x) = \Pr(X \leq x) = Cx^{-\alpha}, \tag{1}$$

for a parameter $\alpha > 1$ and a normalizing constant C . For our purpose, since a power-law is usually found above some small value of x , it is more common to work with a truncated version of power-law distribution with a lower bound x_{\min}

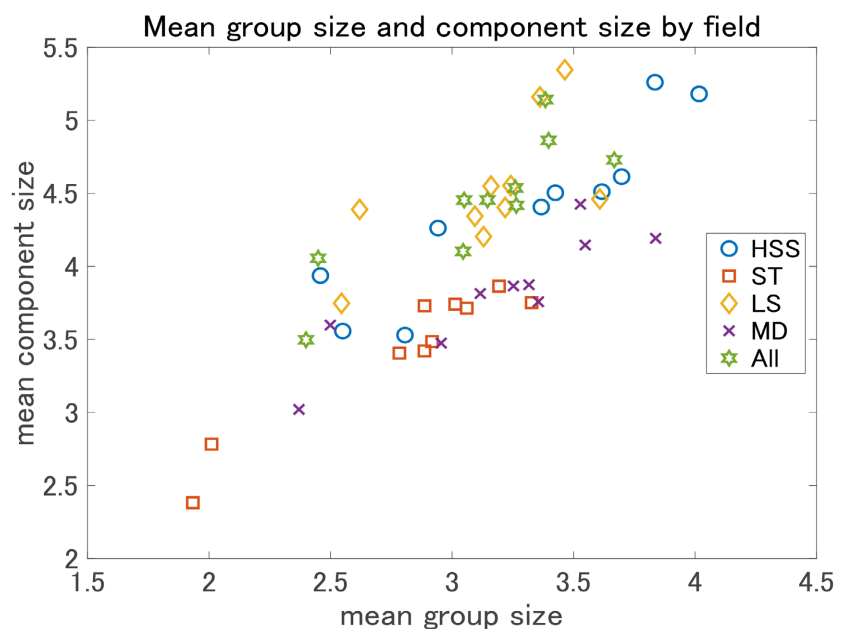


Figure 9. Mean group size and mean component size by field.

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}, \quad (2)$$

where

$$\zeta(\alpha, x_{\min}) = \sum_{n=0}^{\infty} (n + x_{\min})^{-\alpha}, \quad (3)$$

is the Hurwitz zeta function for normalization.

We estimate α and x_{\min} using the method given [13]. Suppose that the observations are x_1, \dots, x_n . First, we fix x_{\min} , then the log-likelihood function is given by

$$\mathcal{L}(\alpha) = -n \log \zeta(\alpha, x_{\min}) - \alpha \sum_{i=1}^n \log x_i. \quad (4)$$

Then the estimates $\hat{\alpha}$ is found by maximizing $\mathcal{L}(\alpha)$:

$$\hat{\alpha} = \arg \max_{\alpha} \mathcal{L}(\alpha). \quad (5)$$

We estimated the lower bound x_{\min} by minimizing the Kolmogorov-Smirnov statistic (K-S statistic).

$$D(x_{\min}) = \max_{x \geq x_{\min}} |S(x; x_{\min}) - P(x; x_{\min})|, \quad (6)$$

where $S(x; x_{\min})$ is the empirical distribution and $P(x; x_{\min})$ is the power-law distribution fitting best to the data for the fixed x_{\min} obtained by the maximum likelihood method explained above. To precisely define them, it is convenient to assume the data x_i are sorted in ascending order: $x_1 \leq x_2 \leq \dots \leq x_n$. The empirical distribution is defined as follows:

$$S(x; x_{\min}) = \frac{\sum_{x_i \geq x_{\min}} \mathbb{I}(x_i \leq x)}{\sum_i \mathbb{I}\{x_i \geq x_{\min}\}}, \quad (7)$$

where $\mathbb{I}\{A\}$ is the indicator function taking the value of 1 if A is true, and 0 otherwise. The power-law distribution is defined as

$$P(x; x_{\min}) = \frac{\sum_{x_{\min} \leq \xi \leq x} \xi^{-\alpha}}{\zeta(\alpha, x_{\min})}. \quad (8)$$

Note that P depends on α which is estimated by (5) with fixed x_{\min} . This implies that α depends on x_{\min} . So P depends on x_{\min} directly and indirectly by estimating α . To avoid complications, we take a practical approach to computing D . First, we assume the range of x_{\min} is between one and some number L : $1 \leq x_{\min} \leq L$. We selected $L = 15$ for our numerical study. Then for each x_{\min} , estimate $\hat{\alpha}$ by (5), namely obtain $\hat{\alpha}(1), \hat{\alpha}(2), \dots, \hat{\alpha}(L)$. They provide power-law distributions $P(x; x_{\min} = 1), P(x; x_{\min} = 2), \dots, P(x; x_{\min} = L)$. Using these distributions, we choose x_{\min} which minimizes the K-S statistic $D(x_{\min})$ defined in (6).

We estimate the power-law parameters $\hat{\alpha}$ and x_{\min} for the degree and component size distribution, respectively, of the researcher's network G_t , $t = 2000, \dots, 2020$. **Figure 10** shows the empirical distribution of the degree and

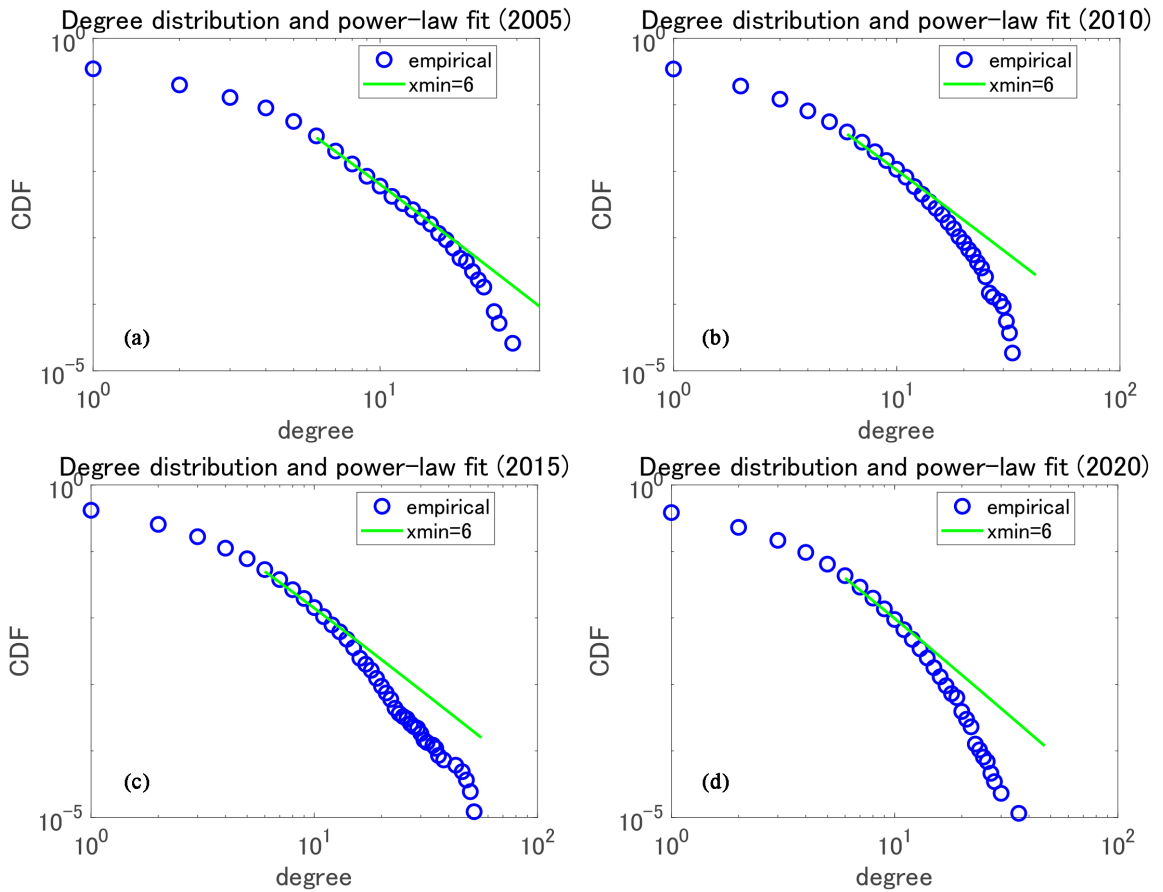


Figure 10. Degree distribution and power-law fitting. (a) All, 2005; (b) All, 2010; (c) All, 2015; (d) All, 2020.

the fitted line for the networks of four years: 2005, 2010, 2015, and 2020. Because we find that the K-S statistics are minimized when the parameter x_{\min} is around six for all years, we set $x_{\min} = 6$ for drawing the graph. It appeared to fit the data well except for the tail. The deviation at the tail is due to outliers caused by the small sample size. **Figure 11** shows the empirical distribution of the component size and the fitted lines for 2005, 2010, 2015, and 2020. The fit appeared to be better than that of the degree distribution. The empirical distribution was on the fitted line except for the extreme tail.

The estimates of power-law exponent α of degree distribution and component size distribution for the field-wise networks for all years are shown in **Figure 12**. They have a value of approximately three, often observed in real-world data.

The empirical distribution of the degree for the four fields in 2020 is shown in **Figure 13**, and that of the component sizes is shown in **Figure 14**. Regarding the degree distribution, HSS and LS do not fit well with the empirical distribution, particularly in the tail. Thus, the number of large-degree vertices is less than that expected by the power law. Meanwhile, ST and MD appeared to fit the power law well.

However, the power law appeared to fit the data perfectly in the component

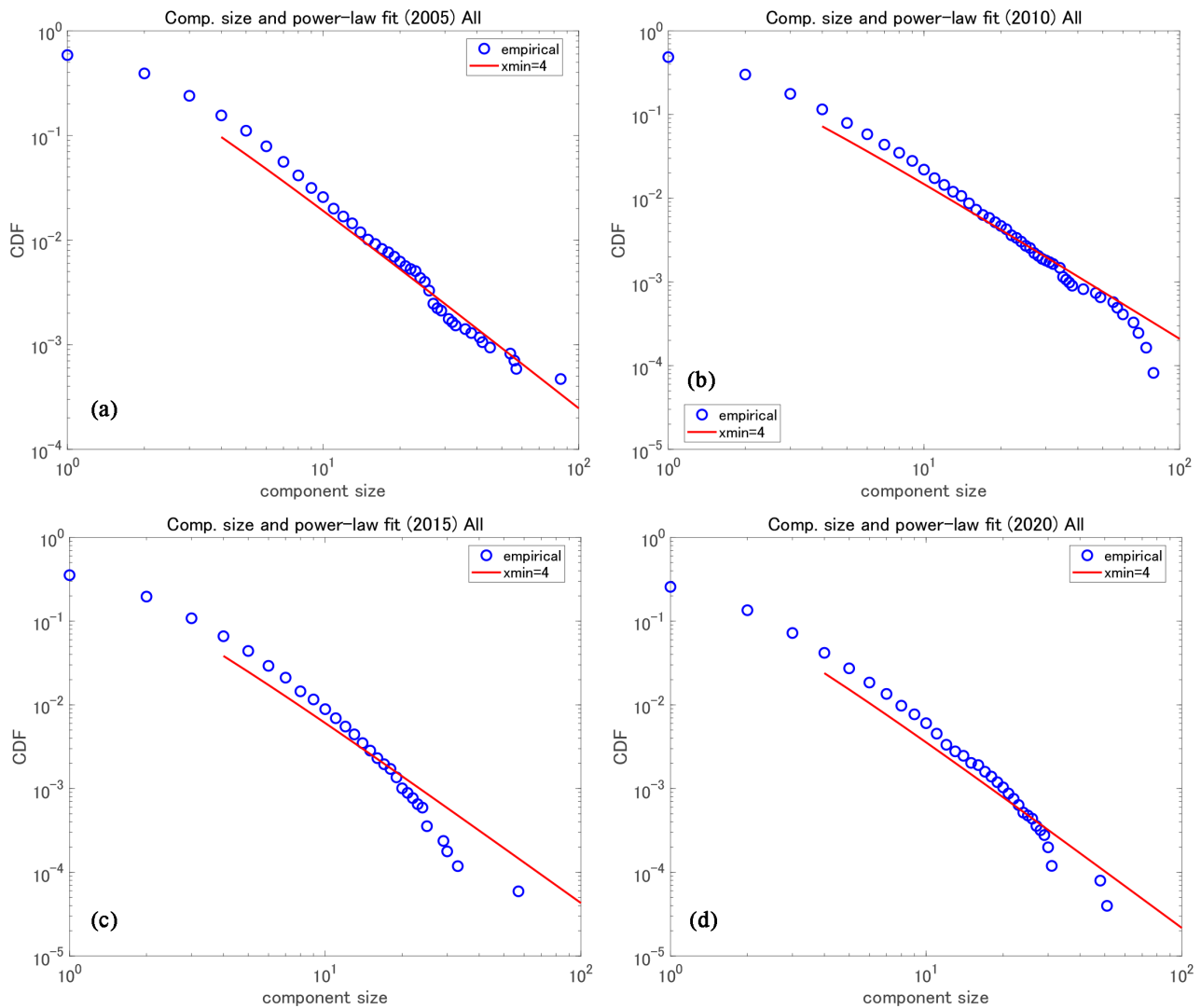


Figure 11. Component size distribution and power-law fitting. (a) All, 2005; (b) All, 2010; (c) All, 2015; (d) All, 2020.

size. The fitting of the law to the data suggests that some network formation models, particularly preferential attachment models, can explain the generation of a network better by expanding the component as a group than by increasing its degree as an individual researcher.

3.3. In-Degree and Out-Degree

For further investigation, we developed a directed graph model $D_t = (V_t, A_t)$, where the vertex set is the same as the undirected one, but A_t is the set of directed edges from the leader to the collaborator. So we define

$A_t = \{(i, j) : \exists r, i = h_r \text{ and } M_r \ni j \text{ and } T_r \ni t\}$. Note that (i, j) is an ordered pair of vertices, where i is the head and j is the tail. Therefore, the edge starts at the leader vertex and ends at the collaborator. Introducing a directed network makes it possible to consider the in-degrees and out-degrees. The in-degree of vertex i is the number of ingoing edges connected to vertex i : $d_t^-(i) = \#\{(j, i) \in A_t\}$. The

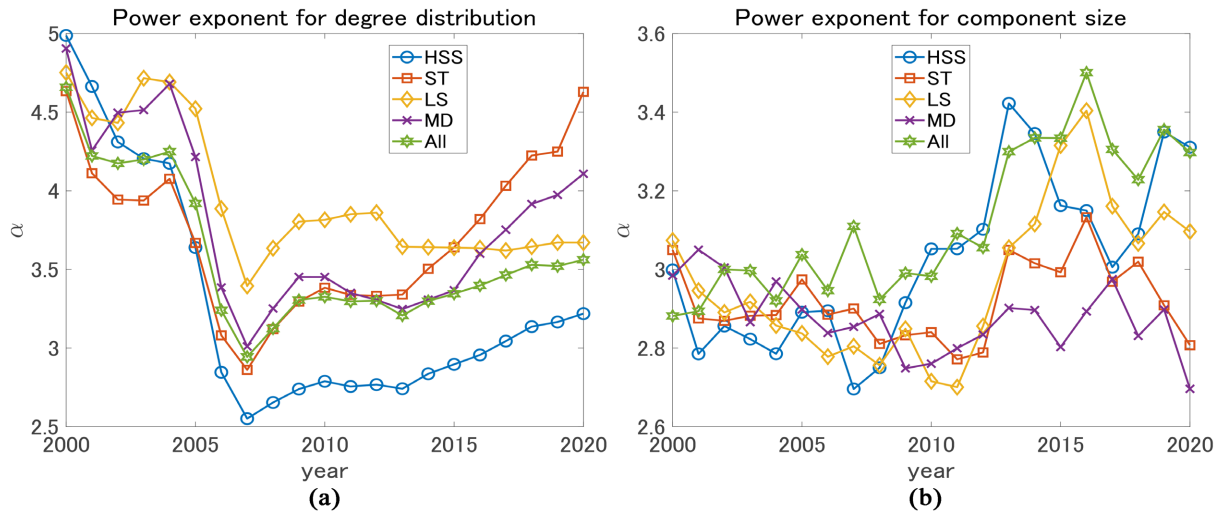


Figure 12. (a) Estimated power-law exponents for GASR researcher’s network degree distribution. (b) Those for component size distribution.

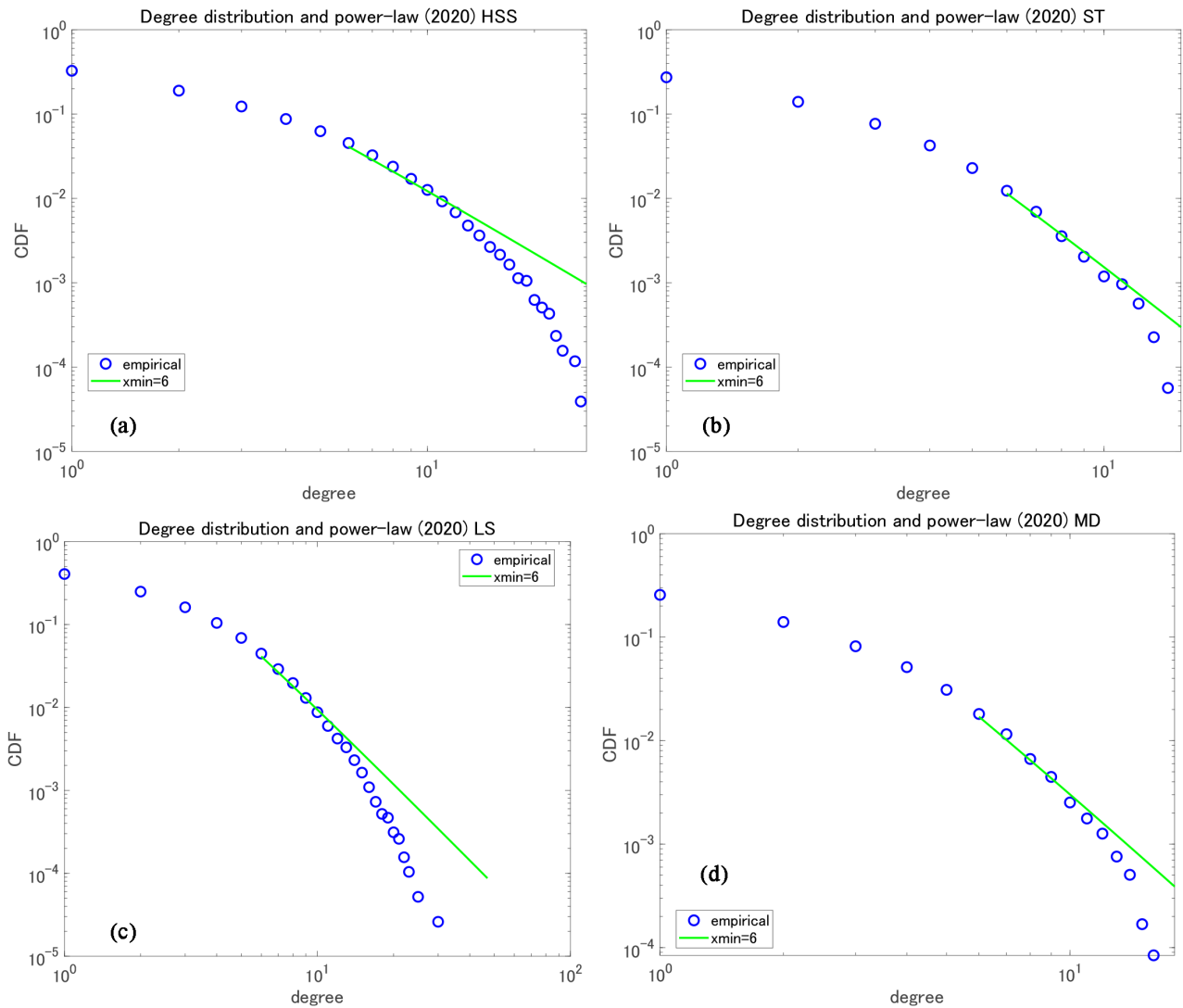


Figure 13. Degree distribution by fields and power-law fitting (2020). (a) HSS; (b) ST; (c) LS; (d) MD.

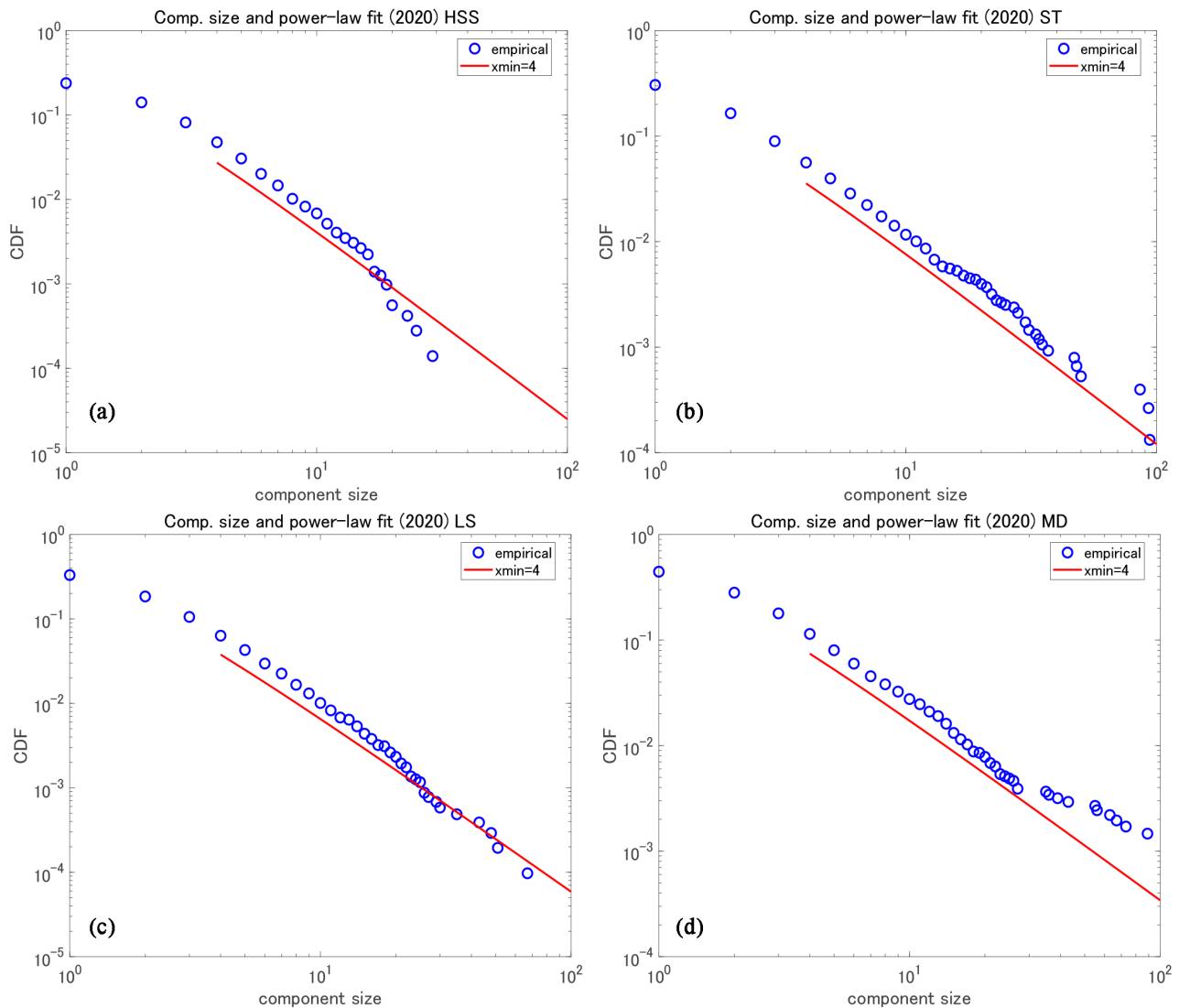


Figure 14. Component size distribution by fields and power-law fitting (2020). (a) HSS; (b) ST; (c) LS; (d) MD.

out-degree of vertex i is the number of outgoing edges connected to vertex i : $d_i^+ (i) = \#\{(i, j) \in A_i\}$.

Figure 15(a) depicts the distribution of in-degree, where the vertices with in-degree $d^- = 1$ are in the majority (The share is approximately 70%.) This implies that most researchers have joined a single GASR project as collaborators. On the other hand, looking at **Figure 15(b)** which shows the distribution of out-degree, vertices with the out-degree one $d^+ = 1$ are still the largest group, although vertices with an out-degree more than two $d^+ \geq 2$ have a large share (The percentage of out-degree one is approximately 30%.) A vertex with an out-degree corresponds to a project leader. The out-degree is the number of collaborators. **Figure 15(b)** indicates that project leaders often have more than two collaborators. Although the components with size one have a large share, as shown in **Figure 7(b)** the components with multiple members maintain their shares and are increase.

Next, we extend the in/out degree to include the institutions to which the researchers belong. The edges were grouped into two types by examining the institutions to which they were connected. We call an edge an internal link when it is between researchers who belong to the same institution and an edge an external link when it is between researchers who are affiliated with different institutions. **Figure 16** shows the number of internal and external links and the share of internal links to the total links. The share of internal links rapidly decreased until 2010 to approximately 50% and maintains that level. We assume that the increase in degrees is caused by an increase in external links. This indicates the researcher’s preference for outside rather than inside collaborators.

The field-wise analysis exhibited different internal and external link share trends, as shown in **Figure 17**. Among the four fields, LS showed the highest

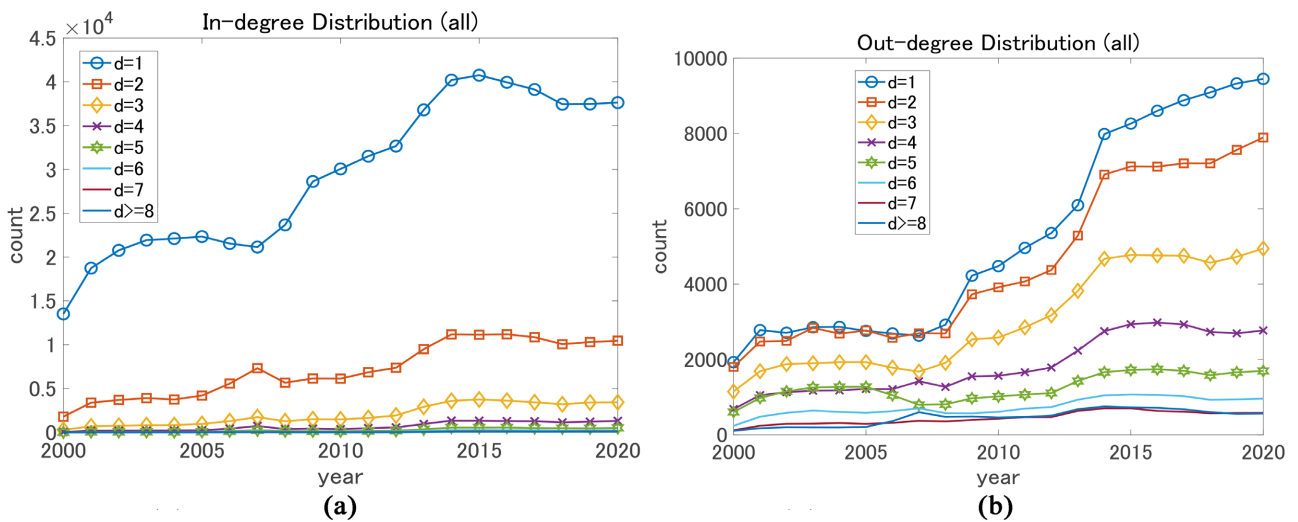


Figure 15. (a) In-degree distribution of researcher’s network; (b) Out degree distribution.

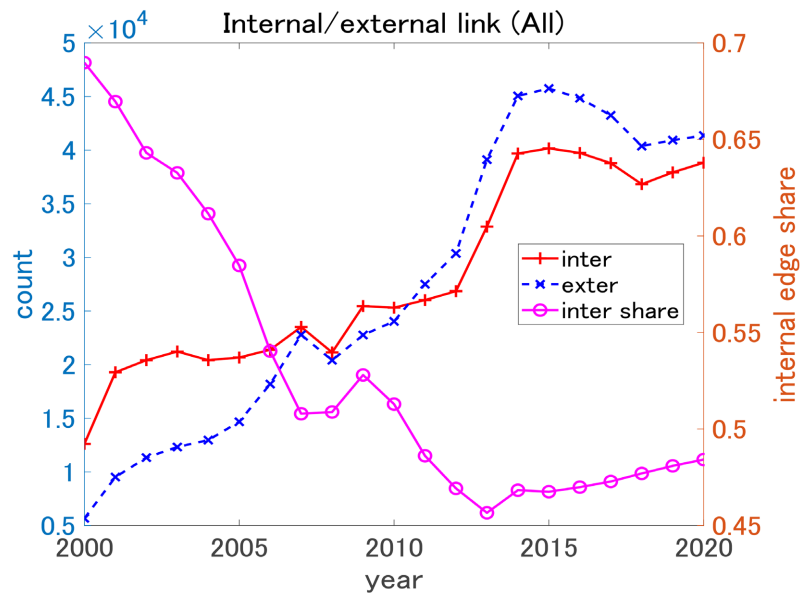


Figure 16. Internal external link ratio.

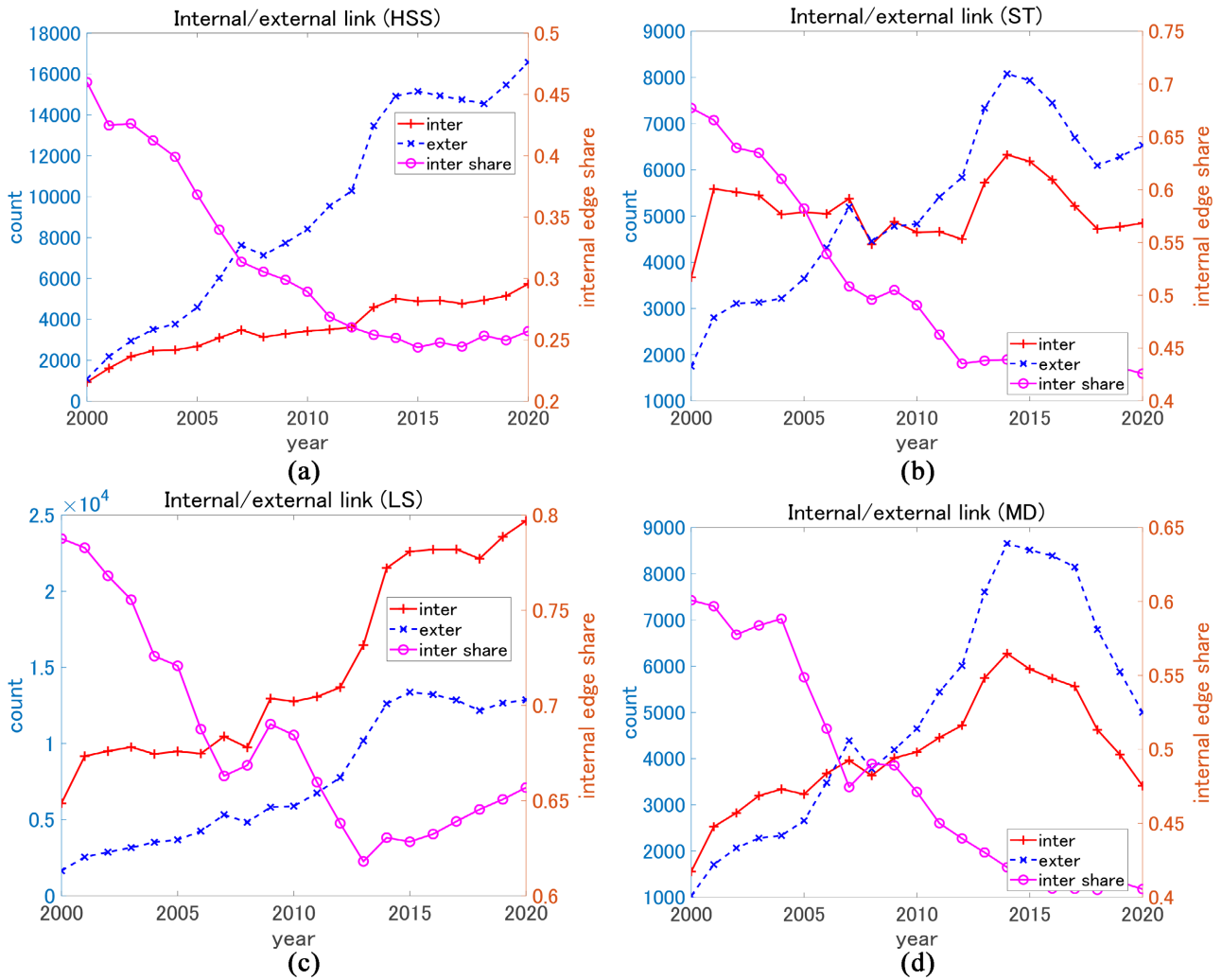


Figure 17. Internal external link ratio of four fields. (a) HSS; (b) ST; (c) LS; (d) MD.

share of internal links with 65%, whereas HS showed the lowest with 25%, in recent years. Therefore, LS researchers are likely to be connected to those in the same institution. In contrast, HSS researchers are less connected to those in the same institution than to other institutions. In Japanese universities, the LS sections often have large departments with many researchers. In contrast, the HSS usually has a small department with few researchers. The difference between them seemingly originates from the differences in the sizes of the departments affiliated with the fields.

3.4. Markov Analysis

We investigated the researcher network evolution by focusing on how the degree of the vertex (how many researchers each researcher has collaborated with) changes over time. The transition probability of degree is estimated as follows:

$$p_{ij} = \frac{\sum_{t,r} \mathbb{I}\{d_{t+1}(r) = j \text{ and } d_t(r) = i\}}{\sum_{t,r} \mathbb{I}\{d_t(r) = i\}}, \quad (9)$$

where p_{ij} is the probability that researchers have degree j in year $t+1$, given that they have degree i in year t . Here we implicitly assume stationarity of the process. The state set consists of non-negative integers and a special symbol NA, meaning that the researcher is not included in any project and does not appear in the network G_t . It is important to distinguish the NA from zero degrees. The former does not appear in the network G_t , whereas the latter appears as a single vertex with no connecting edges. We calculated the transition probabilities of degree d , in-degree d^- , and out-degree d^+ . The results are shown in **Figure 18**.

The in-degree transition probability contrasts significantly with the out-degree transition probability. In both cases, the probability that the degree does not change, that is p_{ii} , is the largest. However, for the in-degree case, the transition probability exhibits a gentle slope, suggesting that a transition to a different

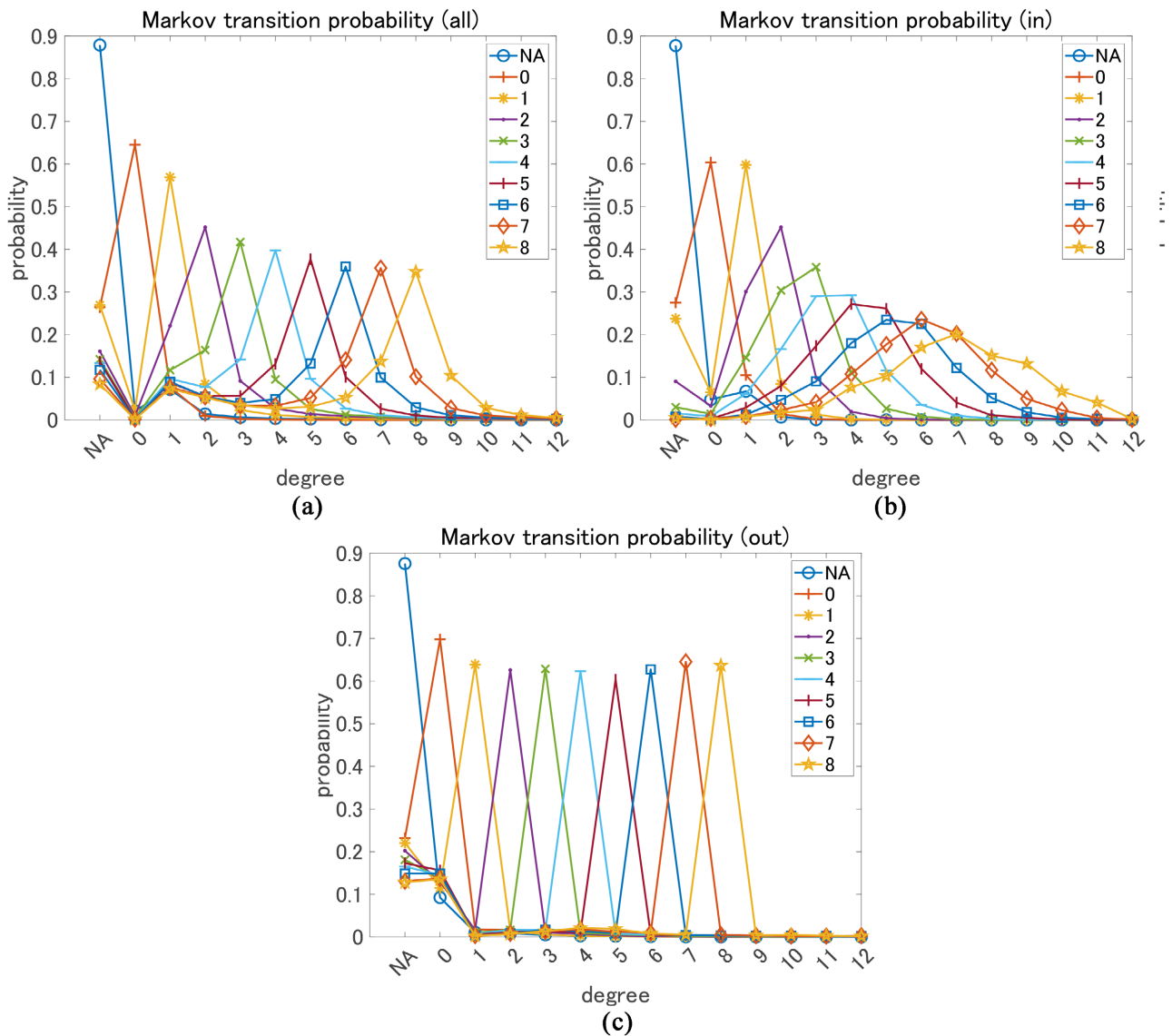


Figure 18. Markov transition probability of degree. (a) in and out (b) in (c) out.

degree often occurs, particularly for relatively large degrees. On the other hand, for the out-degree case, the transition probability shows a sharp peak, and almost all transitions are to NA, which implies that researchers do not transit as a project leader from one project to another continuously, but join the project as a collaborator frequently.

The transition probability can be illustrated by the distribution of the existence time of a researcher in a project, which we define as the number of years that a researcher has a positive degree in the networks G_t , that is, $N_i = \sum_t \mathbb{I}\{d_t(i) > 0\}$ for undirected degrees. Similarly defined are N_i^- and N_i^+ for in-degree and out-degree by replacing $d_t(i)$ by $d_t^-(i)$ and $d_t^+(i)$, respectively. The results are shown in **Figure 19**. The existence time N_i for the undirected degree is shown in **Figure 19(a)**, while N_i^- for in-degree corresponds to the case a researcher

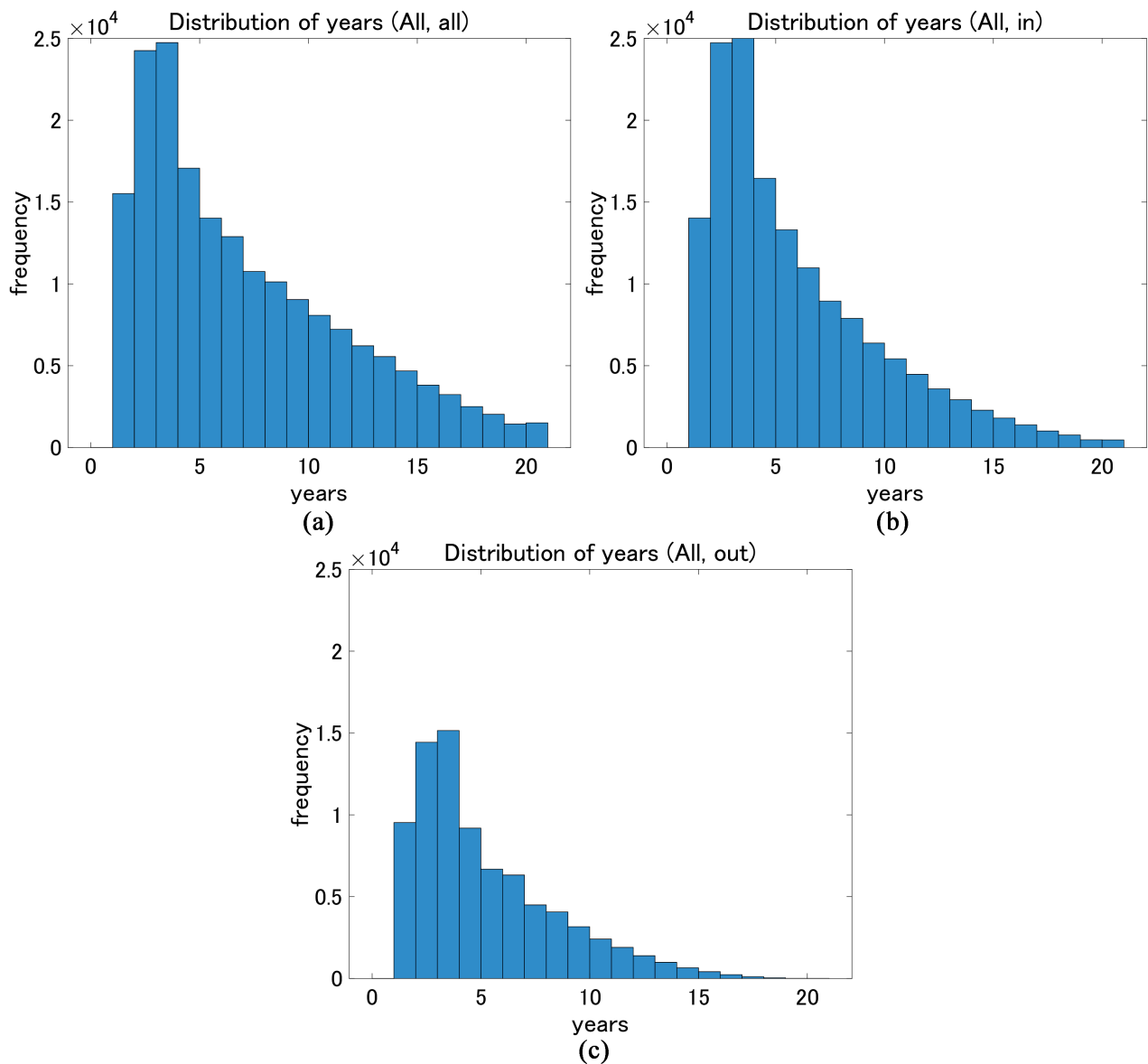


Figure 19. Distribution of years in which researchers join the project. (a) total, (b) as a collaborator, (c) as a leader.

plays the role of collaborator, and N_i^+ for out-degree corresponds to a leader. The first observation made by comparing **Figure 19(b)** and **Figure 19(c)** is that the frequency of leaders is much lower than that of collaborators. This implies that a researcher plays the role of a leader in a project less frequently. There is a peak at approximately three years, which is most likely the research period, as is the case with a collaborator. After the peak, the frequency decreased exponentially in both the in-degree and out-degree cases, with a slightly rapid rate in the leader case. These decays in frequency indicate that a researcher can join the GASR projects a few times, probably due to the tough competition of applications, and that playing as a leader several times is rare. Even if a researcher works as a leader, referring to **Figure 18(c)**, they would take a break to avoid becoming a leader of another project immediately after one project.

4. Concluding Remarks

In the conclusion, we summarize our main findings. Network analysis of the researchers' collaboration revealed that their connection was slightly wider than that of the research group that appeared in the project. The development of networks in different research fields explores distinct structures. HSS and LS researchers were more likely to form large networks, whereas ST and MD networks were small. In the directed graph analysis, we found that the in-degree distribution shows that vertices with an in-degree of one ($d = 1$) are the majority, which implies that most researchers join the GASR project as collaborators in a single project. However, the out-degree distribution, which counts the number of collaborators a leader has, shows that degree one is not necessarily predominant. However, vertices with multiple degrees were frequent. Furthermore, after classifying the links into internal and external institutions, we observed that the share of external links increased until 2013 to more than half, which implies that researchers are more likely to look for their collaborators outside of their institution. Field-wise analysis shows that this trend is observed in the HSS but not in the LS, where the share of internal edges has remained at more than half in recent years.

Finally, Markov analysis was applied to the dynamic evolution of the network to determine the historical changes in vertex degrees. The transition probability of the vertex degree shows a different pattern between the in-degree and out-degree. The transition probability of the in-degree is widespread; therefore, the number of projects that a researcher joins as a collaborator can change frequently. In contrast, the out-degree transition probability has a sharp peak, which suggests that a researcher often plays the role of a leader and avoids continuously becoming a leader from one project to another.

In our study, we observed that the component size of the researcher's network was slightly larger than that of the project group, but we did not find a widely spread network. Research projects supported by the Japanese GASR have been small-size over these twenty years in nominal descriptive and in-depth network

data analysis. This trend has increased in recent years. Currently, we do not have a good explanation for this observation. However, we recognize that the network analysis method is a powerful tool for obtaining in-depth knowledge of complicated system structures. Since our study is limited to GASR, we cannot discuss whether our results can apply to other research fund cases. However, GASR is the largest research fund in Japan, and it involves most Japanese researchers. Our finding would not be special but have generality for the network structure of researchers in Japan.

Acknowledgements

The author thanks Mr. Tadashi Nakagawa, Mr. Takuro Matsumoto, Mr. Hitoshi Koshiba, and Prof. Tatsuo Oyama for their kind support and valuable discussions. This research was supported by the SciREX Coevolution Project.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research. <https://www.jpsps.go.jp/english/e-grants/>
- [2] National Institute of Informatics, Grants-in-Aid for Scientific Research Database. <https://kaken.nii.ac.jp/en/>
- [3] Cummings, J.N. and Kiesler, S. (2005) Collaborative Research across Disciplinary and Organizational Boundaries. *Social Studies of Science*, **35**, 703-722. <https://doi.org/10.1177/0306312705055535>
- [4] Koontz, T.M. and Thomas, C.W. (2006) What Do We Know and Need to Know about the Environmental Outcomes of Collaborative Management? *Public Administration Review*, **66**, 111-121. <https://doi.org/10.1111/j.1540-6210.2006.00671.x>
- [5] Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002) Evolution of the Social Network of Scientific Collaborations. *Physica A*, **311**, 590-614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)
- [6] Newman, M. (2004) Coauthorship Networks and Patterns of Scientific Collaboration. *PNAS*, **101**, 5200-5205. <https://doi.org/10.1073/pnas.0307545100>
- [7] Frey, B.B., Lohmeier, J.H., Lee, S.W. and Tollefson, N. (2006) Measuring Collaboration among Grant Partners. *American Journal of Evaluation*, **27**, 383-392. <https://doi.org/10.1177/1098214006290356>
- [8] Fujita, M., Okudo, T., Terano, T. and Nagane, H. (2021) Analyzing Two Approaches in Interdisciplinary Research: Individual and Collaborative. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **25**, 301-309. <https://doi.org/10.20965/jaciii.2021.p0301>
- [9] Yamashita, Y., Giang, H.-N. and Oyama, T. (2018) Investigating the Performance of Japan's Competitive Grant Grants-in-Aid for Scientific Research System. *International Journal of Higher Educations*, **7**, 167-184. <https://doi.org/10.5430/ijhe.v7n5p167>
- [10] Corley, E.A., Boardman, P.C. and Bozeman, B. (2006) Design and the Management

of Multi-Institutional Research Collaborations: Theoretical Implications from Two Case Studies. *Research Policy*, **35**, 975-993.

<https://doi.org/10.1016/j.respol.2006.05.003>

- [11] Lee, S. and Bozeman, B. (2005) The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, **35**, 673-702.

<https://doi.org/10.1177/0306312705052359>

- [12] Newman, M. (2018) *Network*. Second Edition, Oxford University Press, Oxford.

<https://doi.org/10.1093/oso/9780198805090.003.0018>

- [13] Clauset, A., Shalizi, C.R. and Newman, M.E.J. (2008) Power-Law Distributions in Empirical Data. *SIAM Review*, **51**, 661-703. <https://doi.org/10.1137/070710111>