論文 / 著書情報
Article / Book Information

題目(和文)	連携・知識構造特性に基づく科学運営のためのデータ計装
Title(English)	Data instrumentation for management of science based on collaboration and knowledge-structural features
著者(和文)	宮下修人
Author(English)	Shuto Miyashita
出典(和文)	学位:博士（技術経営）, 学位授与機関:東京工業大学, 報告番号:甲第12139号, 授与年月日:2021年9月24日, 学位の種別:課程博士, 審査員:仙石 慎太郎,梶川 裕矢,橋本 正洋,笹原 和俊,西山 伸宏
Citation(English)	Degree:Doctor (Management of Technology), Conferring organization: Tokyo Institute of Technology, Report number:甲第12139号, Conferred date:2021/9/24, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

PhD Thesis

Data instrumentation for management of science based on collaboration and knowledge-structural features

Shuto MIYASHITA

Department of Innovation Science

School of Environment and Society

Tokyo Institute of Technology

ABSTRACT

This thesis, entitled ‘data instrumentation for management of science based on collaboration and knowledge-structural features’, aims at the instrumentation to enable data-driven decision-making for the management of scientific research, which consists of five Chapters.

Scientific research, particularly publicly-funded research and development (R&D) for innovations and solving multifaced complex societal issues, is significantly increasing in scale and complexity. However, the literature on strategic and organizational management of scientific research (so-called ‘the management of science’) at universities or public research institutes is still underdeveloped compared to the literature of management of technology that mainly includes industrial R&D. Furthermore, methodology that enables the acquisition of prescriptive knowledge in the management of science need to be developed alongside the descriptive knowledge that has been nurtured based on conventional scientometrics. To address these outstanding issues, it is essential to develop a methodology for instrumentation of structural features in basic and scientific research by focusing on the collaboration and knowledge structures in large-scale R&D programs and their descendant projects. Based on the understanding, a methodology for capturing the real-time status of a scientific research project was developed to support data-driven decision-making by defining activity indicators for scientific research based on network science and conventional scientometrics.

Chapter 1 provides a background of the research including the concept of the management of science, its relation to scientometrics and the significance of instrumentation. The research objective is disseminated in this chapter as to develop a methodology for instrumentation of structural features of research activities by focusing on the collaboration and knowledge structures in large-scale R&D projects centered on science research. Further, a general framework is provided as the micro-meso-macro architecture that corresponds to phenomena at the individual (or team), project and program levels, respectively.

Chapter 2 focuses on the system structure of interdisciplinary R&D program and its descendant projects, based on empirical observation of the FIRST Program and its descendant NanoBio First project. In the program-level analysis, the linkage of research funds and human resources to the outcomes is examined using a structural equation model, which suggest that different mechanisms work in the creation of academic outcomes and intellectual properties. In the project-level analysis, the factors that promote or impede the creation of intellectual properties are explored. The analyses adopt mixed-methods approach combining quantitative network analysis and qualitative interview. It was found that the collaboration between a central university and a startup firm (which has a complementary role) established a knowledge logistics system to realize not only social implementation of research results through joint patent applications but also cultivated a reciprocal mindset and flexible resource management according to the R&D stage.

Chapter 3 explores the interactive dynamics of collaboration and knowledge structures in interdisciplinary R&D projects by visualizing the co-author and the co-word network, respectively,

based on bibliographic analyses of publications. By introducing graph entropy as a measure of the complexity of these weighted network structures, the instrumentation of individual and institutional R&D activities can be visualized. Co-author and co-word network analyses constructed upon bibliographic information from publications in the COINS project confirmed that that these networks have distinctive structural features and different modes of time-course evolution.

Chapter 4 examines a qualitative case study on how changes in the collaboration and knowledge structures at the team and individual levels can influence those at the project level in achieving R&D project goals based on the COINS project. Based on the information from the key participants in the COINS project and the qualitative and time-course observation of the co-author and co-word networks at the individual level, an R&D strategy taken by each observed individual was assigned to one of four quadrants comprised with two independent axes: the production/consumption of core/peripheral knowledge, which was in line with the result of the graph entropy-based collaboration and knowledge structures in Chapter 3 subtheme. Furthermore, based on the observations and analysis of the COINS project, these results were verified to be consistent with the R&D strategy and organization management in the real world.

Finally, in Chapter 5, the insights obtained from the previous chapters were summarized. The instrumentation of the collaboration and knowledge structures in R&D activities at the program, project, team, and individual levels, respectively, enable the observation of organizational and managerial phenomena at each level and the identification of mechanisms implemented across levels, which can lead to a structural understanding of large-scale R&D programs and projects. Methodological developments in the instrumentation of organizational and managerial phenomena in scientific research will contribute to the systematization of management of science and will help practitioners to support evidence-based management and data-driven decision-making.

CONTENT

CHAPTER 1. Introduction

1. 1. Background

Research and development (R&D) in advanced science and technology is expected to lead to new discoveries and inventions that will provide innovative solutions to various social issues. The sophistication of basic and academic scientific research in universities and public research institutes as well as technological development in industry are imperative in pursuing sustainable development, as referred in the sustainable development goals (SDGs). This is because R&D is a source of innovation that impacts the creation of economic value, especially in science-intensive industries such as pharmaceuticals (Cohen, Nelson, and Walsh, 2002; Grigss et al., 2013). Put differently, in recent years, scientific research has not been just a purely fundamental and academic pursuit of results; it is also expected to contribute to industrialization and commercialization (Etzkowitz, 1998; Etzkowitz and Leydesdorff,2000).

Mission-oriented R&D projects, which are funded by government-sponsored programs aimed to solve social problems, are an important part of recent science, technology, and innovation policy. Such projects emphasize transdisciplinarity and a key issue to be addressed is the open innovation through cooperation between academia, industry, and government, that is, industry-university-government collaboration (Bogers, Chesbrough, and Moedas, 2018; Perkmann and Walsh, 2007). Subsequently, basic science research activities in universities and public research institutes are transitioning from small group activities to large-scale projects involving groups with different specialties; simultaneously, encouragement of the integration of different disciplines to create interdisciplinary knowledge broadens the complexity of R&D (Anzai et al., 2012). Furthermore, basic science research should achieve high cost-effectiveness and perform appropriate risk management just like corporate investment activities because it is usually subsidized by public fund. Thus, it is challenging to organize a body of knowledge for the strategic and organizational management of scientific research to realize sustainable R&D against such increasing size and complexity.

Management of technology (MoT) is a well-established discipline that deals with the strategic and organizational management of innovation processes. It largely deals with strategic and organizational management theories for R&D practiced in companies, and has contributed to new product development or intellectual property management. However, the methodology for the management of basic and academic scientific research in universities and public research institutes, the so-called “management of science” is less developed than the established MoT (Anzai et al., 2012; Kodama, Watatani, and Sengoku, 2013). Strategic and organizational management is essential for the operation of modern basic science research, such as the increasingly large and complex interdisciplinary R&D projects. Therefore, this study focuses on how to systematize the management of science, which is the theoretical basis for dealing with the strategic and organizational management of basic science research.

1.2. Management of science and scientometrics

Existing disciplines closely related to management of science include scientometrics, science of science (Fortunato et al., 2018; Wang and Barabasi, 2021), or research on research (RoR) (Ioannidis, 2018). In response to the aforementioned environment change surrounding scientific research, the knowledge for conducting scientific research as a team, or "the science of team science" is increasingly expected due to particularly close area in management of science (Börner et al., 2010; Stokols et al., 2008a, b). Scientometrics focuses on quantitative methods, especially bibliometric methods, to measure and evaluate various activities in science, and also to explain the observed phenomena. On the contrary, management of science is a type of management science, which differs from the natural sciences, such as physics and biology, because it not only describes phenomena but also prescribes means of problem-solving (Van Aken, 2004; Van Aken, 2005). Thus, it focuses on strategic planning and organizational design for scientific research, employing scientometric methodology.

In scientometrics, quantitative indicators are usually used to assess the impact and interdisciplinarity of scientific research. Some of these indicators are calculated based on networks built from bibliographic data of academic publications. In this regard, science can be viewed as an ecosystem—a complex network of ideas, researchers, and papers as elements. These indicators, derived from networks that represent the structure of the system of science, can be described as a summarized quantity, because the networks' adjacency matrix corresponds to the design structure matrix (DSM) (Browning, 2001; Browning, 2016) representing the relationship between the elements of science.

Scientometric methodologies allow us to quantitatively assess the outcomes of scientific research. However, several issues must be overcome before the insights can be applied to the ongoing scientific research practice. Although the utilization of scientometric indicator in decision-making process has advantages in terms of objectivity and quantitative evaluation, overestimating technical aspect of scientometric indicators creates a gap between the reality of scientific research and practice (Rafols, 2019). Furthermore, some of the traditional scientometrics are not suitable for ongoing research activities because of the time frames. For example, citation analysis, which is often used to evaluate the impact of scientific research, has a time lag between publication and citation, making it difficult to apply the results to ongoing processes (Kodama, Watatani, and Sengoku, 2013).

In this regard, the methodological development for management of science is expected, especially for decision-making in ongoing R&D. The present study focuses on a network representing the system structure of scientific research, which can be built by information obtained from bibliographic data of published articles, and explores the means for facilitating real-time decision-making in scientific research.

1.3. Instrumentation for management of science: How to obtain evidence for management

To achieve the objectives of large-scale interdisciplinary R&D programs and projects, selection of desirable actions based on scientific procedures, that is, practice of evidence-based management (Rousseau, 2006), is necessitated. The insights for organizational and managerial phenomena will be extracted from the underlying data and data-driven decision-making will be executed using these insights. However, a means for measurement of the data that quantitatively represent the characteristics of these phenomena is essential to generate relevant insights. Without access to such data, the observation of phenomena and the interpretation of their results become impossible.

The main focus of this thesis is the instrumentation to enable data-driven decision-making in transdisciplinary R&D projects. The history of science is also the history of the development of instruments for scientific research: for example, advances in analytical chemistry were brought about by the progress of measuring instruments (Baird, 1993). Similarly, the motivation of the present study is to provide a methodology for the instrumentation of organizational and managerial phenomena that are observed in the activity of scientific research in interdisciplinary R&D projects, by developing an instrument, which is an application of the methodology for network science and scientometrics. In this thesis, the term “measurement” and “instrumentation” are clearly distinguished in the process of decision-making for project management. The word “measurement” refers to the assignment of concrete numerical values as (quantitative) data that reflect the characteristics of the phenomenon. Meanwhile, the word “instrumentation” is intended to develop techniques or methods for characterization and quantification to enable data measurement of the phenomenon. Furthermore, as instrumentation implies that it allows measurement to achieve automatic control (Trimmer, 1949; Wildhack, 1954), the intention is providing a methodology for prescribing solutions to organizational and managerial problems, which is characteristic of management science. Therefore, measurement is an activity in the explanatory science paradigm, while instrumentation is an activity in the design science paradigm (Van Aken, 2004; Van Aken, 2005) in this thesis.

1.4. Objective

The present study aims to develop a methodology for instrumentation of structural features of research activities by focusing on the structure of collaboration and knowledge in large-scale R&D projects centered on basic science research. Through the effective use of assets available to research institutions, such as scholarly literature databases, ex ante activity indicators which are required in the strategic and organizational management of scientific research to capture real-time status and to execute data-driven decision-making are constructed and their usefulness are examined, rather than ex post performance indicators such as the number of citations.

1.5. Research approach

As the target of the present study is large-scale R&D projects funded by government programs, it is

necessary to decompose the system into appropriate granularity to reduce its complexity for analysis. Here, a micro–meso–macro architecture (Börner et al., 2010) is adopted to focus on the phenomena observed at the micro (individual), meso (team), and macro (project) levels, and the mechanisms in which such phenomena appear have been discussed.

Furthermore, this study assumes that there are two types of system structures in scientific research: collaboration structures and knowledge structures. The collaboration structure represents the cooperation of the actors involved in scientific research, while the knowledge structure represents the linkage between the knowledge possessed by these actors. Considering the purpose of interdisciplinary R&D, improving interdisciplinarity entails the manipulation of knowledge structures. However, intentionally or directly manipulating knowledge structures would be difficult and the usually observed phenomenon involves organizing specific research teams. Here, the relationship between collaboration and knowledge structures is investigated.

1.6. Chapter structure

In this chapter, as a background to establish management of science, the increasing size and complexity of publicly-funded R&D and the similarities or differences between management of science and scientometrics are explained. In the subsequent chapters, the observations and interpretations of the phenomenon at the micro, meso, and macro levels in large-scale and publicly-funded R&D programs and projects have been discussed.

In Chapter 2, focus is directed on the system structure of interdisciplinary R&D programs and projects and the program- and project-level analyses conducted, respectively. In the program-level analysis, the manner in which the funds and human resources subsidized by the programs are linked to the outcomes is statistically modeled. In project-level analysis, the creation of intellectual property, is analyzed from the perspective of knowledge logistics at the individual level. By adopting a mixed-methods approach, which combines quantitative network analysis and qualitative interview analysis, the factors that promote and impede the creation of intellectual property are explored.

In Chapter 3, graph entropy, a measure of the complexity of network structure, is introduced. Collaboration and knowledge structures are represented by co-author and co-word networks of paper, respectively. The complexity of these network structures is measured and cross-sectional and longitudinal analyses are conducted.

In Chapter 4, how the generative mechanism of interventions for research collaboration is implemented from project level to team level and from team level to individual level, as a research strategy to achieve the project objectives is discussed through the arrangement of the configuration among the context, interventions, mechanisms, and outcomes of individual R&D activities in a large-scale project. Using the insights obtained from the previous chapters and the information accumulated through the participant observation, a qualitative analysis of the generative mechanism of interventions to realize research collaboration is conducted.

In the final chapter, Chapter 5, the insights obtained from the previous chapters are summarized and how the insights in management of science can be used for decision-making in scientific research are discussed.

References

Anzai, T., Kusama, R., Kodama, H. & Sengoku, S. (2012). “Holistic observation and monitoring of the impact of interdisciplinary academic research projects: An empirical assessment in Japan”. *Technovation* 32(6):345–357.

Baird, D. (1993). Analytical Chemistry and the ’Big’ Scientific Instrumentation Revolution. *Annals of Science* 50(3):267-290.

Bogers, M., Chesbrough, H. & Moedas C. (2018). “Open innovation: Research, practices, and policies”. *California Management Review* 60(2):5–16.

Börner K., Contractor N., Falk-Krzesinski H. J., Fiore S. M., Hall K. L., Keyton J., Spring B., Stokols D., Trochim W., & Uzzi B. (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine* 2(49):1–6, 49cm24-49cm24

Browning., T. R. (2001). “Applying the Design Structure Matrix to System Decomposition and Integration Problems: A Review and New Directions”. *IEEE Transactions on Engineering Management* 48(3):292–306.

Browning, T. R. (2016). “Design Structure Matrix Extensions and Innovations: A Survey and New Opportunities”. *IEEE Transactions on Engineering Management* 63(1):27–52.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). “Links and Impacts: The Influence of Public Research on Industrial R&D”. *Management Science* 48(1):1–23.

Etzkowitz, H. (1998). “The norms of entrepreneurial science: Cognitive effects of the new university-industry linkages”. *Research Policy* 27(8):823–833.

Etzkowitz, H. & Leydesdorff L. (2000). “The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations”. *Research Policy* 29:109–123.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen A. M., Radicchi F., Sinatra R., Uzzi B., Vespignani A., Waltman L., Wang D., & Barabási A. L. (2018). Science of science. *Science* 359(6379)

Grigss, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). “Sustainable development goals for people and planet”. *Nature* 495(7441):305–307.

Ioannidis, J. P. (2018). “Meta-research: Why research on research matters”. PLoS Biology 16(3):1–6.

Kodama, H., Watatani, K., & Sengoku, S. (2013). “Competency-based assessment of academic interdisciplinary research and implication to university management”. *Research Evaluation* 22(2):93–104.

Perkmann, M., & Walsh, K. (2007). “University-industry relationships and open innovation: Towards a research agenda”. *International Journal of Management Reviews* 9(4):259–280.

Rafols, I. (2019). “S&T indicators in the wild: Contextualization and participation for responsible metrics”. *Research Evaluation* 28(1):7–22.

Rousseau, D. M. (2006). “Is There Such a Thing As “Evidence-based Management”?” *Academy of Management Review* 31(2):256–269.

Stokols D., Hall K. L., Taylor B. K., & Moser R. P. (2008a). The Science of Team Science. Overview of the Field and Introduction to the Supplement. *American Journal of Preventive Medicine* 35(2) S77-S89

Stokols D., Misra S., Moser R. P., Hall K. L., & Taylor B. K. (2008b). The Ecology of Team Science. Understanding Contextual Influences on Transdisciplinary Collaboration. *American Journal of Preventive Medicine* 35(2) S96-S115

Trimmer, J. D. (1949). “Instrumentation and Cybernetics”. The Scientific Monthly 69(5):328–331.

Van Aken, J. E. (2004). “Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules”. *Journal of Management Studies* 41(2):219–246.

Van Aken, J. E. (2005). “Management research as a design science: Articulating the research products of mode 2 knowledge production in management”. *British Journal of Management* 16(1):19–36.7

Wang, D., & Barabasi, A.-L. (2021). “The Science of Science”. Cambridge University Press.

Wildhack, W. A. (1954). “Instrumentation - Revolution in industry, science, and warfare”. *Science* 120(3121).

CHAPTER 2. Structural relationship and intellectual property creation in publicly funded R&D program and project management

2.1. Introduction

The promotion of innovation based on cutting-edge science and technology is a driving force for economic growth and industrial development as well as a powerful tool for solving the various societal challenges facing our world. In addition, the significance and benefits of innovation have recently been emphasized in the sustainable development goals (SDGs) (Grigss et al., 2013; Voegtlin and Scherer, 2017). In particular, the promotion of open innovation (Bogers, Chesbrough, and Moedas, 2018) based on cooperation against academia, industry, and government — the so-called university–industry–government collaboration — is a key agenda in science, technology, and innovation policy, and governmental research and development (R&D) programs are continuously being implemented (Beck et al., 2020; Etzkowitz, 2004). At the same time, the increasing size and complexity of R&D projects pose a managerial issue (Anzai et al., 2012; Lauto and Valentin, 2013; Milojevic,2014; Wuchty, Jones, and Uzzi, 2007). Specifically, construction of a management system to promote smooth cooperation among actors with different specialties and sectors is a pressing issue in R&D management (Katoh et al.,2018; Lauto and Sengoku, 2015; Sauermann et al., 2020).

In innovation-oriented R&D projects, strategic creation of intellectual property, especially patent management, has been emphasized in recent years (D'Este and Patel, 2007; Perkmann et al., 2013). Such projects not only call for publication of research results (which many scientists prefer) but also simultaneous pursuit of patents. Pursuit of both patents and publication must be strategic, as patent applications must be filed prior to publication of the paper (Gans, Murray, and Stern, 2017). Creation of intellectual property is essential for sustainable science, technology, and innovation; transforming R&D outcomes into transferable goods such as patents promotes knowledge dissemination and further innovation in society (Lee and Lin, 2020; Siegel, Veugelers, and Wright, 2007).

In fact, it has been pointed out that intellectual property can influence many SDGs and contribute to solving many societal issues, such as food, health, energy, and climate change (Alos-Simo, Verdu-Jover, and Gomez-Gras, 2020; Bannerman, 2020). Bridging the technologies created by universities and public research institutions for industry is an essential part of sustainable development; therefore, appropriate technology and knowledge transfers (Battistella, De Toni, and Pillon, 2016; Bozeman, 2000; Perkmann et al., 2013) are required. In other words, the creation of intellectual property is based on reciprocal relationships among different sectors and requires strategic and organizational management.

Although creation of intellectual property is strongly encouraged in publicly funded R&D programs, making the R&D outcomes fruitful in terms of industrialization and commercialization is still a challenging problem. Specifically, in the case of state-of-the-art R&D, there is information

asymmetry and mismatch in the sense of purpose between knowledge producers (i.e., academia) and knowledge consumers (i.e., practitioners). To elaborate further, a knowledge logistics system (Wijnhoven, 2003; Wijnhoven, 1998) that can autonomously and continuously supply high-quality knowledge to knowledge consumers must be established.

In the present study, the structure of a large-scale R&D project as a system that includes various actors from different disciplines and sectors, and their subsequent inter-organizational collaborations are analyzed. In particular, the focus is biopharmaceutical field, which relies on direct interactions between academic institutions and industry. Then the key elements that contribute to generating innovative outcomes are explored and the roles and responsibilities of the intermediatries that catalyze for R&D activities are discussed. The present study also aims to develop a comprehensible theory on the mechanisms that promote or impede creation of intellectual property and to develop a sustainable strategic and organizational management system for creation of intellectual property. This study carries out a case study using the mixed-methods approach (Eisenhardt,1989b; Jick, 1979). An exploratory study was conducted to investigate the factors that promote or impede creation of intellectual property by using a combination of quantitative analysis based on collaboration networks built from bibliographic data and qualitative analysis of a semi-structured interview.

2.2. Literature review

In university–industry collaboration, technology transfer remains an important topic (Bozeman, 2000). Delivering research outcomes generated in universities to industry (e.g., through the form of patent licensing) should have the potential to solve societal problems. However, technology transfer is a bidirectional process with feedback rather than a unidirectional move from academia to industry (Meyer-Krahmer and Schmoch, 1998). Previous studies of the innovation process suggest that the innovation process is not linear as represented by “technology push” or “needs pull.” Rather, there are interactions between the fundamental research in research institutions and technology development in industry (Rothwell, 1992).

In particular, in science-intensive sectors such as the biopharmaceutical industry, fundamental research in universities strongly influences industry (Cohen, Nelson, and Walsh, 2002). In addition, citations between patents and scientific publications tend to be strong in science-intensive sectors, leading to strong relationships between the outcomes of academic research and technologies used in industry. Furthermore, users of equipment for scientific experiments (i.e., scientists) contribute to innovation. Indeed, to be precise, the process known as technology transfer should be called “information exchange” to emphasize this bidirectional interaction (Meyer-Krahmer and Schmoch, 1998).

Given this background, the importance of efforts to realize open innovation is widely recognized because it is distinguished from technology transfer in a narrow sense (e.g., the transfer of

intellectual property such as patents) and means that university-industry relationships for research collaboration are built earlier (Perkmann and Walsh, 2007). In the context of open innovation, there are two types of university-industry relationships: research partnerships (joint research and research centers) and research services (contract research and academic consulting) (Perkmann and Walsh, 2007). Research partnerships are formal R&D arrangements of the collaborations among organizations in R&D, mainly realized with political effort. Research services provide knowledge or techniques to external clients (often in industry) from researchers in universities. The Fraunhofer Institute in Germany is a successful model, where R&D activities are conducted based on industry needs (Beise and Stahl, 1999).

In research collaboration, academia and policymakers agree that interdisciplinary research has the potential to influence the established paradigm and to offer a breakthrough for problems unable to be solved only using the knowledge or techniques from traditional monodisciplinary research. Hence, many policies to promote interdisciplinary research, including governmental initiatives, have been launched. However, the term "interdisciplinary" has multiple meanings and ambiguous criteria, and there is no broadly accepted consensus about what interdisciplinary is or is not (Huutoniemi et al., 2010; Stember, 1991). Thus, evaluations of interdisciplinarity have been difficult to perform and have mainly relied on conventional peer review. The problems with this process include the high workload for reviewers and difficulty excluding reviewer bias (Kostoff, 1994). In recent years, bibliometric methods, namely evaluations using quantitative indicators of interdisciplinarity, have been adopted (Stirling, 2007; Wagner et al., 2011). In practice, however, bibliometric indicators using information on publications or patents are unsuitable for managing ongoing R&D projects because of the time lag until they are published (Kodama, Watatani, and Sengoku, 2013). Moreover, Lauto and Sengoku (2015) pointed out issues in terms of human resource development, finding that the reward system for those engaged in interdisciplinary research is insufficient.

In such transdisciplinary research, since stakeholders with various interests are usually involved, principal–agent relationship (Eisenhardt, 1989a; Holmstrom, 1979; Jensen and Meckling, 1976) among them can result conflicts of interest which can cause undesirable agent behaviors, such as adverse selection and moral hazard, for the principals because of information asymmetry (Akerlof, 1970; Bergh et al., 2019; Bergmann and Friedl, 2008; Holmstrom, 1979; Ollier and Thomas, 2013; Sappington, 1991). These problems can inhibit the efficient creation of outcomes and, furthermore, incur additional costs to prevent them. Adverse selection (Akerlof, 1970) is a phenomenon in which the inherently desirable option for the principal is driven out due to the information asymmetry before transactions. In general, there are two types of methods to prevent adverse selection: screening and signaling (Connelly et al., 2011; Delfgaauw and Dur, 2007; Riley, 2001). Screening is that the principals obtain information about the agent's type by presenting them with multiple choices and make them choose according to the agent's type. Signaling is to solve information asymmetry disclosing agent's own type to the principals. On the other hand, moral hazard is a phenomenon in

which the expected performance is undermined by the agent's opportunistic behavior because the principals cannot observe the agent's behavior. The means of preventing moral hazard include monitoring and incentive contracts (Arnott and Stiglitz, 1991; Levin, 2003; Lewis and Bajari, 2014). Monitoring is a means of inhibiting agent's opportunistic behavior by the principal's observation how agent performs. Incentive contracts are a way that the agent to select the desirable behavior for the principal by aligning the interests of the principal and the agent.

2.3. Focus of the case study

As the focus of case study, the governmental R&D support initiative that was implemented in Japan over a five-year period from 2009 to2013 was selected: The Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)[1] and Development of Innovative Diagnostic and Therapeutic Systems Based on Nanobiotechnology, known as NanoBio First[2].

2.3.1 FIRST Program

The FIRST Program is a large-scale governmental R&D program that was implemented in Japan over a five-year period. This program provided large R&D funds for projects organized by 30 Japan's top researchers were selected for sustainable growth in international industrial competitiveness by promoting world-class advanced R&D. This program has five major categories: "life science," "medical engineering," "material science," "mathematics, physics, and information science," and "machinery and system development."

The FIRST Program aims to liberate researchers from troublesome tasks unrelated to research activity. For example, each subsidized project has an Operational Support Institution that coordinates inter-organizational collaborations among participating organizations and assists with the strategic creation of intellectual property from research outcomes. In addition, enabling researchers to spend research funds in the following year provides them with financial freedom during an R&D activity. This research fund has the following classification: Expense A (direct expenses), Expense B (expenses for the Operational Support Institution), and Expense C (indirect expenses). It also provides an additional research fund to accelerate R&D further. Expense Band Expense C must be less than 20% of Expense A

2.3.2 NanoBio First

NanoBio First is one of the projects funded by the FIRST Program and aimed to create a system for realizing medical care from early diagnosis to curative treatment for intractable diseases, such as

[1] https://www.jsps.go.jp/english/e-first/index.html [last visited on 18 October 2020]

[2] http://park.itc.u-tokyo.ac.jp/nanobiof/en/index.html [last visited on 18 October 2020]

cancer, by using nanobio devices that integrate the various functions of medical devices and drugs at the nanoscale. For provisioning such system, a seamless and comprehensive process in which there are various technologies across multiple disciplines is required. Examples include the detection of the warning signs and lesions by early diagnosis using imaging techniques, the combination of surgeries and drug therapies for safe and effective treatment, and the reconstruction of the lesion after treatment.

Because of the interdisciplinarity in the process, this project demands close collaboration with engineering, medicine, pharmacies, and industry. To promote interdisciplinary or cross-field R&D, this project had smaller research domain (called subthemes there) to create a (i) nano-diagnosis system, (ii) nano-drug delivery system, (iii) minimally invasive nano-treatment system, and (iv) nano-reconstruction system. Each of these four subthemes has a leader (called the subtheme leader, including project leader). Subtheme leaders manage the R&D in the domain in which they in charge. They coordinate R&D activities cooperating with universities, large companies, national research institutes, startup firms, and the Japan Science and Technology Agency, which is the Operational Support Institution in this project. In addition, how to implement technology seeds from their results (i.e., research outcomes) to contribute to society was investigated by monitoring global trends in nanotechnology R&D because the commercialization of advanced medical technology affects medical and social systems as a whole.

2.4. Method

Exploratory analyses were conducted according to the following procedure.

1. Data on each project in the FIRST Program for program-level analyses and bibliographic data on papers and patents created in NanoBio First for project-level analyses were collected.
2. Quantitative analysis: Program-level statistical modeling and project-level network analyses were conducted using the collected data.
3. Qualitative analysis: Semi-structured interview with project leader were conducted to deepen the insights obtained from the quantitative analysis.

2.4.1 Data collection for quantitative analyses

Data collection for program-level statistical modeling

The dataset used in the program-level analysis was obtained from the ex post evaluation report published by the Japan Cabinet Office[3].

Data collection for constructing co-author network of papers

First, internal documents that was described list of published papers in NanoBio First during project

[3] Ex post evaluation of Funding Program for World-Leading Innovative R&D on Science and Technology, available from https://www8.cao.go.jp/cstp/sentan/sentan_jigo.html [accessed on 18 October 2020]

period (2009-2013) was obtained. Then, based on the list of papers, an adjacency matrix of a co-author network and a list of information about each node were generated after downloading bibliographic information from Web of Science (Clarivate Analytics).

Data collection for constructing co-inventor network of patents

As s same manner of the papers, list of applied patents in NanoBio First during project period was obtained. On the basis of this internal documents, a list of patents created in NanoBio First was compiled and aggregated patent information by family. PDF files of the patent documents were downloaded from the Japan Platform for Patent Information (J-PlatPat)[4], a database of the National Center for Industrial Property Information and Training (INPIT) in Japan, and Espacenet[5], an online database of the European Patent Office. Information on the inventors and their affiliation was extracted from the documents of each patent family and compiled for each family. Finally, 67 patent families data were used as the dataset for the analysis.

2.4.2 Quantitative program-level statistical modeling and project-level network analysis

Program-level statistical modeling

Using this dataset, descriptive statistics, visualizing the observations and an exploratory factor analysis were conducted. Then observations were classified into two types of variables: resources for R&D projects (input variables) and performances resulting from R&D activities (output variables). Research funds and the number of researchers were the input variables. By contrast, the number of academic publications, academic presentations, patent applications, and patent registrations were set as the output variables. After the exploratory factor analysis, the structural relationships of the FIRST Program were schematically drawn using structural equation modeling. Although this method is somewhat unsuitable for this dataset because of the sample size and skew of distributions, it is useful for theory building from a case study.

Project-level network analysis based on co-author network of papers

A co-author network is a network in which nodes represent authors and edges represent the co-authorship. In this analysis, the co-author network was constructed using the following procedure: (i) information about the authors and their affiliations was extracted from the paper data and a list of author information was generated; (ii) an adjacency matrix was generated by setting edges between authors (i.e., nodes) for all combinations of co-author in each paper; (iii) after anonymizing for privacy protection, the adjacency matrix was transformed into a network object to add information about the affiliation of each node. (iv) To identify the key actors in this co-author network, the node's degree

[4] https://www.j-platpat.inpit.go.jp [last visited on 18 October 2020]

[5] https://worldwide.espacenet.com [last visited on 18 October 2020]

and betweenness centrality was calculated.

Project-level network analysis based on co-inventor network of patents

A co-inventor network is a network in which nodes represent inventors and edges represent the interrelationship of co-inventors. In this analysis, the co-inventor network was constructed using the following procedure: (i) information about the inventors and their affiliations was extracted from the patent family data and a list of inventor information was generated; (ii) an adjacency matrix was generated by setting edges between inventors (i.e., nodes) for all combinations of co-inventors in each patent; (iii) after anonymizing for privacy protection, the adjacency matrix was transformed into a network object to add information about the affiliation of each node. (iv) To identify the key actors in this co-inventor network, the node's degree and betweenness centrality was calculated.

2.4.3 Qualitative Analysis: Semi-structured interview with project leader

The next step of the analysis is a qualitative study aimed at enriching the results of the quantitative analysis with additional in-depth and contextualized insights into the actual understanding of NanoBio First and the reasons that promote patent creation. A 60- minute semi-structured interview with the project leader was conducted and asked following four questions.

- How could NanoBio First achieve early creation and exploitation of intellectual property, such as patent registration and licensing?
- What was expected of collaborations between a university and a startup firm in NanoBio First?
- What kind of actor had the high centrality as extracted from the network analysis?
- Why was another startup firm established after NanoBio First when there already existed one before?

These questions were selected as results of narrowing the scope of the interview through reading of publicly available documents such as post evaluation reports, field research such as preparatory interviews with project leaders and other key actors, and the results of our quantitative network analysis (see the next section), which showed that actors in startups have a high degree centrality and betweenness centrality and that collaboration between universities and startups would play an important role in the creation of intellectual property. This interview was conducted on October 2, 2020.

2.5. Results

2.5.1 Results of program-level statistical modeling

Table 2.1 summarizes the descriptive statistics of observations. To capture the overview of the observations related to the R&D projects subsidized by the FIRST Program, the data on input and output variables were presented using a bar chart. Figures 2.1, 2.2, 2.3, and 2.4 show the research funds for each R&D project, the number of researchers, the number of published papers (mostly in

peer-reviewed journals), and the number of academic presentations, respectively. Compared with published papers, more academic presentations were submitted without peer review. Figure 2.5 shows the number of patent applications and registrations. There were too few patent registrations despite the many applications in most of these projects. However, NanoBio First exhibited a relatively high number of patent registrations, implying a better mechanism for creating intellectual property.

Table 2.1 Descriptive statistics of observations

Observation	Mean	S.D.	Max.	Min.	Med.
Expense A	2815.7	638.04	4366	1500	2752
Expense B	242.7	108.42	617	71	255
Expense C	269.97	93.15	577	144	260.5
Additional expense	322.87	333.23	1195	0	195
Researcher	106.87	68.58	317	25	88.5
Post doctral fellow	9.83	10.28	43	0	6.5
Total paper	154.67	137.56	518	10	118.5
Peer reviewed paper	146.33	133.17	506	10	109.5
Academic presentation	627	533.92	2793	25	528
Peer reviewed academic presentation	239.53	214.41	1072	0	188
Domestic patent application	42.97	35.56	136	2	32.5
Fireign patent application	24.13	29.36	119	0	16
Domestic patent registration	2	3.01	14	0	1
Foreign patent registration	0.8	2.37	12	0	0

Figure 2.1 Research funds for each R&D project.

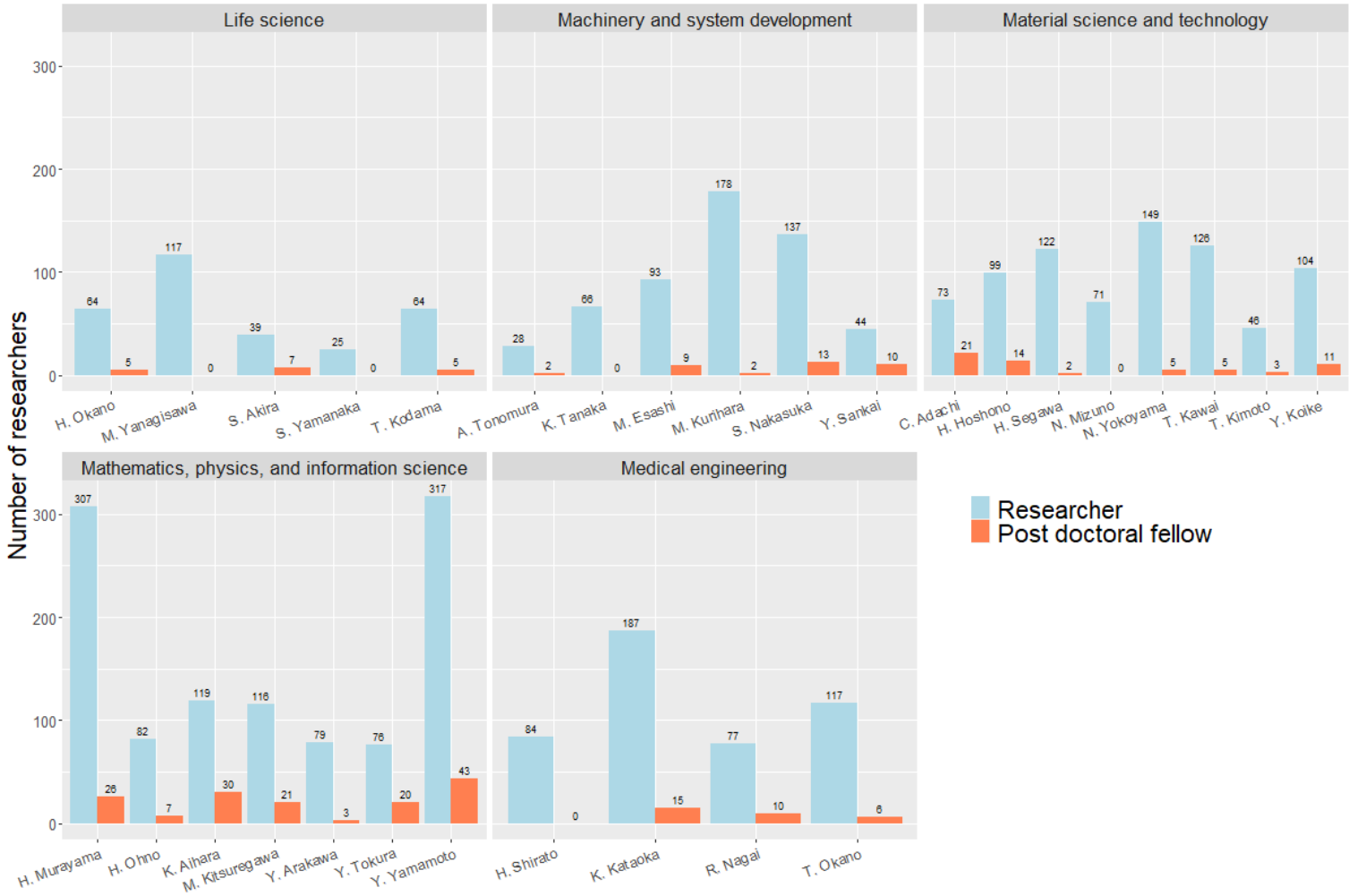

Figure 2.2 Number of researchers.

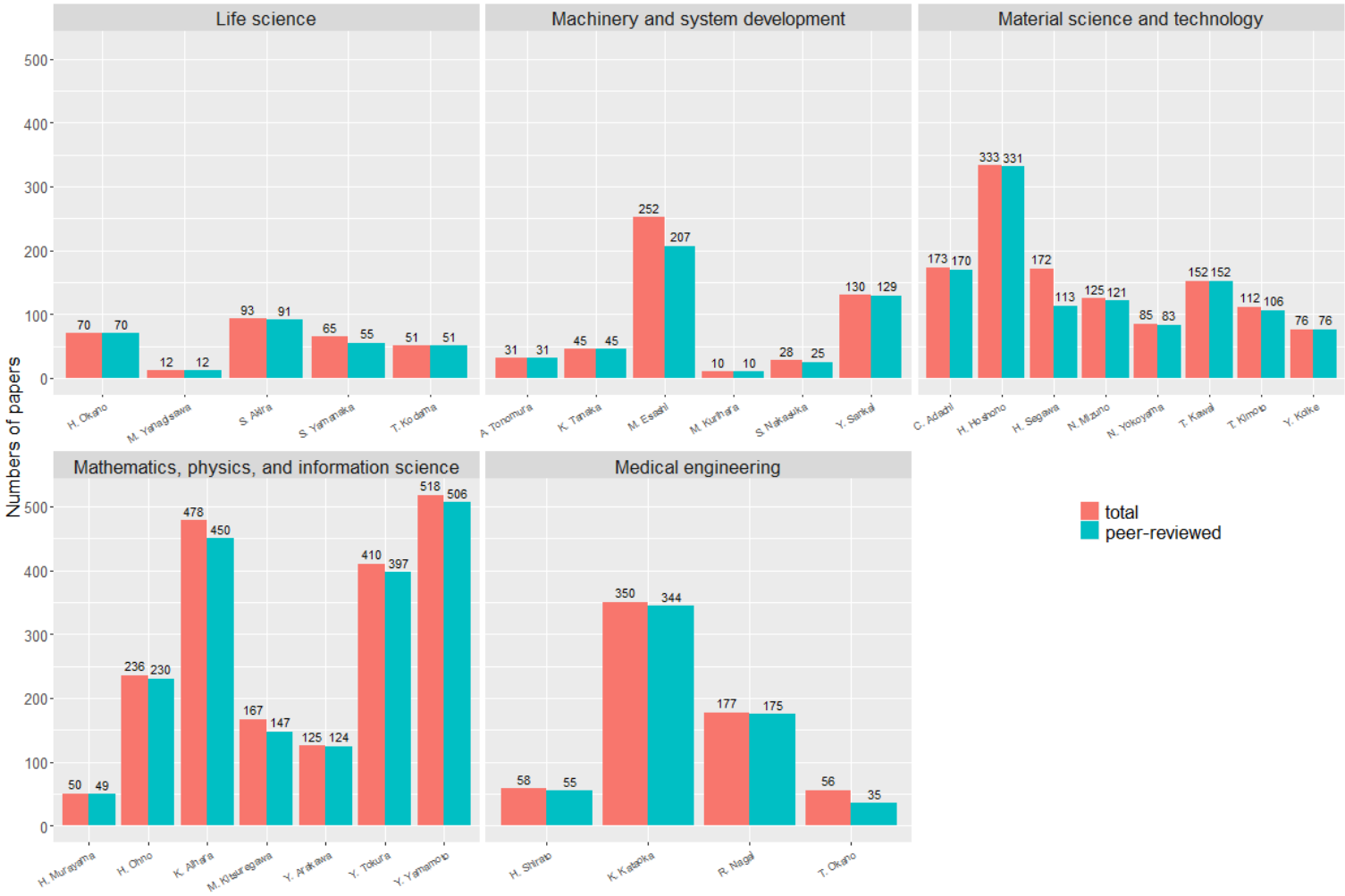

Figure 2.3 Number of published papers.

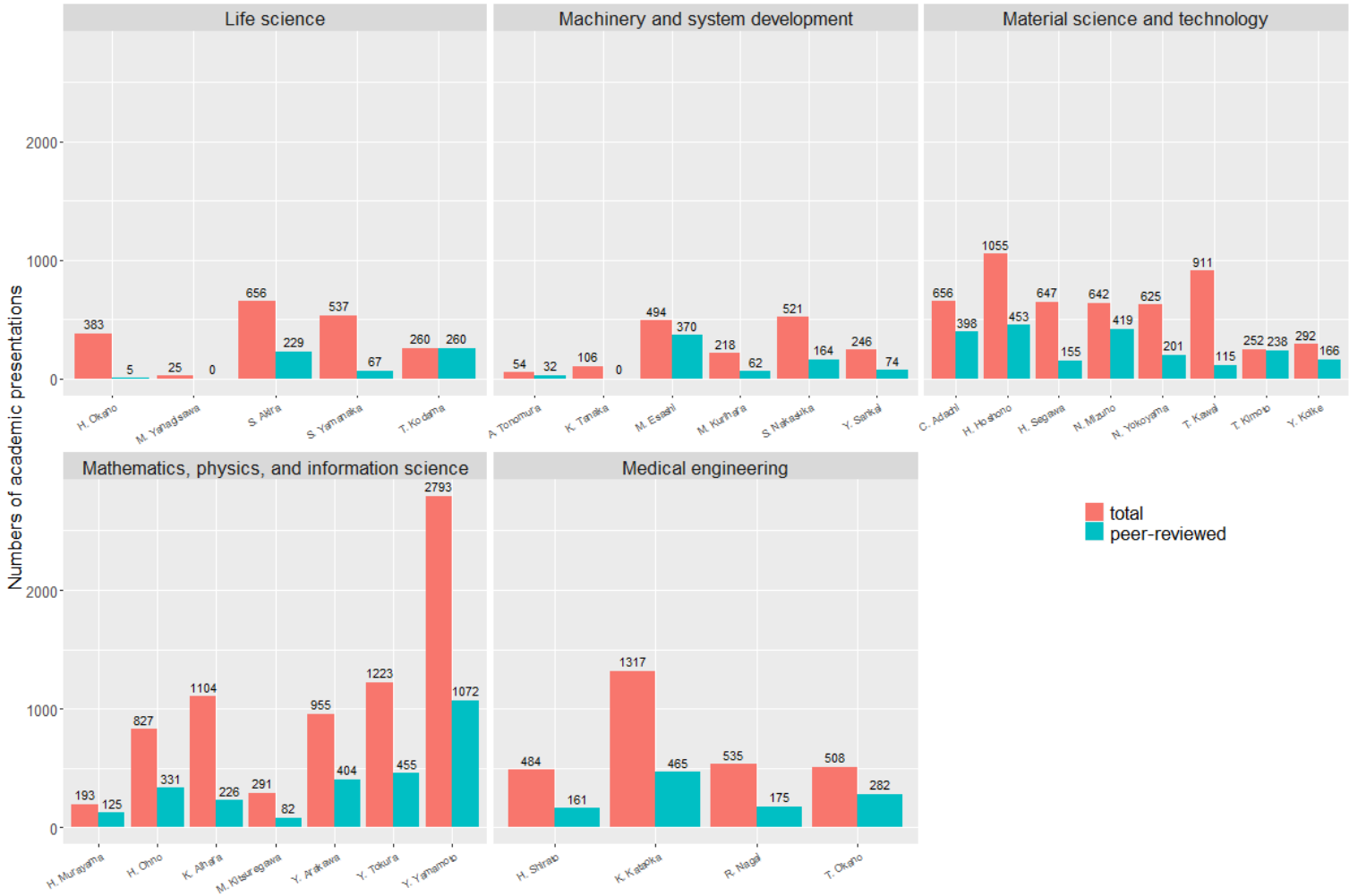

Figure 2.4 Number of submitted academic presentations.

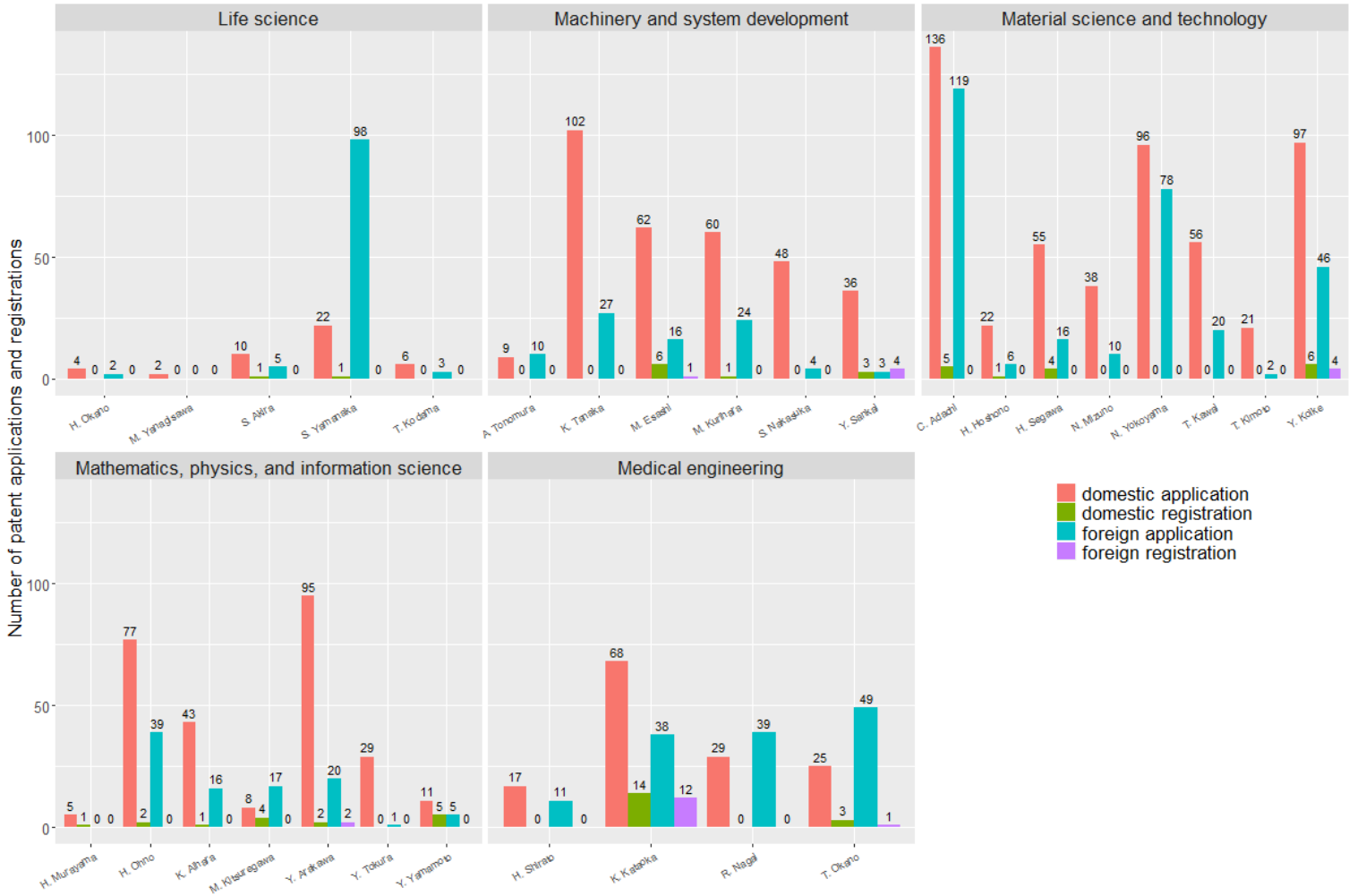

Figure 2.5 Number of patent applications and registrations.

As a result of the factor analysis, factor loadings estimated were shown in Table 2.2. Then, the number of factors was assumed to be 3, generalized least square method was selected as the estimation method and the rotation was set at Oblimin (Oblique rotation) in the analytical process. The structural relationships of the R&D project were modeled using structural equation modeling, assuming a path from the input variables to the output variables through certain factors. The transformation $x' = \log(x + 0.5)$ was applied to all observations (Yamamura, 2002). Figure 2.6 shows the modeled path diagram. Additionally, Table 2.4 shows the values on measures of fit for the model shown in Figure 2.6. The results indicate that the factors related to the number of researchers directly affected academic outcomes and were simultaneously influenced by direct expenses (Expense A). Therefore, the amount of research funds has an indirect influence on academic outcomes. This finding suggests that the more funding is used to employ full-time researchers such as post-doctoral fellows, the greater are the academic outcomes, especially peer-reviewed papers.

Table 2.2 Factor loadings

Output variable	Academic outcomes	Patent registrations	Patent applications
Total paper	0.89	0.065	-0.066
Peer reviewed paper	0.897	0.059	-0.066
Academic presentation	0.984	-0.064	0.037
Peer reviewed academic presentation	0.907	-0.024	0.063
Domestic patent application	0.01	0.09	0.809
Foreign patent application	-0.03	-0.054	0.831
Domestic patent registration	0.128	0.837	0.096
Foreign patent registration	-0.092	0.886	-0.051

Table 2.3 Correlations among the factors

	Academic outcomes	Patent registrations	Patent applications
Academic outcomes	1	0.323	0.005
Patent registrations	0.323	1	0.277
Patent applications	0.005	0.277	1

Table 2.4 Measure of fit for structural equation modeling

	p value	CFI	RMSEA	DOF	GFI	AGFI
Value	0.359	0.982	0.045	50	0.837	0.703

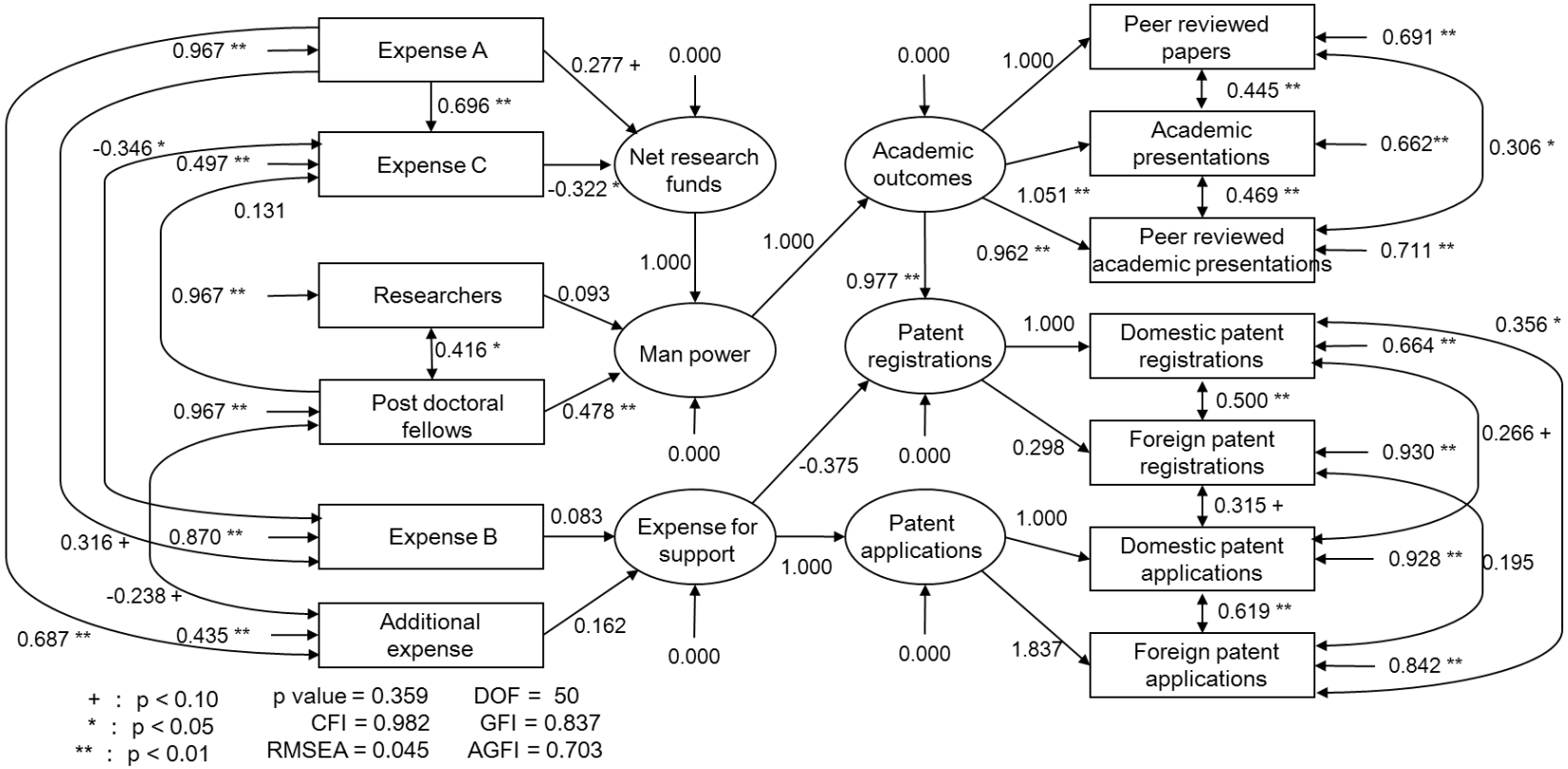

Figure 2.6 Path diagram of the structural equation modeling results.

2.5.2 Result of network analysis based on co-author network

In the co-author network of papers in Figure 2.7, each node represents an author and the color of each node indicates the actor's affiliations: pink represents university, blue represents industry, and yellow represents public research institution.

The co-author network from the bibliographic information of papers created in from NanoBio First had 728 nodes, 5596 edges, and 5 connected components. A large cluster formed around the project leader as close relationship with the project leader has been observed. As can be seen in the co-author network, the edges are concentrated in a few actors, that is, only certain actors have high centrality. In other words, there are both dense and sparse parts in the whole network. Most of the actors with high centrality were researchers at the university, in particular principal investigator (PI).

However, there are some actors belonging to public research institutions and firms who have high centrality. In particular, actors of research institutes for applied R&D, especially the National Cancer Center, tend to have a higher betweenness centrality than degree centrality (data not reported). Hence, clinicians who belonged to the national institute play the role of a bridge between the small clusters. In addition, the industrial actors with relatively high centrality in the co-author network of the paper are confirmed to belong to startup companies (NanoCarrier).

Figure 2.7 Co-author network of papers created in NanoBio First. The color of a node represents its attributes: purple, university; light blue, industry; yellow, public institute; blue, no data.

2.5.3 Result of network analysis based on co-inventor network

The composition of the patent family created in NanoBio First used in the analysis is summarized in the Table 2.4. The “Number of affiliations” column represents the number of patent applicants or patentee per sector. The “Number of created intellectual property” column shows the subtotal number of created patent families for each sector, where the counts overlap in the case of joint applications. The number of patent families filed in a single sector is shown in parentheses. The “Subtotal number of inventors” column represents the subtotal number of inventors listed in the patent document for each sector. The row of “Others” represents inventors whose affiliation could not be identified from the patent documents.

Table 2.8 The composition of the patent family created in NanoBio First used in the analysis

Researcher’s affiliation	Number of affiliations	Number of created intellectual property	Subtotal number of inventors
University	6	41 (15)	83
Industry	10	46 (19)	26
Public	5	12 (3)	17
Other	-	-	7

The co-inventor network constructed with patent data from NanoBio First had 133 nodes, 543 edges, and 10 connected components. This co-inventor network is illustrated in Figure 2.8. One of the structural features of the co-inventor network was that each cluster always included actors belonging to the industrial sector. Although some patents were applied by only a single institute (university or public research institutes), cluster formation in the co-inventor network clearly show collaborative relationships among inventors at NanoBio First associated with the product (i.e., patents); this indicates synthesis of knowledge between researchers in academia and practitioners in industry. This suggests that actors who belong to industrial area deliver knowledge to develop enterprises through R&D outcomes to actors who belong to academia, thereby serving as a complementary knowledge source.

Tables 2.5 and 2.6 list the top 25 actors with high degree centrality and betweenness centrality, respectively. The actors with high degree centrality were researchers from a university and were actively involved in creation of the patents in this project. Furthermore, apart from university researchers, public research actors also had relatively high betweenness centrality. These actors were mainly principal investigators (PIs) at public research institutes and they had higher betweenness centrality than degree centrality because they provide brokerage between members inside and outside of their organizations to collaborate with other organizations. Furthermore, as a common feature, only a few actors had high degree and betweenness centralities. This high centrality concentration of specific actors suggests that this co-inventor network has a core/periphery structure (Borgatti and Everett, 1999; Fontiand Maoret, 2016).

Notably, inventor 33, who belongs to the industrial sector, had high degree and betweenness

centralities compared to other inventors in the industrial sector. Inventor 33 served as a chief science officer (CSO) at that time for the university-originated startup firm founded by the project leader. Considering the intellectual property strategy in R&D projects based on university–industry–government collaboration, since this actor had a high centrality, in other words, was included in the core, the importance of the collaborative relationship between the university and startup firm was indicated. In fact, the startup firm to which this actor belongs filed many patent applications jointly with a university and contributed to creation of intellectual property.

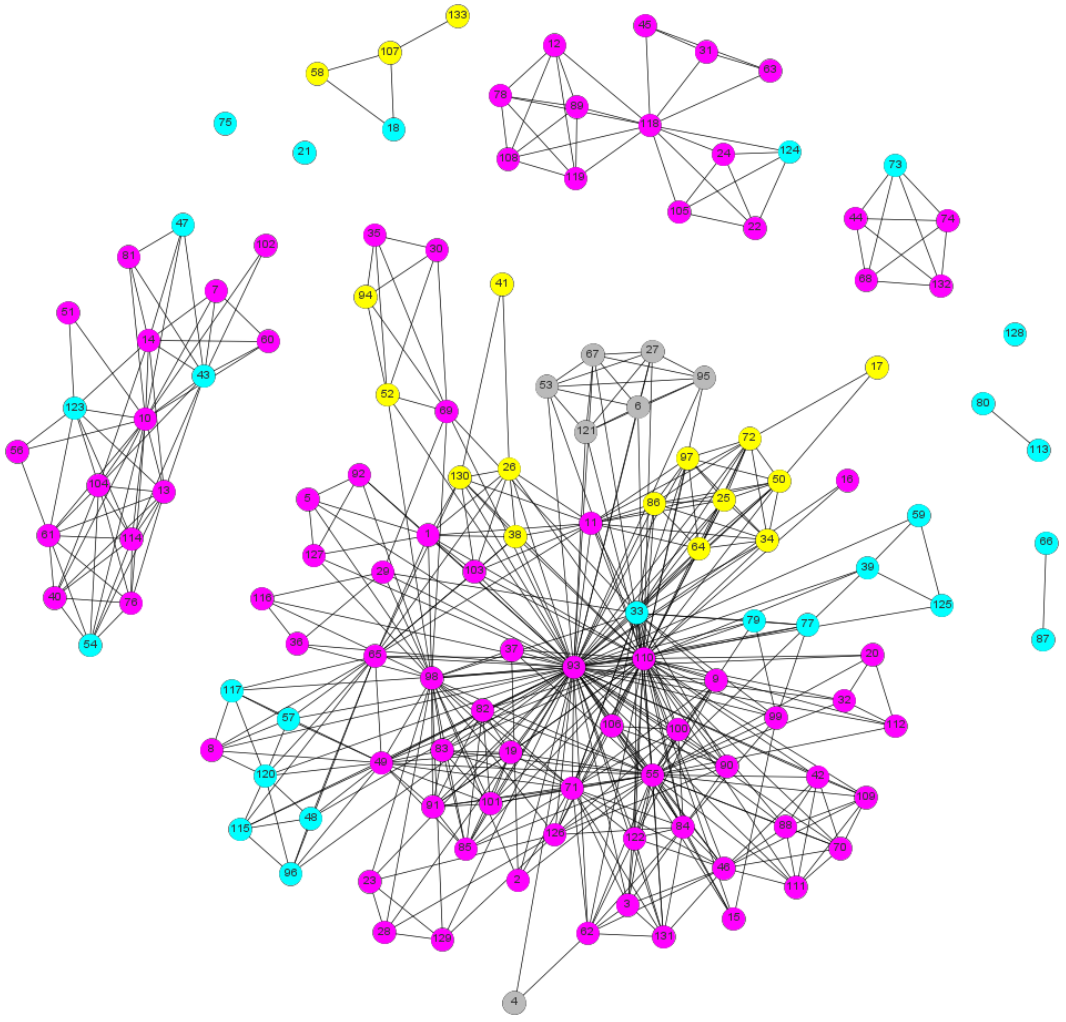

Figure 2.8 Network of co-inventors of patents created in NanoBio First. Each node is color-coded according to the sector to which the inventor belongs: purple, university; light blue, industry; yellow, public institute; gray, no data.

Table 2.5 Top 25 actors with high degree centrality in the co-inventor network of patents

Inventor	Sector	Degree centrality
Inventor 93	University	80
Inventor 110	University	59
Inventor 55	University	37
Inventor 98	University	27
Inventor 33	Industry	25
Inventor 71	University	20
Inventor 49	University	19
Inventor 10	University	17
Inventor 11	University	16
Inventor 65	University	16
Inventor 84	University	16
Inventor 46	University	14
Inventor 1	University	12
Inventor 118	University	12
Inventor 122	University	12
Inventor 126	University	12
Inventor 19	University	11
Inventor 50	Public	11
Inventor 72	Public	11
Inventor 82	University	11
Inventor 83	University	11
Inventor 85	University	11
Inventor 91	University	11
Inventor 100	University	11
Inventor 101	University	11

Table 2.6 Top 25 actors with high betweenness centrality in the co-inventor network of patents

Inventor	Sector	Betweenness centrality
Inventor 93	University	2025.55
Inventor 110	University	629.10
Inventor 55	University	140.94
Inventor 98	University	139.83
Inventor 69	University	135.97
Inventor 52	Public	107.70
Inventor 33	Industry	88.94
Inventor 10	University	57.25
Inventor 118	University	47.00
Inventor 26	Public	41.50
Inventor 50	Public	41.50
Inventor 72	Public	41.50
Inventor 130	Public	41.50
Inventor 49	University	39.00
Inventor 65	University	31.33
Inventor 11	University	20.92
Inventor 71	University	19.41
Inventor 1	University	13.67
Inventor 84	University	10.21
Inventor 43	Industry	10.20
Inventor 126	University	9.73
Inventor 46	University	7.50
Inventor 14	University	6.20
Inventor 123	Industry	4.95
Inventor 62	University	4.00

2.5.2 Results of qualitative semi-structured interview

Table 2.7 is a summary of semi-structured interview's questions and responses to project leader. First, the project leader had an unwavering mindset that outcomes of R&D can be made available for societal benefit through patenting. By sharing this mindset, NanoBio First was able to achieve agile intellectual property management, such as publishing academic papers and applying for patents concurrently. Therefore, it is thought to have led to the early commercialization of the intellectual property that researcher at the NanoBio First closely consulted with a patent attorney hired by using funds from the newly approved budget for R&D support to practice the creating a "strong patent" capable of industrial application even in the middle stages of the R&D.

NanoBio First also adopted a strategy that emphasized collaboration with startups rather than existing large companies, such as licensing the intellectual property created to startups on a priority basis. The collaboration between universities and startups was expected to provide opportunities for researchers in academia to learn experientially the process of social implementation of their research outcomes and to function as a platform for the aforementioned mindset sharing. Although startups have limited resources compared to existing large companies, it seems that the startups were expected as collaborative partners for educational objectives such as the acquisition of tacit knowledge rather than benefits in terms of resource.

Inventor 33 was a CSO at NanoCarrier, which is a university-originated startup firm founded by the project leader before the commencement of NanoBio First project. Inventor 33 had previously worked at a large biotech company where he carried out research related to drug delivery (an expertise of the project leader), and was the director of the corporate laboratory there. Inventor 33 is a specialist with a deep understanding of intellectual property practices and their importance, and was one of the key decision makers on intellectual property serving as a CSO of NanoCarrier. This fact supports the suggestion from the results of the quantitative network analysis that a person who has dual skills in science and business administration plays an important role in the creation of intellectual property.

After the FIRST Program, the project leader established a new startup, AccuRna Inc. (Kawasaki, Japan), to focus resources on an R&D organization specializing in nucleic acid drugs, while an existing startup company (NanoCarrier) was already established. AccuRna entered into a capital and business alliance with NanoCarrier and set up a system to secure and utilize management resources in the early stages and aquired exclusive rights to the intellectual property created in NanoBio First, which gave it a competitive advantage. In addition, the establishment of AccuRna was a great opportunity to attract people with expertise in nucleic acids or investors who are only interested in new technologies. It was expected to involve new stakeholders that NanoCarrier had not been able to approach in the past. In other words, talent was pooled into a new organization specializing in a specific domain, and the accompanying storage of knowledge tied to the person became feasible. The stored knowledge could be transferred from one organization to another through organizational

consolidation, such as a corporate acquisition. AcuRna was finally acquired by NanoCarrier in September 2020, which integrated the human resources of both human resources and the knowledge, such as know-how tied to them and knowledge transfer between organizations was achieved.

Table 2.7. Summary of semi-structured interview's questions and responses to project leader

Question	Observed fact	Implication
Q 1: How could NanoBio First achieve early creation and exploitation of intellectual property, such as patent registration and licensing?	• Conventional patenting scheme is problematic because university researchers seek assistance with patenting from technology licensing organizations (TLOs) only during the final stages of R&D • Researcher at the NanoBio First closely consulted with a patent attorney hired by using funds from the newly approved budget for R&D support to practice the creating a "strong patent" capable of industrial application even in the middle stages of the R&D • Not only research leaders but also practitioners should identify the similarities and differences between publishing and patenting their own R&D outcomes, which are not contradictory	• Shared mindset that was recognized reciprocity between R&D and social implementation contributed to I agile creation of intellectual property
Q 2: What was expected of collaborations between a university and a startup in NanoBio First?	• Through the collaboration with the startups, it was expected that the researchers in academia experientially learn the seriousness of business and the process of implementing their own research results into society • Interacting with existing large companies hardly to provide the opportunity of such a real case study	• The Acquisition of tacit knowledge that can only be obtain through on-the-job-training in real business environment was emphasized, rather than benefits in terms of R&D resources
Q 3: What kind of actor had the high centrality (i.e., Inventor 33 in Figure 2.8) as extracted from the network analysis?	• Inventor 33 was the chief scientific officer of NanoCarrier's at the time and made important decisions regarding intellectual property, who had been the head of a research laboratory at a major biotechnology company prior to joining NanoCarrier	• A person who has dual skills in technological development and business administration played an important role in the creation and dissemination of intellectual property.
Q 4: Why was another startup (i.e., AccuRna) established after NanoBio First when there already existed one before?	• AccuRna was established as an organization to conduct R&D specializing in new modalities (i.e., nucleic acid drugs), which intended to concentrate the management resources • NanoCarrier achieved advancement of an anticancer drug (small molecule drug) under development to the clinical trial stage at that time, but R&D in a new field was undesirable due to the diversification of management resources and a risk of backlash from investors	• A new startup specializing in a new modality/application formed a specific intellectual resource pool that accelerated the product development process

2.6. Discussion

2.6.1 Management at the program level

The results of the program-level analysis suggest different mechanisms in R&D activities between the publication of papers for theoretical advancement and creation of intellectual property such as patent applications and patent registrations. Furthermore, the important finding that patent applications and registrations are affected by different factors implies different mechanisms behind these two processes too. When the FIRST Program is considered to be a system, these implications suggest the existence of three subsystems: (i) a subsystem to promote the creation of academic outcomes, (ii) a subsystem to promote patent applications, and (iii) a subsystem to promote patent registrations. Moreover, a common structure should provide interfaces to coordinate these subsystems. Put simply, an R&D program such as the FIRST Program should provide a holistic system to create synergy between the publications from academic research institutions and intellectual property expected as part of industrialization and commercialization activities.

2.6.2 Core/periphery structure in a co-author network of papers and co-inventor network of patents

In the co-author network and co-inventor network obtained from the analysisof NanoBio First, the strength of the link between nodes, or the relationships between actors, depends on the destination. In other words, each actor has different relationships as collaborators such as a “strong tie” or a “weak tie” (Granovetter, 1973). To better understand the structural feature of NanoBio First, the focus should be strength of the tie (i.e., the closeness of the relationship between actors).

Observing the centrality of each actor in the co-author and co-inventor networks, ties tend to be concentrated on some actors such as the project leader and subtheme leaders. These actors with high centrality are members of NanoBio First and most built close relationships with each other before the start of the project. In each network, some actors aggregated and formed a cluster using these close relationships, while others are loosely connected in clusters. Therefore, a core/periphery structure (Borgatti and Everett, 1999) can be observed in these networks, which indicates that some clusters are composed of actors connected by strong ties, whereas others have weak ties. In the present study, the set of members of an R&D project who share close relationships are defined as the core, and the periphery otherwise.

2.6.3 Building a knowledge logistics system and the contribution of a startup firm

The results of the network analysis suggest that collaboration between actors from both academia and industry can promote creation of intellectual property by synthesizing the complementary knowledge inherent in each sector. Hence, embedding such actors with complementary knowledge into the core in NanoBio First would accelerate knowledge sharing, and forming a transaction-free zone (Baldwin, 2008) would facilitate efficient knowledge exchange across the sectors. Reduced communication costs

within a core would accelerate R&D cycle and efficiently create R&D outcomes (Milojevic, 2014; Zhang et al., 2007).

On the other hand, the results of the interview indicated that researchers were encouraged to acquire practical knowledge (mindset, business diligence, skills, and know-how) on creation of intellectual property as well as on its social implementation through an on-the-job-training (Mincer, 1962) with a startup firm. In other words, NanoBio First practiced cultivation of human resources skilled in both science and business administration (Nakajima, Miyashita, and Sengoku, 2018). To cultivate such talent, NanoBio First particularly highlighted the transformation of the mindset of researchers in academia. In this study, the terms "sequential mindset" and "reciprocal mindset" were defined to distinguish between mindsets that should be broken out and that should be cultivated in such a transformation. A sequential mindset is a mode of thought assuming that technology and knowledge transfer proceeds from basic research to applied research and development in stages, as in a linear model of innovation (Godin, 2006), and that social implementation is not considered until the final stage of research and development. Meanwhile, the reciprocal mindset is a mode of thought that actively seeks to acquire and store tacit knowledge in domains closer to social implementation to achieve concurrent creation of academic and practical outcomes with considerations of the interaction between R&D and social implementation. Cultivating such reciprocal mindset contributed to early acquisition of "strong patents" deserving industrial applications.

To summarize the above discussion, NanoBio First incorporated a process of learning (Wijnhoven, 2003) to acquire explicit and tacit knowledge across different domains or sectors. Rapid knowledge production cycle in the transaction-free zone inside the core enabled development of a knowledge logistics system for agile generation of "strong patents" and early creation of intellectual property. In particular, the involvement of the startup firm NanoCarrier was crucial in the implementation of the above-mentioned system.

2.6.4 Information asymmetry in intellectual property management

The present study focuses on the principal–agent relationship (Eisenhardt,1989a; Holmstrom, 1979; Jensen and Meckling, 1976) caused by information asymmetry between knowledge producers (project body) and knowledge consumers (industry or program body as a mediator) in knowledge logistics, and considers the possibility that conflicts of interest impede efficient creation of intellectual property. In this assumption, the knowledge consumers correspond to the principal and the knowledge producer to the agent and the agent has either sequential or reciprocal mindset, as private information. Figure 2.9 shows a schematic representation of the payoff relationship between knowledge producers and knowledge consumers based on the framework of the principal–agent model. When information asymmetry exists in the process of intellectual property creation, the principal–agent relationship can cause problems, such as adverse selection and moral hazard.

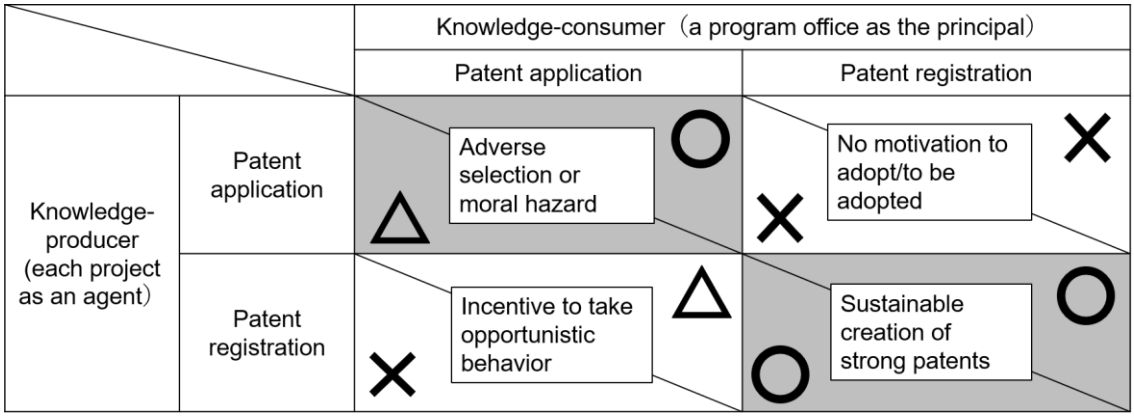

Figure 2.9. Schematic diagram of the payoff relationship in the principal–agent relationship between knowledge producers and knowledge consumers. A circle in the symbol indicates that it is a beneficial choice for the relevant principal or agent's stance, a triangle indicates that it is difficult to determine either, and a cross indicates that it is an unbeneficial one.

Utilization of a startup firm for preventing adverse selection

When principals evaluate whether an agent will appropriately create intellectual property under information asymmetry, they tend to adopt the number of patent applications as evaluation criteria because patent registrations require a certain amount of time, right from patent application to registration, and involve additional costs and uncertainty. Based on this evaluation criteria, agents with a sequential mindset tend to be judged as “superior” to agents with a reciprocal mindset, which would be desirable for creation of intellectual property rather than the because of the greater number of patent applications. Therefore, there exists a problem of adverse selection. In the worst scenario, agents with a sequential mindset drive out agents with a reciprocal mindset, resulting in a flood of low-quality patents that are unsuitable for industrial use.

In general, there are two measures to prevent adverse selection: screening and signaling; (Connelly et al., 2011; Delfgaauw and Dur, 2007; Riley, 2001); in a case like the FIRST Program, the establishment of a startup firm before calling for projects can contribute to both. Specifically, screening corresponds to designing R&D programs with two types of grant plans: a plan in which only projects that have been undertaken through industry-academia collaboration with established startup firms obtain generous budgets and support; and another plan in which projects are requested few for commercialization but limited budgets and support. This would prevent the adoption of projects that have little interest in social implementation of R&D outcomes. Simultaneously, signaling agents’ “seriousness” about innovation through establishment of a startup firm offers a useful way to manifest the value of investment to principals.

Monitoring and incentive mechanisms for avoiding moral hazard

Owing to information asymmetry originating from highly specialized knowledge, another problem,

namely moral hazard, can arise between principals and agents in cutting-edge R&D projects. Even if an agent displayed opportunistic behavior, for instance, submitting an application without requesting patent prosecution, the principal would not know the reasons in detail without understanding the content of individual patents, which require highly specialized knowledge. Hence, agents would have a strong incentive to select such opportunistic behavior because achieving patent registrations requires costs and effort, not only at the time of patent application but also at the time of patent prosecution request.

Enhanced monitoring and incentive contracts would be effective in preventing moral hazard (Arnott and Stiglitz, 1991; Levin, 2003; Lewis and Bajari, 2014). In the FIRST Program, milestone evaluations such as periodic monitoring were conducted, which contributed to prevention of moral hazard to a certain degree. However, program designers should be aware of the limitations of monitoring enhancement offered by the increasing size and complexity of an R&D project. On the other hand, a simple incentive contract ensures that only projects that actively create intellectual property would receive additional R&D funding.

To exercise these measures properly, a mechanism design must satisfy the incentive compatibility for agents. For young researchers, for example, contracts that lead to stable employment, such as offering permanent posts and priority hiring, will be able to provide stronger incentives than financial induction because there is a certain amount of risk involved in engaging in activities for intellectual property creation instead of their academic research (Lauto and Sengoku, 2015).

2.6.5 Limitations and future perspectives

The present study has several limitations. First, the generalizability cannot be guaranteed because the present study performed a theory building from a single case study. Further theoretical development is expected by expanding the applicable object to other R&D programs or projects in the future. This requires that availability of data for comparative analysis among different cases should be ensured. In other words, policy-makers should be aware of institutional design that enables to conduct analysis for so-called evidence-based policy making after the completion of an R&D program or project.

Second, there is the problem of the statistical validity of the used dataset (e.g., sample size) in the analysis at program level. The number of cases used in this study is relatively high (30), but it is not sufficient to guarantee statistical robustness and most data did not satisfy normality. The present study compromised pursuing statistical validity to execute a practical analysis because it was difficult to avoid these problems completely and apply each analytical method strictly.

Third, the analysis in this study was static: the dynamics within the R&D project could not be considered because the information used was a summary at the end of the project. Managing an ongoing R&D project may require a deeper understanding of its dynamic processes.

Fourth, since the present study intended to reveal a mechanism that promotes or impedes creation of intellectual property, the proposed model that explains adverse selection and moral hazard

in the process of creating intellectual property is not mathematically rigorous. A mechanism design for mission-oriented R&D program or project will require a mathematically rigorous approach. The development of mathematical models representing incentive mechanisms can also contribute to empirical analysis (e.g., statistical analysis) for the process of intellectual property creation.

Fifth, the R&D outcomes, such as trade secrets or know-how, that are strategically hidden to prevent imitation by competitors are not considered because the present study mainly analyzed patent information. To consider more strategic decision-making, the proposed model should modify the behavior or types of players. An improved model will facilitate analysis of decision-making in more complex situations. In addition, to deal with the realistic decision-making process, or to avoid the bounded assumptions of players in principal–agent theory (or game theory), introducing a psychological or behavioral economic approach to the analysis would be desirable.

2.7 Conclusion

Present study investigated the structural relationships among the elements of complex and large-scale R&D projects in which various actors participate and interact to promote innovation. In Addition, to reveal the mechanisms that promote or impede creation of intellectual property in a large-scale publicly funded R&D project, a case study using a mixed-methods approach, combining quantitative and qualitative analyses was conducted. Focusing on the FIRST Program, which is a national policy for promoting innovative and world-leading R&D, how the NanoBio First promoted intellectual property creation was examined. The quantitative analysis using the co-author and co-inventor networks had a feature of core/periphery structure. The qualitative analysis of the management policy on intellectual property confirmed that the project leader emphasized cultivating university researchers' mindset towards social implementation through collaboration with a startup firm and to establish systems for flexible resource management according to the R&D stage.

Based on these results, using the framework of principal–agent theory, the decision-making process with information asymmetry between knowledge producers and knowledge consumers was modeled for systematic explanation of the mechanism that promotes or inhibits the creation of intellectual property in large-scale R&D projects involving diverse actors. To handle adverse selection and moral hazard in the process of intellectual property creation, the effectiveness of solutions such that agents should plan the R&D strategy, including collaboration with a startup firm at the time of project adoption, was discussed. Moreover, principals should stipulate incentive contracts that provide additional R&D funding depending on the degree of intellectual property creation.

Finally, the theoretical and practical implications of the present study are described. The theoretical implication is that bridging the gap between two scholarly streams, policy science research and management science research, developed a methodological basis for dealing with program and project management measures in an integrated manner based on a mixed-methods approach. In addition, from the perspective of principal–agent theory, the mechanisms that promote and impede the

creation of intellectual property in large-scale R&D projects were explained. Meanwhile, the practical implication is the contribution to the institutional design of future R&D programs and projects. The utilization of startups and the incentive design for sustainable intellectual property creation were highlighted, especially in transdisciplinary research oriented program and projects.

References

Akerlof, G. A. (1970). "The Market for "Lemons": Quality Uncertainty and the Market Mechanism". *Quarterly Journal of Economics* 84(3):488–500.

Alos-Simo, L., Verdu-Jover, A. J., & Gomez-Gras J. M. (2020). "Knowledge transfer in sustainable contexts: A comparative analysis of periods of financial recession and expansion". *Sustainability* 12(12):4916.

Anzai, T., Kusama, R., Kodama, H., & Sengoku, S. (2012). "Holistic observation and monitoring of the impact of interdisciplinary academic research projects: An empirical assessment in Japan". *Technovation* 32(6):345–357.

Arnott, B. R., & Stiglitz, J. E. (1991). "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out of Peer Monitoring". *The American Economic Review* 81(1):179–190.33

Baldwin, C. Y. (2008). "Where do transactions come from? Modularity, transactions, and the boundaries of firms". *Industrial and Corporate Change* 17(1):155–195.

Bannerman, S. (2020). "The World Intellectual Property Organization and the sustainable development agenda". *Futures* 122:102586.

Battistella, C., De Toni, A. F., & Pillon, R. (2016). "Inter-organisational technology/knowledge transfer: a framework from critical literature review". *Journal of Technology Transfer* 41(5):1195–1234.

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T. M., Conradsen, M. L. Di Marco, D., Distel, A. P., Dobusch, L., Dörler, D., Effert, A., Fecher, B., Filiou, D., Frederiksen, L., Gillier, T., Grimpe, C., Gruber, M., Haeussler, C., Heigl, F., Hoisl, K., Hyslop, K., Kokshagina, O., LaFlamme, M., Lawson, C., Lifshitz-Assaf, H., Lukas, W., Nordberg, M., Norn, MT., Poetz, M., Ponti, M., Pruschak, G., Pujol Priego, L., Radziwon, A., Rafner, J., Romanova, G., Ruser, A., Sauermann, H., Shah, S. K., Sherson, J. F., Suess-Reyes, J., Tucci, C. L., Tuertscher, P., Vedel, JB., Velden, T., Verganti, R., Wareham, J., Wiggins, A., Xu, & Sunny M. (2020). "The Open Innovation in Science research field: a collaborative conceptualisation approach". *Industry and Innovation* 1–50.

Beise, M., & Stahl, H. (1999). "Public research and industrial innovations in Germany". *Research Policy* 28(4):397–422.

Bergh, D. D., Ketchen, D. J., Orlandi, I., Heugens, P. P.M.A.R., & Boyd, B. K., (2019). "Information Asymmetry in Management Research: Past Accomplishments and Future Opportunities". *Journal of Management* 45(1):122–158.

Bergmann, R., & Friedl, G. (2008). “Controlling innovative projects with moral hazard and asymmetric information”. *Research Policy* 37(9):1504–1514.

Bogers, M., Chesbrough, H., & Moedas, C. (2018). “Open innovation: Research, practices, and policies”. *California Management Review* 60(2):5–16.

Borgatti, S. P., & Everett, M. G. (1999). “Models of core/periphery structures”. *Social networks* 21(4):375–395.

Bozeman, B. (2000). “Technology transfer and public policy: a review of research and theory”. *Research Policy* 29(4-5):627–655.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). “Links and Impacts: The Influence of Public Research on Industrial R&D”. *Management Science* 48(1):1–23.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). “Signaling theory: A review and assessment”. *Journal of Management* 37(1):39–67.

Delfgaauw, J., & Dur, R. (2007). “Signaling and screening of workers’ motivation”. *Journal of Economic Behavior and Organization* 62(4):605–624.

D’Este, P., & Patel, P. (2007). “University-industry linkages in the UK: What are the factors underlying the variety of interactions with industry?” *Research Policy* 36(9):1295–1313.

Eisenhardt, K. M. (1989a). “Agency Theory: An Assessment and Review”. *Academy of Management Review* 14(1):57–74.

— (1989b). “Building Theories from Case Study Research”. *Academy of Management Review* 14(4):532–550.

Etzkowitz, H. (2004). “The evolution of the entrepreneurial university”. *International Journal of Technology and Globalisation* 1(1):64–77.

Fonti, F., & Maoret, M. (2016). “The direct and indirect effects of core and peripheral social capital on organizational performance”. *Strategic Management Journal* 37(8):1765–1786.

Gans, J. S., Murray, F. E., & Stern, S. (2017). “Contracting over the disclosure of scientific knowledge: Intellectual property and academic publication”. *Research Policy* 46(4):820–835.

Godin, B. (2006). “The linear model of innovation: The historical construction of an analytical framework”. *Science, Technology, & Human Values* 31(6):639–667.

Granovetter, M. (1973). “The strength of weak ties: A network theory revisited. Sociological theory”. *American Journal of Sociology* 78(6):1360–1380.

Grigss, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). “Sustainable development goals for people and planet”. *Nature* 495(7441):305–307.

Holmstrom, B. (1979). “Moral Hazard and Observability”. *The Bell Journal of Economics* 10(1):74-91.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). “Analyzing interdisciplinarity: Typology and indicators”. *Research Policy* 39(1):79–88.

Jensen, M. C., & Meckling, W. H. (1976). “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure”. *Journal of Financial Economics* 3(4):305–360.

Jick, T. D. (1979). “Mixing Qualitative and Quantitative Methods: Triangulation in Action”. *Administrative Science Quarterly* 24(4):602–611.

Katoh, S., Lauto, G., Anzai, T., & Sengoku, S. (2018). “Identification of factors to promote interdisciplinary research: A trial at COINS”. *2018 Portland International Conference on Management of Engineering and Technology (PICMET).* Portland International Conference on Management of Engineering and Technology, Inc. (PICMET) 1–11.

Kodama, H., Watatani, K., & Sengoku, S. (2013). “Competency-based assessment of academic interdisciplinary research and implication to university management”. *Research Evaluation* 22(2):93–104.

Kostoff, R. N. (1994). “Federal research impact assessment: State-of-the-art”. *Journal of the American Society for Information Science* 45(6):428–440.

Lauto, G., & Sengoku, S. (2015). “Perceived incentives to transdisciplinarity in a Japanese university research center”. *Futures* 65:136–149.

Lauto, G., & Valentin, F. (2013). “How Large-Scale Research Facilities Connect to Global Research”. *Review of Policy Research* 30(4):381–408.

Lee, D., & Lin, K. C. (2020). “How to transform sustainable energy technology into a unicorn start-up: Technology review and case study”. *Sustainability* 12(7):3018.

Levin, J. (2003) “American Economic Association Relational Incentive Contracts”. *The American Economic Review* 93(3):835–857.

Lewis, G., & Bajari, P. (2014). “Moral hazard, incentive contracts, and risk: Evidence from procurement”. *Review of Economic Studies* 81(3):1201–1228.

Meyer-Krahmer, F., & Schmoch, U. (1998). “Science-based technologies: University-industry interactions in four fields”. *Research Policy* 27(8):835–851.

Milojevic, S. (2014). “Principles of scientific research team formation and evolution”. *Proceedings of the National Academy of Sciences.* 111(11):3984–3989.

Mincer, J. (1962). “On-the-Job Training: Costs, Returns, and Some Implications”. *Journal of Political economy* 70(5):50–79.

Nakajima, Y., Miyashita, S., & Sengoku, S. (2018). “Innovative Advanced Technology through University-Industry Collaboration: Role of Venture Capitals, Entrepreneurs and Process Management in Japan.” *Asian Journal of Innovation & Policy* 7(3):564–580.

Ollier, S., & Thomas, L. (2013). “Ex post participation constraint in a principal-agent model with adverse selection and moral hazard”. *Journal of Economic Theory*148(6):2383–2403.

Perkmann, M., & Walsh, K. (2007). “University-industry relationships and open innovation: Towards a research agenda”. *International Journal of Management Reviews* 9(4):259–280.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D’Este, P., Fini, R., Geuna, A.,

Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., & Sobrero, M. (2013). “Academic engagement and commercialisation: A review of the literature on university-industry relations”. *Research Policy* 42(2):423–442.

Riley, J. G. (2001). “Silver signals: Twenty-five years of screening and signaling”. *Journal of Economic Literature* 39(2):432–478.

Rothwell, R. (1992). “Successful industrial innovation: critical factors for the1990s”. *R&D Management* 22(3):221–240.

Sappington, D. E. M. (1991). “Incentives in Principal-Agent Relationships.” *Journal of Economic Perspectives* 5(2):45–66.

Sauermann, H., Vohland, K., Antoniou, V., Balázs, B., Göbel, C., Karatzas, K., Mooney, P., Perelló, J., Ponti, M., Samson, R., & Winter, S. (2020). “Citizen science and sustainability transitions”. *Research Policy* 49(5):103978.

Siegel, D. S., Veugelers, R., & Wright, M. (2007). “Technology transfer offices and commercialization of university intellectual property: performance and policy implications”. *Oxford Review of Economic Policy* 23(4):640–660.

Stember, M. (1991). “Advancing the social sciences through the interdisciplinary enterprise”. *The Social Science Journal* 28(1):1–14.

Stirling, A. (2007). “A general framework for analysing diversity in science, technology and society”. *Journal of the Royal Society Interface* 4(15):707–719.

Voegtlin, C., & Scherer, A. G. (2017). “Responsible Innovation and the Innovation of Responsibility: Governing Sustainable Development in a Globalized World”. *Journal of Business Ethics* 143(2):227–243.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols I., & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* 5(1):14–26.

Wijnhoven, F. (2003). “Operational knowledge management: Identification of knowledge objects, operation methods, and goals and means for the support function”. *Journal of the Operational Research Society* 54(2):194–203.

Wijnhoven, F. (1998). “Knowledge logistics in business contexts: analyzing and diagnosing knowledge sharing by logistics concepts”. *Knowledge and Process Management* 5(3):143–157.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). “The increasing dominance of teams in production of knowledge”. *Science* 316(5827):1036–1039.

Yamamura, K. (2002). “Transformation using (x + 0.5) to stabilize the variance of populations”. *Population Ecology* 41(3):229–234.

Zhang, Z. X., Hempel, P. S., Han, Y. L., & Tjosvold, D. (2007). “Transactive Memory System Links Work Team Characteristics and Performance”. *Journal of Applied Psychology* 92(6):1722–1730.

CHAPTER 3. Complexity of collaboration and knowledge structure in a transdisciplinary R&D project

3.1 Introduction

In recent years, scientific research has been required not only in the pursuit of purely academic results, but also to contribute to the development of the technology oriented toward industrialization and innovation with the creation of economic value (Etzkowitz 1998; Etzkowitz and Leydesdorff 2000; Leydesdorff 2003). In response to the demand that science should contribute to solving societal problems, public research programs are facing the problem of increasing size and complexity. Scientific research in publicly funded research programs has shifted from the activities of small groups to large-scale research and development projects. Interdisciplinary and transdisciplinary research, which are regarded as sources of innovation, have been promoted simultaneously (Anzai et al. 2012).

These changes have strengthened the importance of the "science of team science," which is the knowledge about the strategic and organizational operation of scientific research (Börner et al. 2010; Stokols et al. 2008a, b). Scientometrics is expected to contribute as a means of providing knowledge for such a "science of team science." However, when applying traditional approaches to actual decision-making, the problem that scientometric indices tend to lose value as the materials for appropriate decision-making has been highlighted. This is due to the separation between these and practice in scientific research activities caused by focusing on only the technical aspects and being applied with an excessively reductionist approach (Barre 2019; Ràfols 2019). In other words, the methodology of management of basic and academic scientific research in universities and public research institutions, or "management of science[6]" is underdeveloped as compared to the established technology and innovation management (Anzai et al. 2012; Kodama et al. 2013).

The motivation for the present research is to design an instrument for the management of interdisciplinary research projects that reflect structural characteristics and its complexity. In particular, the present study is interested in the measurement of the structural complexity of ongoing research projects using scientometric methods. Appropriate measurement and control of the project's complexity are useful means of achieving the project goal (Kapsali 2011, 2013). To contribute to decision-making in scientific research, the objects of measurement and methods on complexity should meet the following requirements: (i) real-time measurement of activities is objectively and quantitatively possible (i.e., without time lags such as backward citations)[7]; and, (ii) the method can be universally applicable to decision-making in scientific research programs or projects of different disciplines.

---

[6] The term "management of science" is used as a contrast to technology management, an established discipline that deals with strategic and organizational management theories for research and development in industry.

[7] For example, citation analysis is a method for the impact assessment based on networks representing the citation structure of academic articles. However, such analysis is not suitable for real-time evaluation because there is a time lag until the time a paper is cited.

The present study aims to propose an analytical framework with indices to measure the complexity of collaboration and knowledge structures to support strategic and organizational management of scientific research in interdisciplinary research projects. Collaboration and knowledge structures are represented as networks that can be constructed from bibliometric data of published articles. The proposed framework provides instruments that assist decision-making using quantitative and objective complexity indicators derived from the structure of these networks. To design an instrument that meets these needs, the study focused on the collaboration structure between actors and the knowledge structure between concepts in interdisciplinary research projects. Then, the present study set the research question: How can the complexity of the collaboration and knowledge structure in interdisciplinary research projects be evaluated? In this study, collaboration structure represents the relationship between diverse actors engaged in scientific activities, while knowledge structure represents the relationship between the concepts described in the scientific publications.

Scientific activities can be represented as a complex network of ideas, scholars, and publications (Fortunato et al. 2018; Wang and Schneider 2020). Hence, in scientometrics, networks constructed from bibliographic data are employed for quantitative evaluations of the impact and interdisciplinarity (Ávila-Robinson and Sengoku 2017b; Wagner et al. 2011). Such network-based quantitative evaluation often uses measures that are computed for individual nodes in the network (e.g., degree centrality). Although they reflect the properties of individual components, indicators defined per node are unsuitable for observing the properties of an entire project. In particular, optimizing the whole system to achieve the project goals will be required in mission-oriented projects such as those supported by government programs. In addition, a project management perspective, which is within the scope of this study, requires that the status of such projects be comparable over time.

Based on the above argument, a summary of the methodology for quantifying the complexity of collaboration and knowledge structures is presented below. In the present study, graph entropy (Mackenzie 1966; Sen et al. 2019, 2020; Gomez-Pilar et al. 2018) is introduced as an indicator of the complexity of the network structure. The present study uses the Shannon entropy (Shannon 1948), which is calculated from a probability distribution of the normalized edge weights of a weighted network. Entropy is often used as measure of the diversity of systems, especially interdisciplinarity in the context of scientometrics (Silva et al. 2013; Wang and Schneider 2020; Wagner et al. 2011). An increase in the complexity of the collaboration and knowledge structures in a project can be viewed as diversification of the collaborator and the knowledge to be utilized. This allows us to measure the complexities of the collaboration and knowledge structure by building a co-author network (a network that represents the collaboration structure) and a co-word network (a network that represents the knowledge structure). Furthermore, the current study adopts cross-sectional and longitudinal analyses. A cross-sectional analysis examines the changes in complexity caused by structural division of the project structure into its components, while a longitudinal analysis examines the changes of complexity due to the evolution of the network over time. In other words, our analysis deals with two

types of networks; static and dynamic networks. The combination of these analytical methods and networks is derived from four types of network analysis. The procedures of analysis are as follows:

1. Co-author and co-word networks are constructed using an aggregation of the bibliographic data of publications created during a research project.
2. Qualitative observations on the structural characteristics and dynamic evolution are made by visualizing the networks.
3. Cross-sectional and longitudinal analyses are conducted through the calculations of the complexity measures of the aforementioned networks.

3.2 Case study and hypotheses

The Center of Open Innovation Network for Smart Health (COINS)[8] was selected as a case study. COINS is a research project funded by the Center of Innovation Science and Technology-based Radical Innovation and Entrepreneurship Program (COI STREAM)[9]. The COI STREAM is a large-scale research program funded by Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEXT). The program intensively supports high-risk and challenging interdisciplinary and cross-disciplinary research and development to establish an innovation platform for sustainable innovations that would be difficult for universities to achieve by themselves.

COINS' vision is to create a "smart life care society" in which patients autonomously become healthy in their daily lives without the psychological, physical, and economic burdens of medical care. COINS aims to achieve this vision by creating an "In-body hospital" through the use of smart nanomachines, which are the size as a virus. Such an application of nano-biotechnology will circulate 24 hours a day within the human body to diagnose and treat diseases. The management regime comprises the core organization of this project that is independent of the university, the Kawasaki Institute of Industrial Promotion, a foundation operated by the City of Kawasaki (i.e., local government). Researchers hired by the Institute are working to realize open innovations in the Innovation Center of NanoMedicine (iCONM).

The project, which was subsidized by the COI Program and runs for nine years (FY 2013–2021), is divided into three phases: Phase 1 (FY 2013–2015); Phase 2 (FY2016–2018); and Phase 3 (FY 2019–2021). This study focuses on the outcomes up to Phase 2 of the COI STREAM. To achieve the vision described above, COINS has six research domains, referred to as "subthemes," summarized in Table 1.

[8] https://coins.kawasaki-net.ne.jp/en/index.html [last visited on 12 Aug 2020]
[9] https://www.jst.go.jp/tt/EN/platform/coi.html [last visited on 12 Aug 2020]

Table 1 Subthemes of the COINS

Subtheme	Research topics
Subtheme 1	Nanomachines that can target and eliminate intractable cancer
Subtheme 2	Innovative technology for the treatment of cranial nerve diseases
Subtheme 3	Nanomachines that carry messenger RNA (mRNA) for tissue reconstruction and nano-sized vaccine
Subtheme 4	The system for in-home cancer diagnosis, which requires no blood sampling
Subtheme 5	Theranostic system, which combines nanomachines and medical equipment ultra-minimally invasive treatment
Subtheme 6	Social system for implementation of the results into society

The present study analyzes the relationship between collaboration and knowledge structure in an interdisciplinary research project. Interdisciplinary research projects are expected to create complex insight by integrating diverse knowledge from individual scientists or small groups. According to previous studies in the field of technology management, a structure of a product, which is an outcome of the activities for innovations, interacts with the structure of an organization that develops that product (Henderson and Clark 1990). Such findings enable us to derive a basic hypothesis that changes in collaboration structure can influence changes in knowledge structure in interdisciplinary research projects:

Hypothesis 1: In an interdisciplinary research project, changes in collaboration structure are correlated with changes in knowledge structure.

At a glance, the qualitative description of this hypothesis seems to be trivial. However, the quantitative statement is not trivial because it cannot be revealed the process of changing collaboration or knowledge structures without any instruments of the quantitative measurement for that process. In other words, quantifying how much a structure changes according to another and how structural change proceeds, would provide empirical evidence for management of interdisciplinary research project. In the research on the management of interdisciplinary research project, which is concerned with strategy planning or organizational design for scientific research, observations for hypothesis building (or data measurement) and analyses for hypothesis testing are inseparable. In this sense, our hypothesis is not a priori, but a data-driven hypothesis (Van Helden 2013; Kell and Oliver 2004) from an observational case study.

As a concept that plays a complementary role to the data-driven approach, the present study proposes hypothesis-driven data instrumentation, in which a series of methods for data measurement is implemented to test hypotheses that have not yet been quantitatively represented. The present study introduces the complexity based on network structure as a quantity examining changes of the collaboration and knowledge structures and measures the structural changes observed when some interventions are applied to these networks.

To conduct the multi-level analysis for the structure of scientific research, the present study adopts a micro-meso-macro architecture (Ávila-Robinson and Sengoku 2017a; Börner et al. 2010) and sets three levels: micro-structure (individual actors), meso-structure (subthemes), and macro-structure (project as a whole)[10]. A cross-sectional analysis investigates the changes in complexity by decomposing the structure into three levels. Based on Hypothesis 1, the following hypothesis is derived, regarding the change of the complexity of collaboration and knowledge structure with the decomposition from macro into meso or meso into micro:

Hypothesis 2: When collaboration and knowledge structures are decomposed into their substructures by the same criterion, their reduced complexity is covariant.

As the project progresses over time, although the collaboration among actors and knowledge related to the research topic become more complex, project management to prevent too complex a structure is required at the end of the project. In other words, since research projects essentially tend to be more complex, the external interventions to suppress excessive increase in complexity can affect the success of the project. This argument leads to the following hypothesis regarding the changes in complexity over time:

Hypothesis 3: If there is no external intervention, the complexity of collaboration and knowledge structures in research projects will increase over time.

3.3 Method

3.3.1 The method to compute graph entropy

The present study uses the graph entropy (Mackenzie 1966; Sen et al. 2019, 2020; Gomez-Pilaret al. 2018) calculated using the weights of the edges of the weighted network as a measure, to quantify the structural complexity of the co-author network—which represents the collaboration structure—and the co-word network—which represents the knowledge structure. Given a weighted network $G = (V, E)$, the graph entropy $H(G)$ of the network G is defined by the following equation

$$H(G) = -\sum_{q_{i,j} \neq 0} q_{i,j} \log_2 q_{i,j}$$

where $q_{i,j}$ is the normalized quantity so that the sum of the weights of the edges is 1. If a weight between node i and j in a graph $G = (V, E)$ denotes $e_{ij} \in E$,

[10] Note that micro, meso, or macro levels are a relative concept. In the present study, the structure of the project as a whole is defined as a macro-structure. However, the research programs funding the project, as well as the science and technology innovation policies, correspond more to a macro-level structure.

$$q_{i,j} = \begin{cases} \dfrac{e_{ij}}{\sum(e_{ij} \in E)} & (i \neq j) \\ 0 & (i = j) \end{cases}$$

As is clear from above equation, the graph entropy $H(G)$ of the network G is the Shannon entropy when the distribution of edge weights normalized so that their sum is 1 is considered as a probability distribution.

The graph entropy is a quantity about the complexity of the whole network. If the network is large, the interest for analysis is often only a part of the whole network. Furthermore, it is useful to be able to define an index that measures the local complexity of the network, that is, a graph entropy for a sub-network. The sub-graph entropy of a sub-network G', which is sub-network of network G, is defined as follows:

$$H(G') = -\sum_{q'_{k,l} \neq 0} q'_{k,l} \log_2 q'_{k,l}$$

where $q'_{i,j}$ is a re-normalized quantity of the edge weights of the sub-network G'. Since the edges of the sub-network G' consist only of parts of the edges of the original network G and the sum of these is not equal to 1, the edge weights need to be re-normalized when calculating the sub-graph entropy. If a weight between node k and l in a sub-graph $G' = (V', E')$ denotes $e'_{kl} \in E'$,

$$q'_{k,l} = \begin{cases} \dfrac{e'_{kl}}{\sum(e'_{kl} \in E')} & (k \neq l) \\ 0 & (k = l) \end{cases}$$

The graph entropy $H(G)$ has a nature that its value depends on the number of edges in the network G. It can intuitively understand that the complexity of a network with a large number of edges is likely to be greater than in a network with a small number of edges. Therefore, when comparing the graph entropy between networks with different numbers of edges, conversion of the graph entropy into a quantity that is independent of the number of edges is required. Fortunately, the graph entropy can be normalized by the theoretically calculatable maximum value to limit within the range from 0 to 1. When the weights of the edges in a weighted network G are uniformly distributed and the all edges have equal weights (i.e., equivalent to an unweighted network), the value of graph entropy is maximized. Then, the maximum value of the graph entropy H_{max} can be calculated using the number of edges N_{edge} in the network as follows:

$$H_{max} = -\frac{1}{N_{edge}} \sum \log_2 \frac{1}{N_{edge}} = \log_2 N_{edge}$$

Hence, a network G's the normalized graph entropy $H_{normalized}(G)$ can becalculated through dividing the graph entropy $H(G)$ by its maximum value as follows:

$$H_{normalized}(G) = \frac{H(G)}{H_{max}} = \frac{H(G)}{\log_2 N_{edge}}$$

These equations lead to an important nature of graph entropy: the graph entropy $H(G)$ of a network G

is expressed in the form of the product of the logarithm of the number of edges N_{edge} in the network and the normalized graph entropy $H_{normalized}(G)$.

$$H(G) = \log_2 N_{edge} \times H_{normalized}(G)$$

This fact means that the graph entropy can be decomposed into two parts: one part is determined by the number of edges in the network, which is independent of the edge weights; and the another one is determined by the shape of the distribution of the edge weights. In other words, when a change in graph entropy is observed, such decomposition into components enable us to investigate the reason for the change.

3.3.2 Data collection

Based on internal documents from COINS, academic outcomes published in the period up to Phase 2 (2013–2019) were listed and bibliographic data on publications were collected from the Web of Science, a scholarly literature database provided by Clarivate Analytics. The bibliographic information for "Article," "Review," and "Proceedings Paper" were used for the analysis. Furthermore, only publications from subthemes 1 to 5 dealing with research topics related to nano-biotechnology were set for the analysis; that is, publications from subtheme 6, which dealt with the social sciences, were excluded. Finally, the bibliographic information of 281 publications were used in the analysis.

3.3.3 Software

To build the co-author and co-word networks, the present study used Python for data processing and calculations of the graph entropy and Gephi to visualize these networks. For the layout of nodes in the network, the Fruchterman-Reingold algorithm (Fruchterman and Reingold 1991) embedded in Gephi was used. For more details, see the Appendix A.

3.3.4 Analytical procedures

This study used the following procedures:

1. To build both static and dynamic networks, the following datasets were constructed from the collected bibliographic information: an aggregated dataset of bibliographic information, a corpus of titles and abstracts by publication, panel data of summarized lists of institutions and publications by author, panel data of adjacency matrices of co-author networks, and panel data of adjacency matrices of co-word networks.
2. The static co-author and co-word networks of COINS as a whole were constructed by converting aggregated panel data of the adjacency matrices for all intervals (2013–2019) into network objects.
3. Qualitative observations on the structural features and time evolution of these networks were made from the static network diagram and the dynamic network animations of the networks.
4. Cross-sectional analysis: For static co-author and co-word networks representing the macro-structure (whole project), the meso-structure (per subtheme), and the micro-structure (per

individual actors), the graph entropy was computed to investigate the changes of the structural complexity caused by structural decomposition in COINS. The co-author/word meso/micro-networks was defined as a sub-network structure of the original co-author/word macro-network. A subset of nodes in a particular meso/micro-network was defined as all the listed authors or identified words in the group of publications that belongs to a particular subtheme or an individual, respectively. Then, each co-author/word meso/micro-network was generated based on each of these subsets. For example, in the case of constructing an author's micro (individual) level network, all the publications by this author were identified, and subsequently, a co-author network was made of all the nodes in these publications. Similarly, a co-word network of an author was made based on the words extracted from titles and abstracts in the author's publications.

5. Longitudinal analysis: For the dynamic co-author and co-word networks, graph entropy and normalized graph entropy by year was computed and the changes in the structural complexity of these networks over time were investigated by depicting time-course plots, whose horizontal axis is year, vertical axis is the graph entropy or the normalized graph entropy of the co-author or co-word networks, for co-author and co-word networks, respectively, and a trajectory, whose horizontal axis is the graph entropy or normalized graph entropy of the co-author network, vertical axis is the graph entropy or normalized graph entropy of the co-word network.
6. A semi-structured interview was conducted with the research leader to interpret the results of the analysis and validate the data-driven hypotheses derived from it.

For details of these processes, see the Appendix A.

3.4 Results

3.4.1 Overview of the co-author and co-word networks

Three types of co-author and co-word networks were constructed, representing the structure of the project as whole (macro-structure), per the subtheme (meso-structure), and per actor (micro-structure). The network constructed for the project as a whole network was referred as a macro-network, per subtheme as the meso-network, and per actor as a micro-network.

Figure 3.1 shows the co-author macro-network and Figure 3.2 shows the co-word macro-network. The constructed co-author macro-network had 1025 nodes and 8386 edges, and the co-word macro-network had 7432 nodes and 22,013 edges. The nodes in the co-author macro-network shown in Fig. 3.1 represent authors, while the colors of the nodes represent their affiliations by sector: purple for universities, red for the core organization (iCONM), light blue for industry, yellow for public institutions, and blue for no data. The co-author macro-network comprises several clusters that include university researchers, as well as actors from public institutions and industry. This confirms the characteristics of transdisciplinary research projects where various stakeholders collaborate to address common issues. The static network diagram of the co-author and co-word macro-networks showed the following structural features: the former was globally connected except

one isolated sub-network whereas the latter had one large network of connected components as well as numerous isolated sub-networks. Nodes of the co-author macro-network tended to form multiple modules, whereas those of the co-word macro-network spread outward from the core in cluster-like branches.

The co-author and co-word meso-network were constructed by extracting sub-networks from the original network (i.e., macro-network) based on subtheme attributes assigned to each node. Figures 3.3-3.7 and 3.8-3.12 show the co-author and co-word meso-networks, respectively. The structural features of networks were as follows: those for subthemes 1, 2, 3, and 5 had both common and specific clusters for each subtheme. The network for subtheme 4 had a distributed network with many small clusters and without a core cluster, as observed in the other subthemes.

Figures 3.13 and 3.14 show the time evolution of the co-author and co-word macro-networks, respectively. An observation of the time evolution of dynamic co-author and co-word networks revealed different characteristics: the co-author network was gradually connected through the emergence of brokerages between clusters, after nodes initially formed clusters and agglomerated. In the co-word network, new concepts evolved out of the core cluster and small clusters emerge simultaneously. Furthermore, the emergence of new concepts that connected branches and existing small clusters propagated the knowledge in other fields.

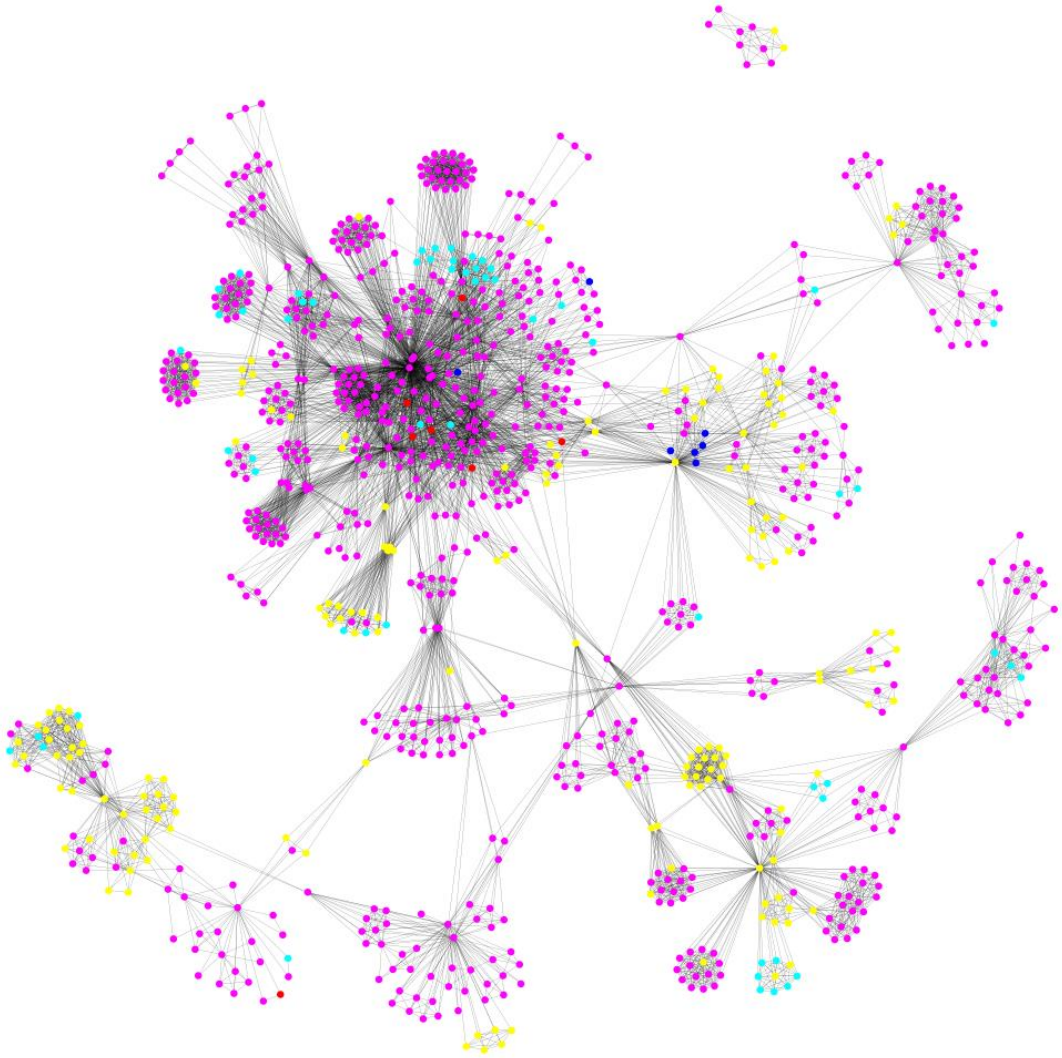

Figure 3.1 Network structure of co-authorship of the publications from the whole COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

Figure 3.2 Network structure of co-word relations of the publications from the whole COINS project.

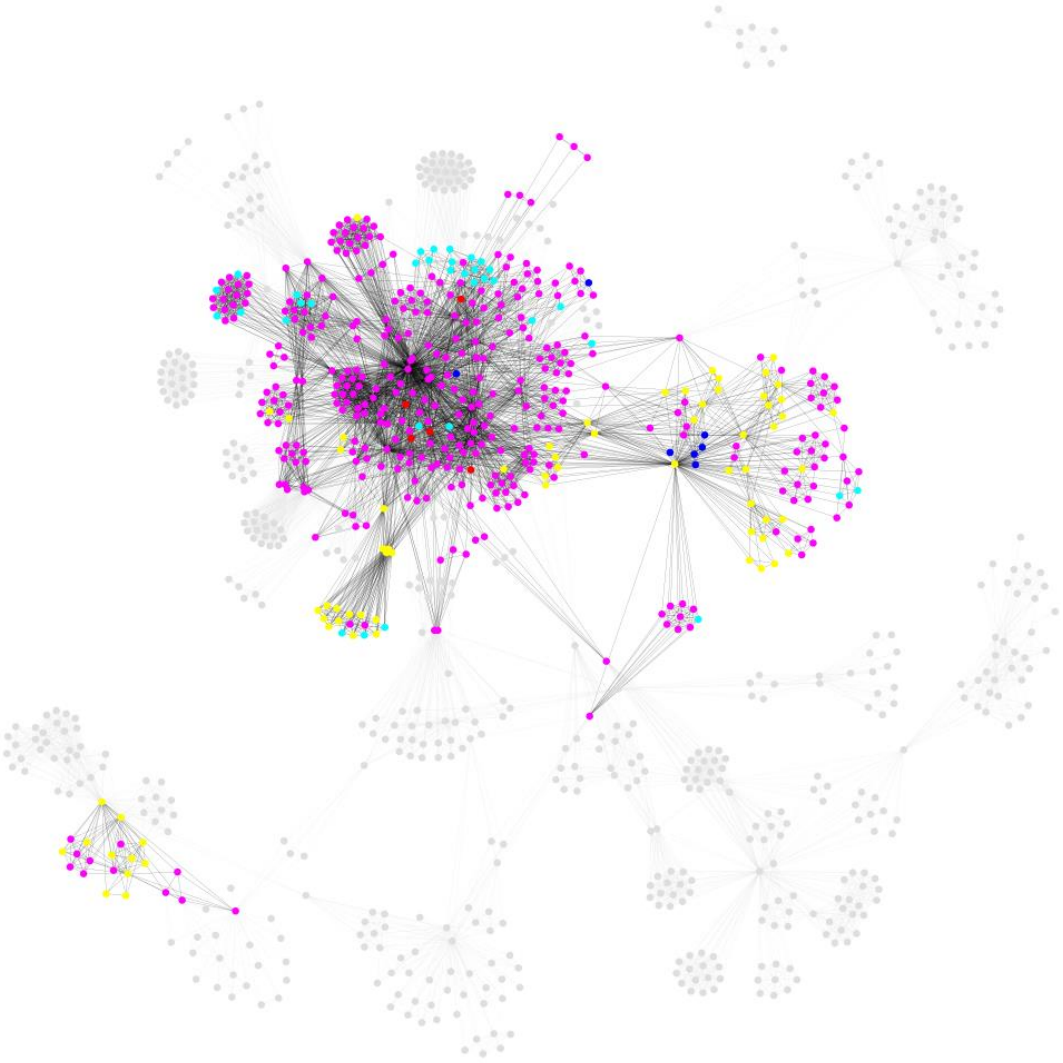

Figure 3.3 Co-author meso-network of the publications from subthemes 1 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

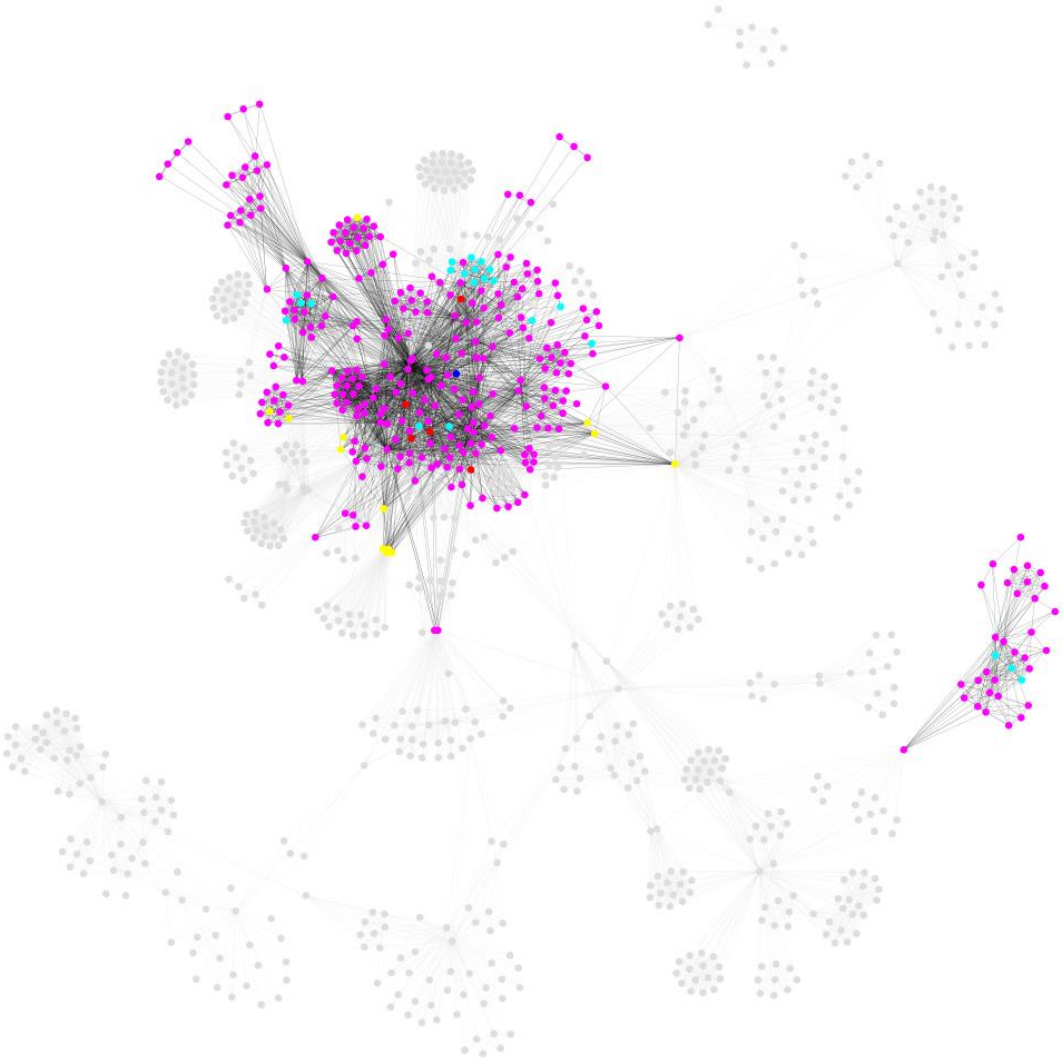

Figure 3.4 Co-author meso-network of the publications from subthemes 2 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

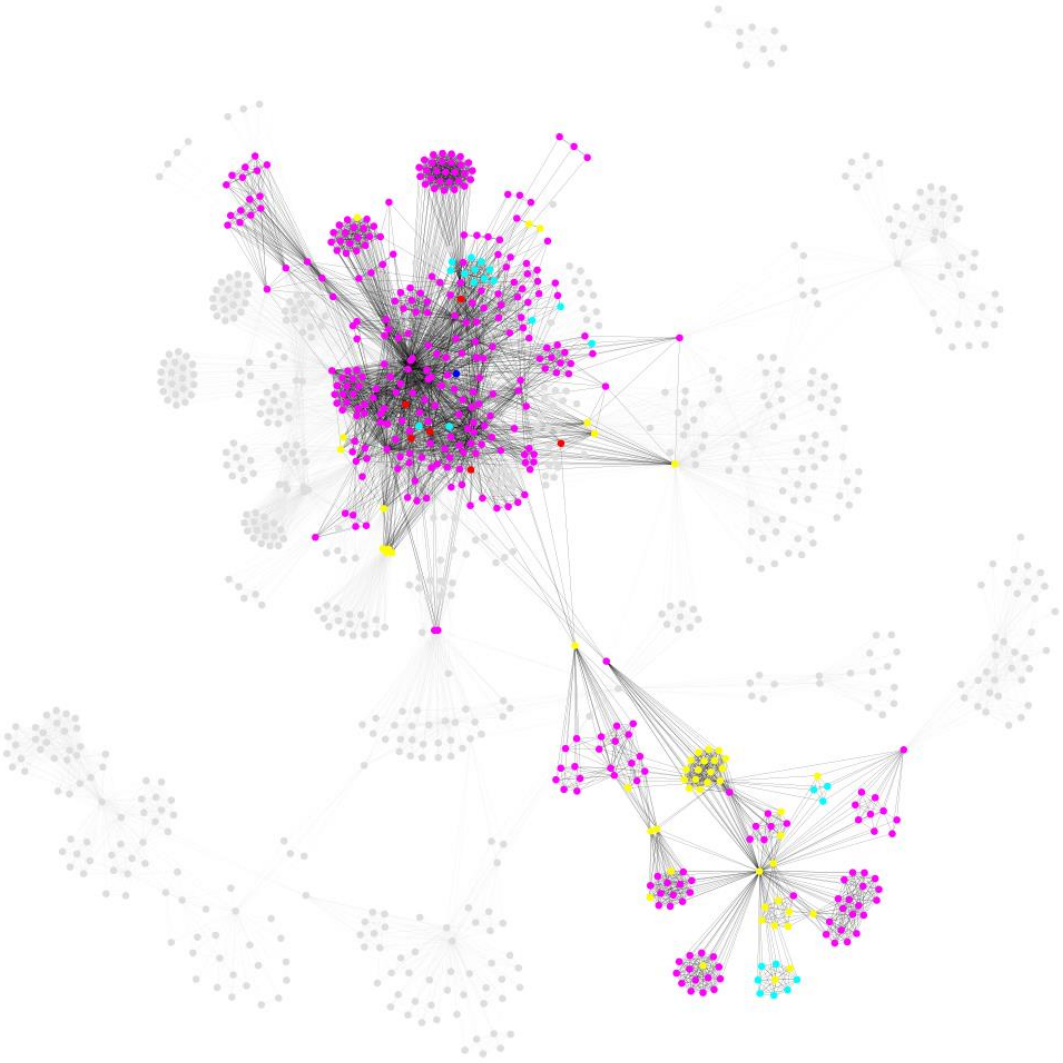

Figure 3.5 Co-author meso-network of the publications from subthemes 3 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

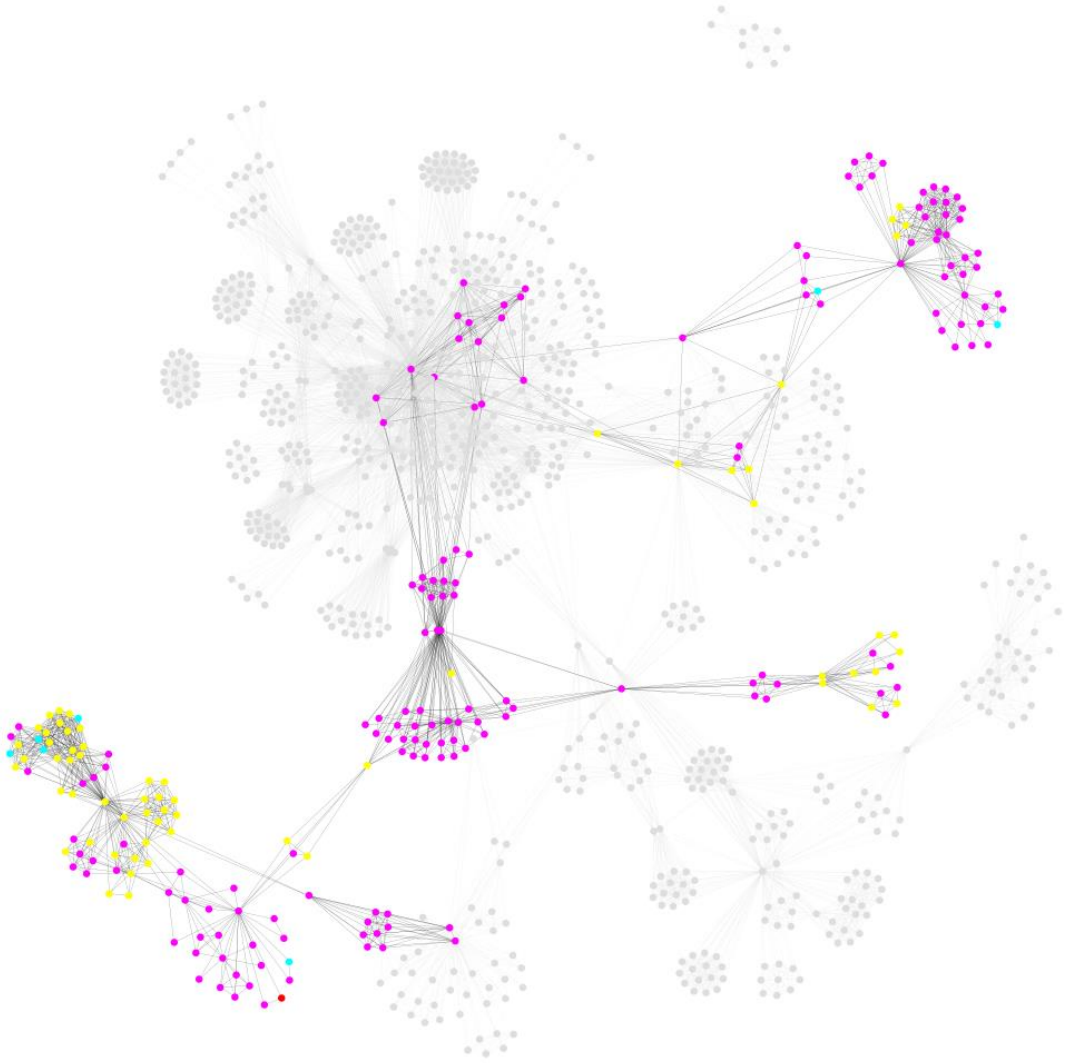

Figure3.6 Co-author meso-network of the publications from subthemes 4 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

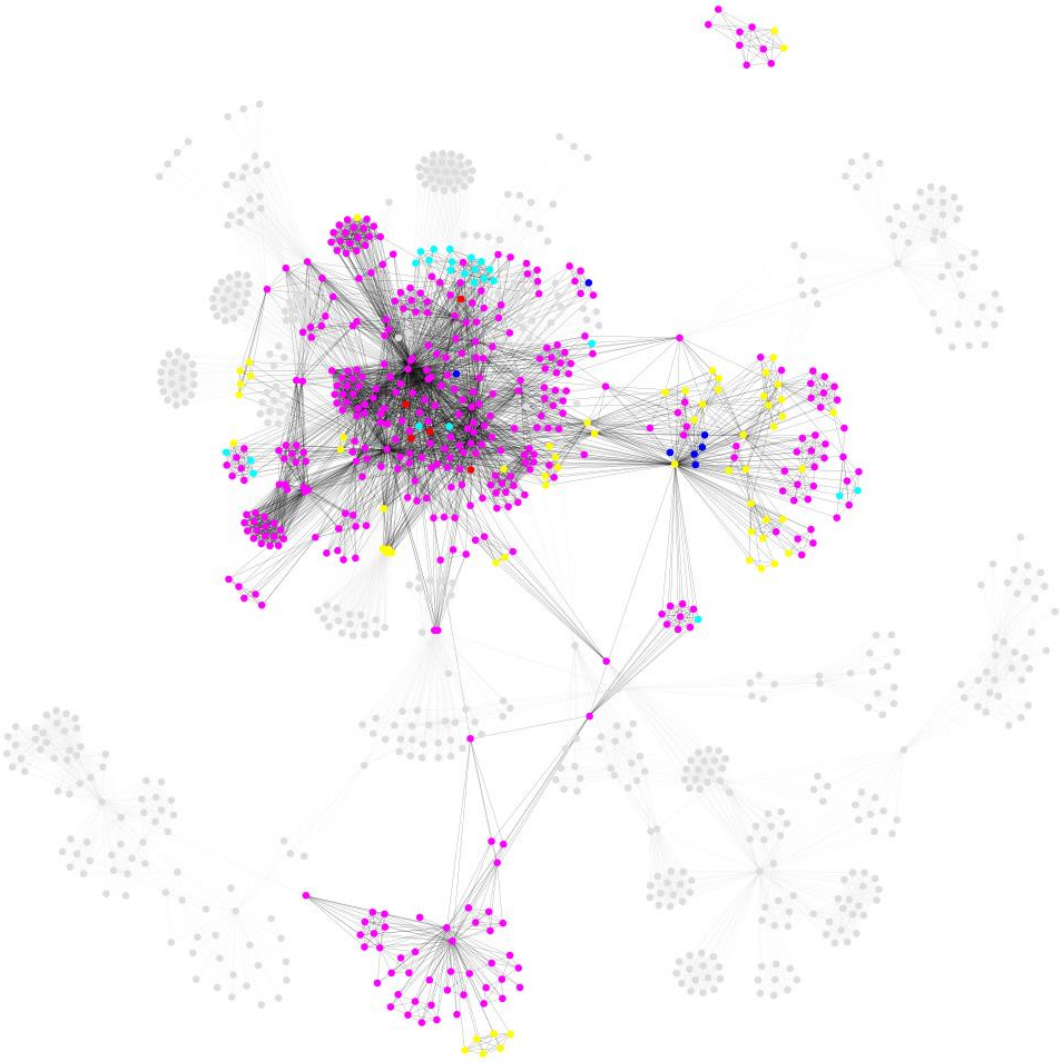

Figure 3.7 Co-author meso-network of the publications from subthemes 5 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

Figure 3.8 Co-word meso-network of the publications from subthemes 1 of COINS project.

Figure 3.9 Co-word meso-network of the publications from subthemes 2 of COINS project.

Figure 3.10 Co-word meso-network of the publications from subthemes 3 of COINS project.

Figure 3.11 Co-word meso-network of the publications from subthemes 4 of COINS project.

Figure 3.12 Co-word meso-network of the publications from subthemes 5 of COINS project.

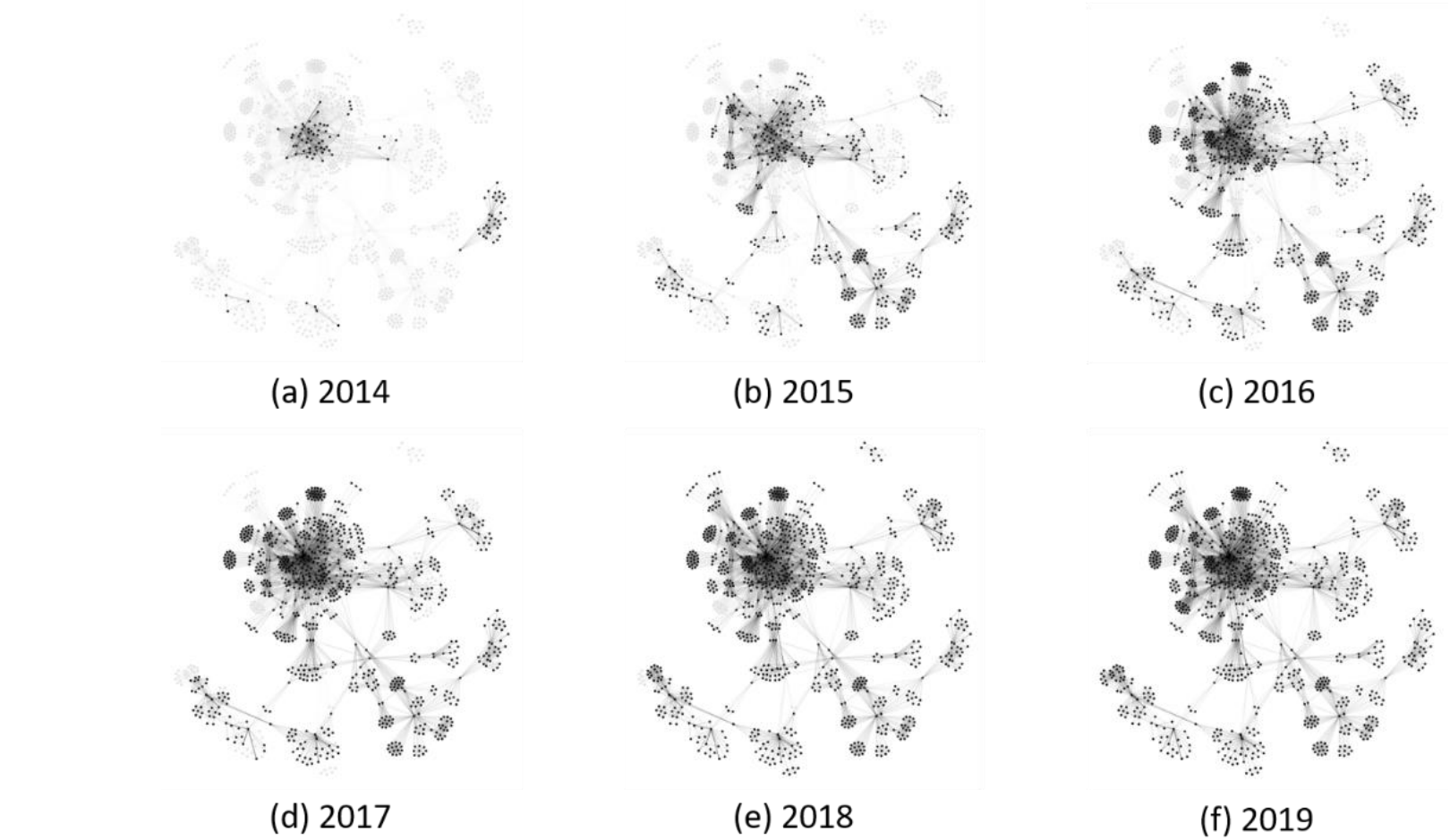

Figure 3.13 Time evolution of the co-author macro-network. The upper row represents the period of Phase 1 (FY 2013-2015) and the lower row represents the period of Phase 2 (FY 2016-2018) in the COI Program. The nodes and edges in this network cumulatively increase over time.

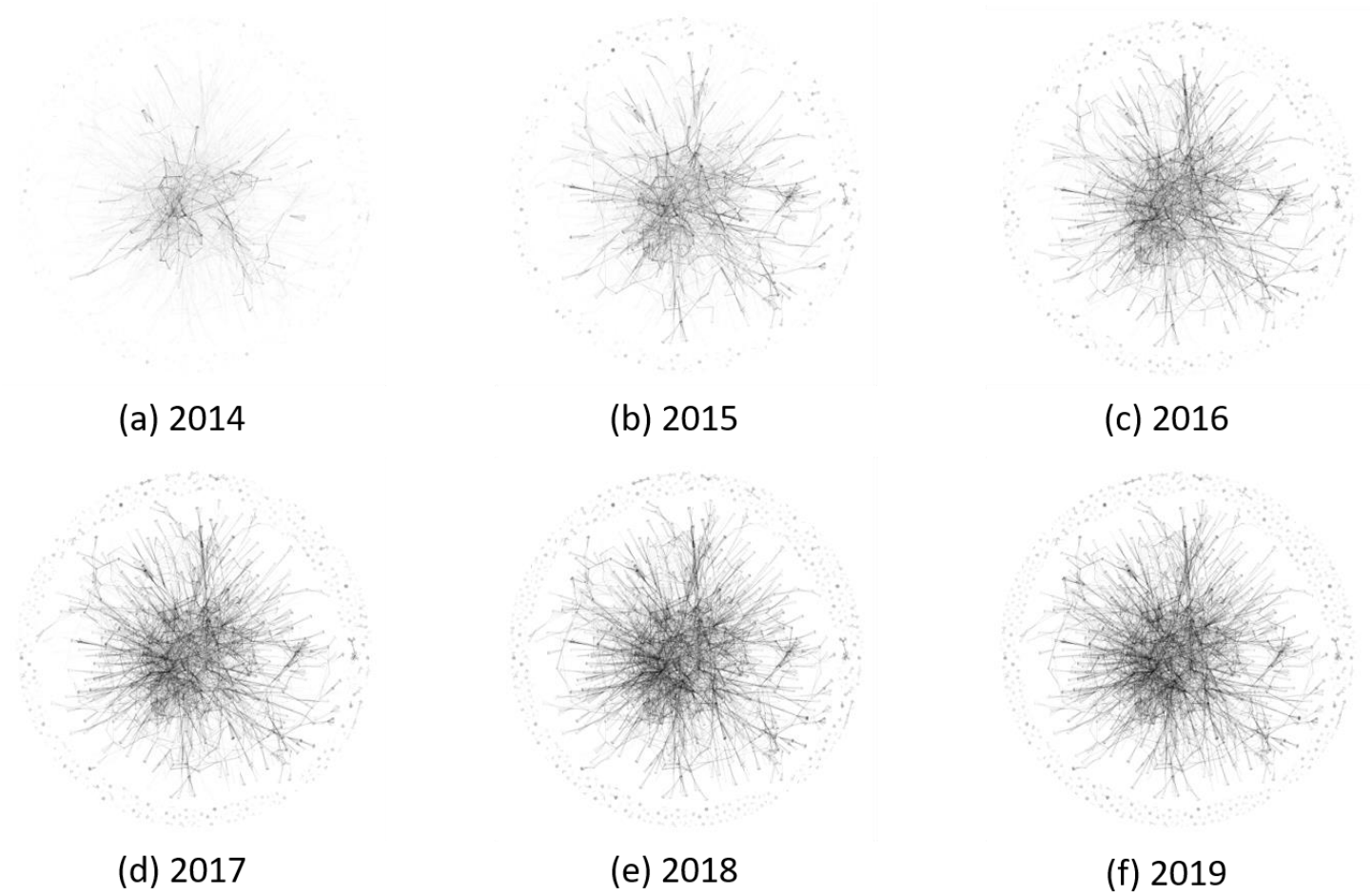

Figure 3.14 Time evolution of the co-word macro-network. The upper row represents the period of Phase 1 (FY 2013-2015) and the lower row represents the period of Phase 2 (FY 2016-2018) in the COI Program. The nodes and edges in this network cumulatively increase over time.

3.4.2 Cross-sectional analysis of the complexity of collaboration and knowledge structures

Graph entropy of the co-author and co-word network were calculated for each static macro-, meso-, and micro-network. Since the calculated graph entropy is a quantity defined for the static network itself, changes in the complexity of these network structures caused by the structural decomposition should be measured by their differences. To measure such changes, the difference in the graph entropy between the macro- and the meso-network, and between the macro- and the micro-network was used.

Figure 3.15 plots the difference in the graph entropy between the macro- and the meso-network. The horizontal axis represents the difference in the co-author network and the vertical axis represents the difference in the co-word network. Each point plotted in Figure 3.15 represents each subtheme, and the value of graph entropy was calculated from the co-author and co-word meso-networks shown in Figure 3.3-3.7 and 3.8-3.12.

Figure 3.16 plots the difference between the graph entropy of the macro- and the micro-network. The horizontal axis represents the difference in the co-author network and the vertical axis represents the difference in the co-word network. Each point plotted in Figure 3.16 represents an individual, and the values of graph entropy are calculated from the co-author and co-word micro-networks, which are defined for each individual, as well as the meso-network defined for each subtheme shown in Figure 3.3-3.7 and 3.8-3.12. According to Figure 3.16, a tendency for the points representing each actor to align linearly was observed.

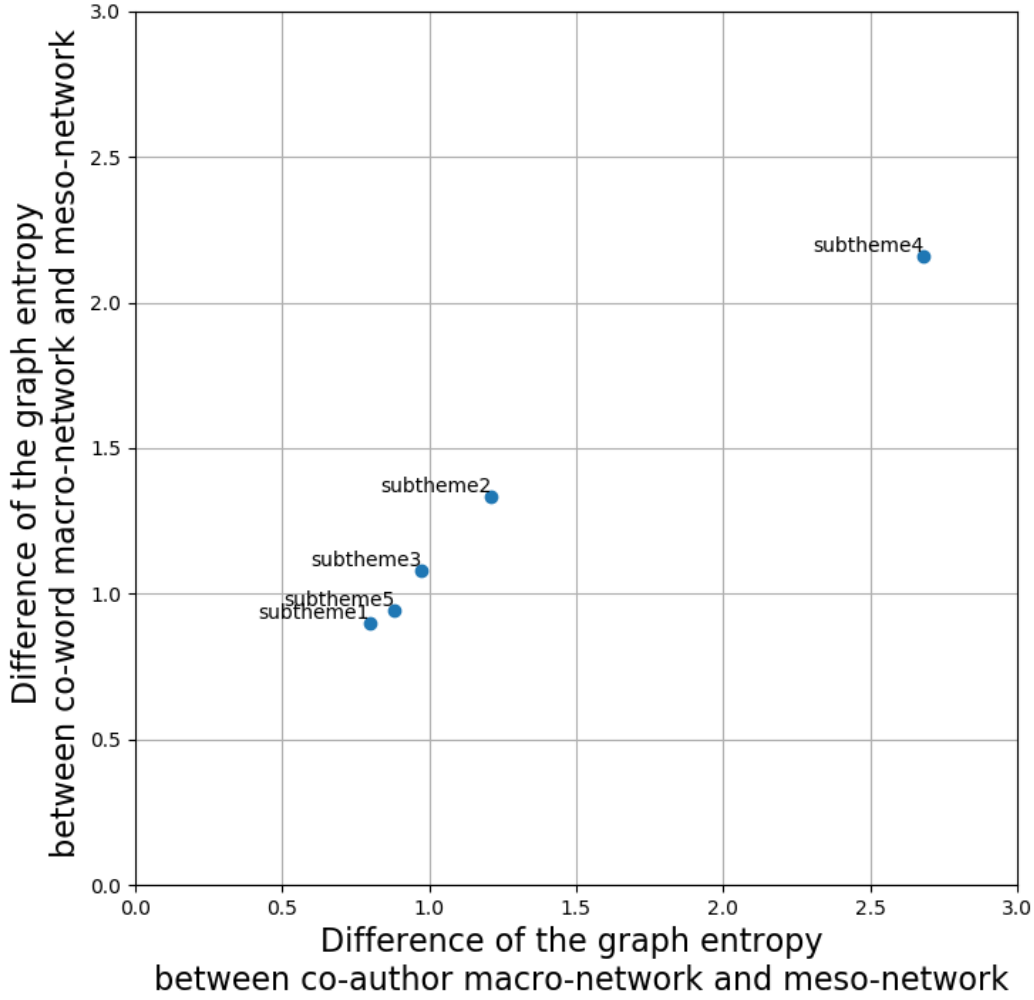

Figure 3.15 Two-dimensional plot of subthemes according to two types of graphical entropy values. The horizontal axis represents the difference of graph entropy between the macro and meso-network of co-authorship whereas the vertical one represents the difference of those of co-word network.

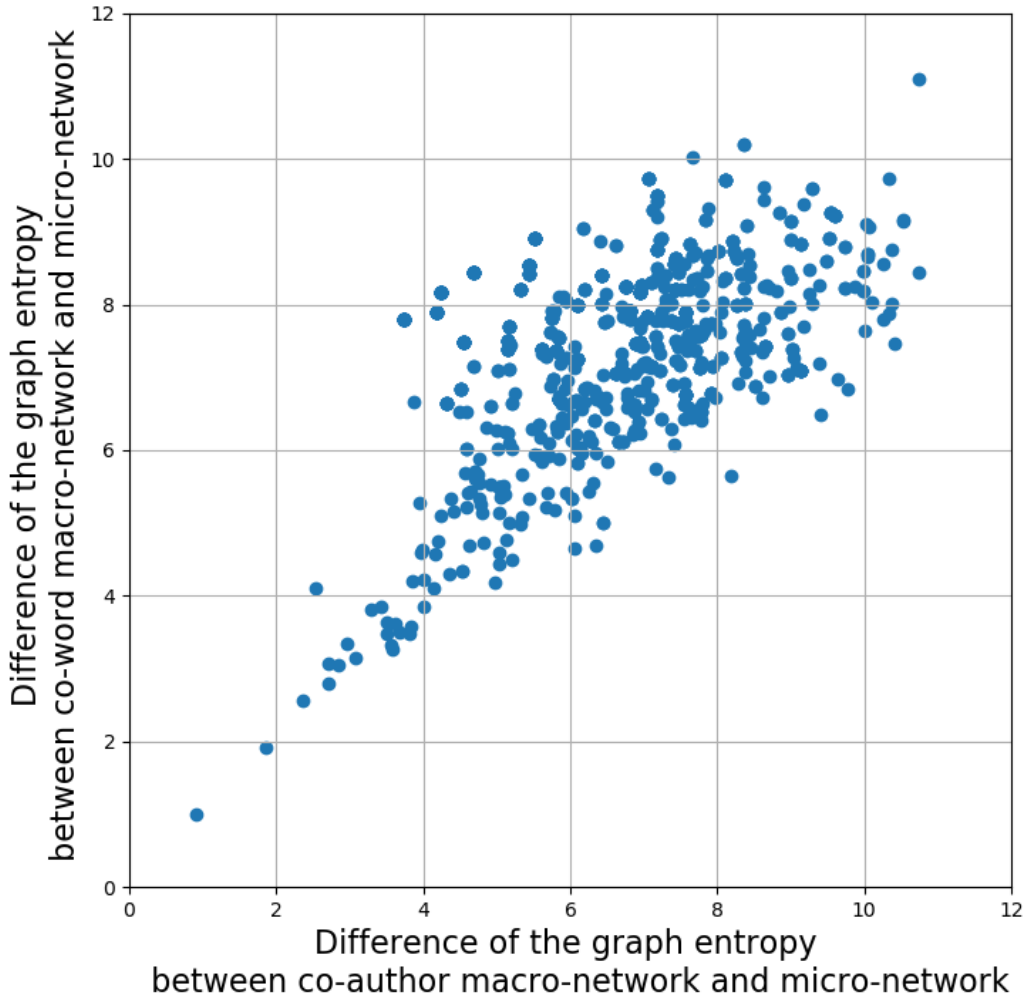

Figure 3.16 Two-dimensional plot of individuals according to two types of graphical entropy values. The horizontal axis represents the difference of graph entropy between the macro and micro-network of co-authorship whereas the vertical one represents the difference of those of co-word network.

3.4.3 Longitudinal analysis of the complexity of collaboration and knowledge structures

For the dynamic co-author and co-word networks, graph entropy and normalized graph entropy by year were computed from the panel data of the adjacency matrices. Figure 3.17-3.25 shows the time-course plot and trajectories of graph entropy and normalized graph entropy over time for the co-author and co-word macro-networks and meso-network. Figure 3.21, 3.23, 3.25 were scaled for easier to see. The results reveal the following trend: although the graph entropy of the co-author network increased from the beginning of the project, the increment gradually decreased as the project progressed. Moreover, a decreasing trend in the latter half of the project was observed. Meanwhile, the graph entropy of the co-word network also increased after the project started and its increment became smaller. However, it continued to increase over time during the observed period. In addition, for both the co-author and the co-word networks, the variations in the graph entropy of the meso-networks were large at the beginning of the project and became smaller as the project progressed. Finally, the time variations in the graph entropy of the meso-networks became equal to those of the macro-network.

Looking at the trajectory of graph entropy, the fact that the graph entropy of the co-author and the co-word networks increases monotonically with time was shown. On the other hand, trajectory of the normalized graph entropy showed a different tendency from the (not normalized) graph entropy, which confirms the differences among the subthemes. First, for subthemes 1, 3 and 5, the normalized entropy of the co-author network decreases over time, but the normalized entropy of the co-word network decrease once and then increases. Subtheme 2 shows that the normalized entropy of the co-word network continues to rise while the normalized entropy of the co-author network decreases. In subtheme 4, the normalized entropy of the co-word network tends to rise, but the normalized entropy of the co-author network decreases after rising once. (These tendency in the trajectory of normalized graph entropy looks like similar to the classification derived by the qualitative analysis in Chapter 4, suggesting some relationship.)

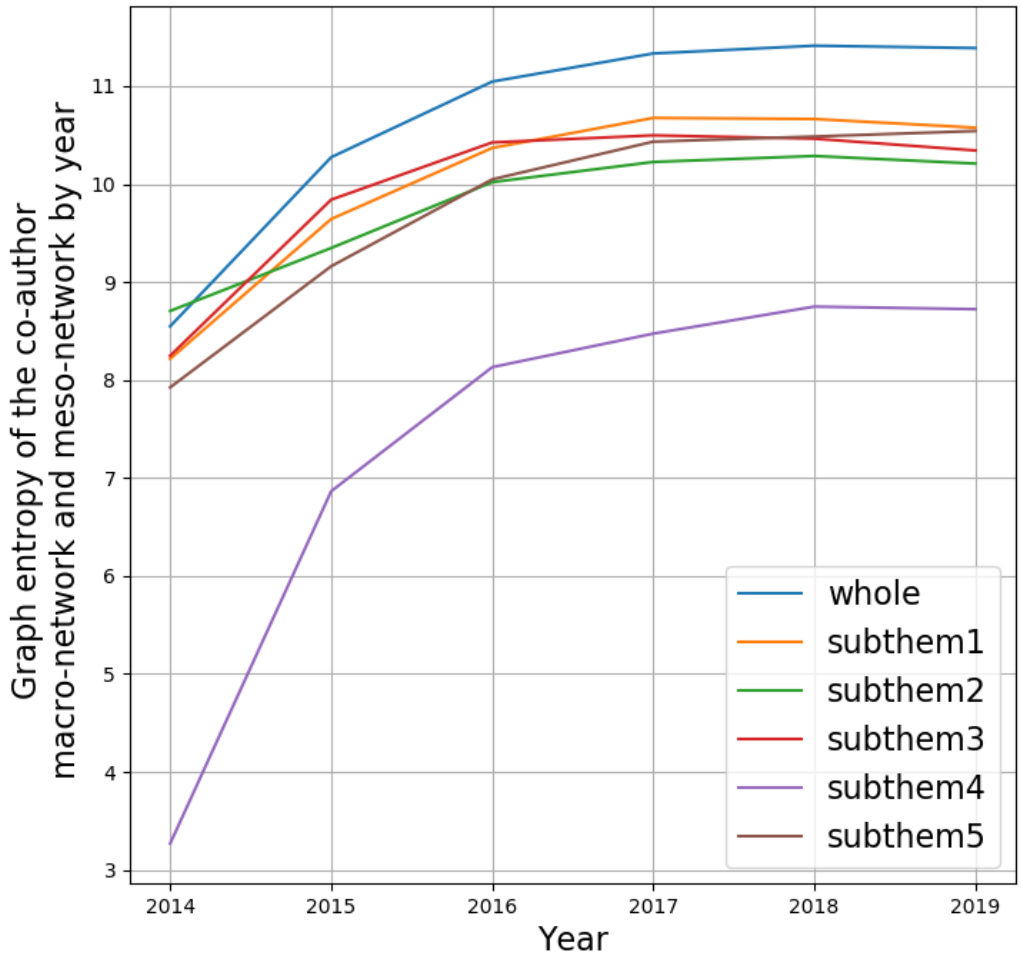

Figure 3.17 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents graph entropy over time for the co-author macro-network and meso-network.

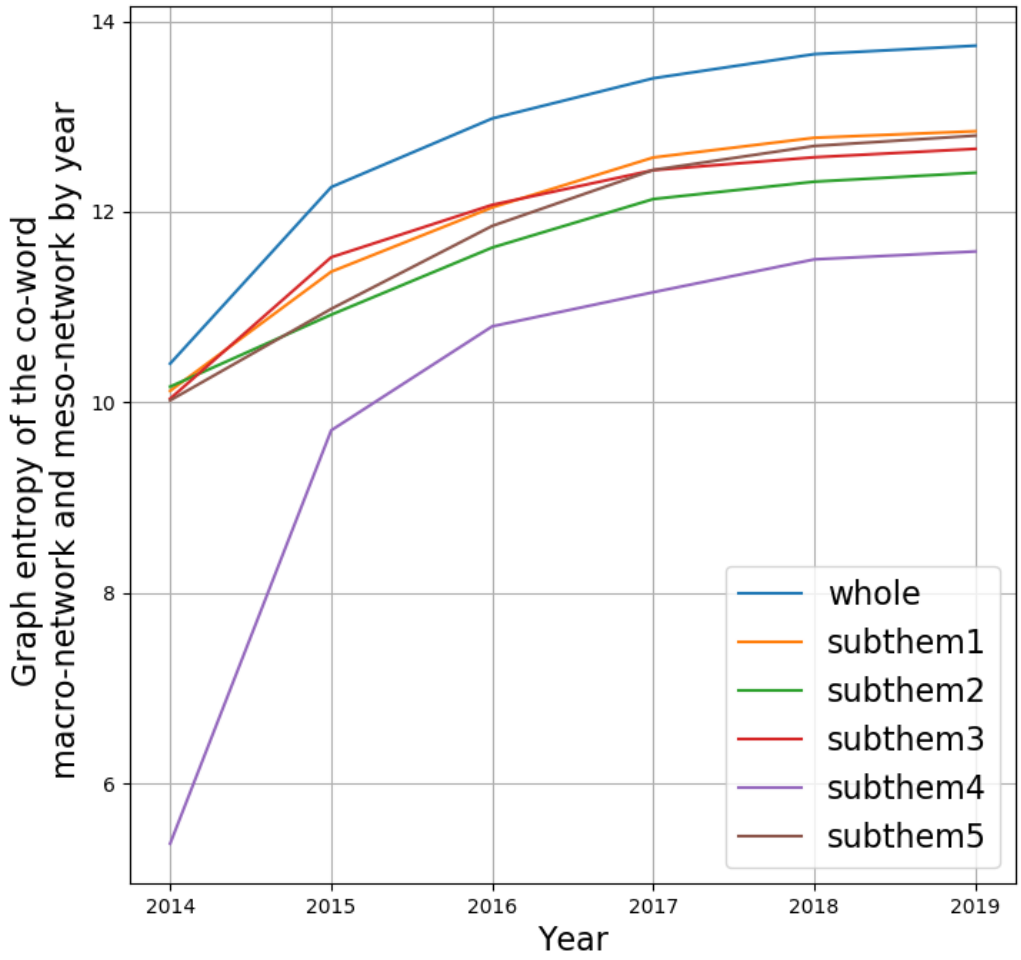

Figure 3.18 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents graph entropy over time for the co-word macro-network and meso-network.

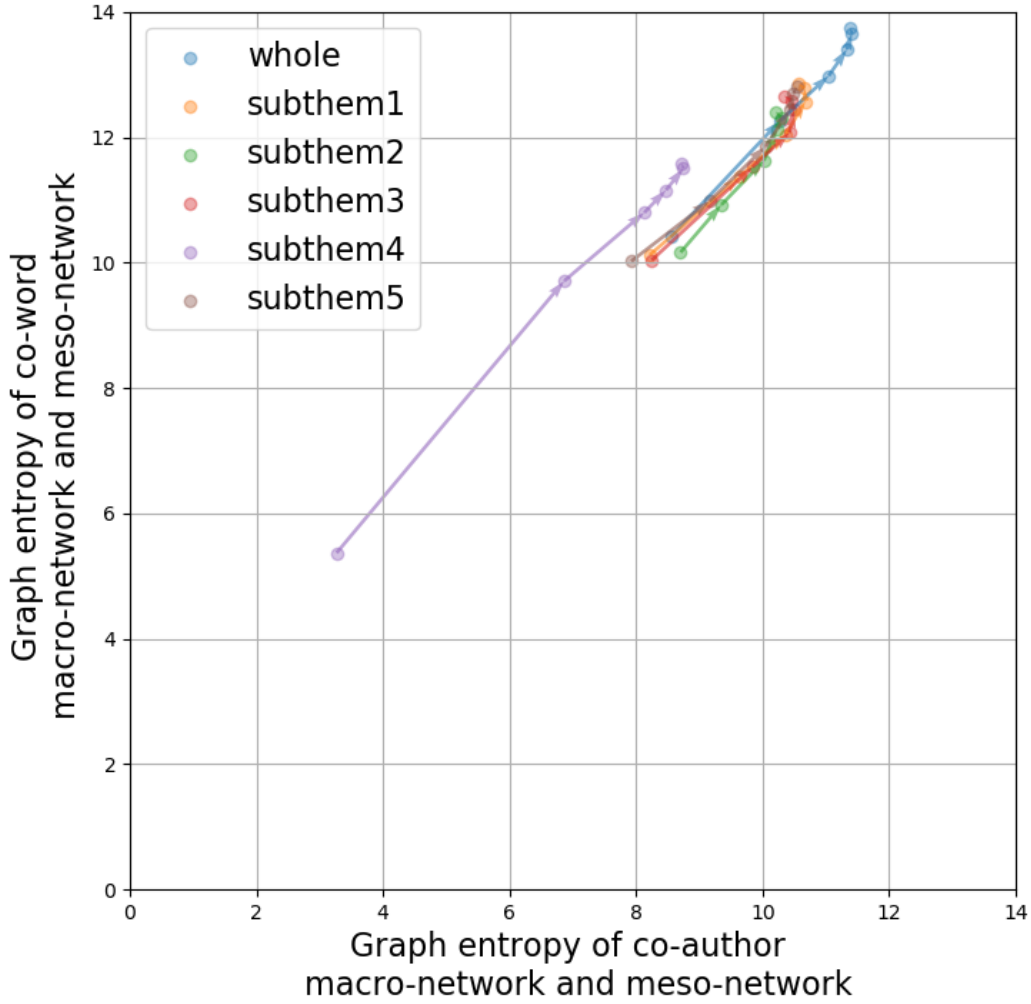

Figure 3.19 Trajectory of the graph entropy of co-author and co-word networks. The horizontal axis represents the graph entropy for the co-author macro-network and meso-network. The vertical axis represents the graph entropy for the co-word macro-network and meso-network.

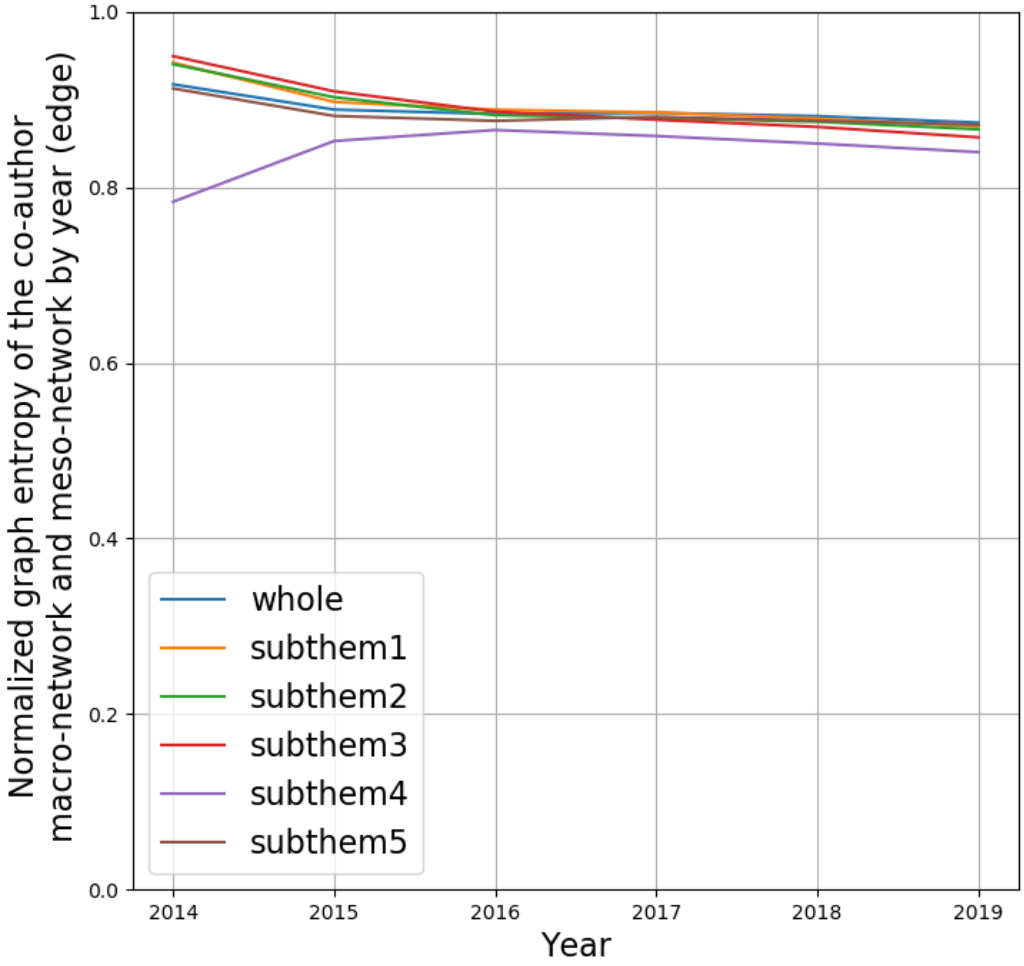

Figure 3.20 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents normalized graph entropy over time for the co-author macro-network and meso-network.

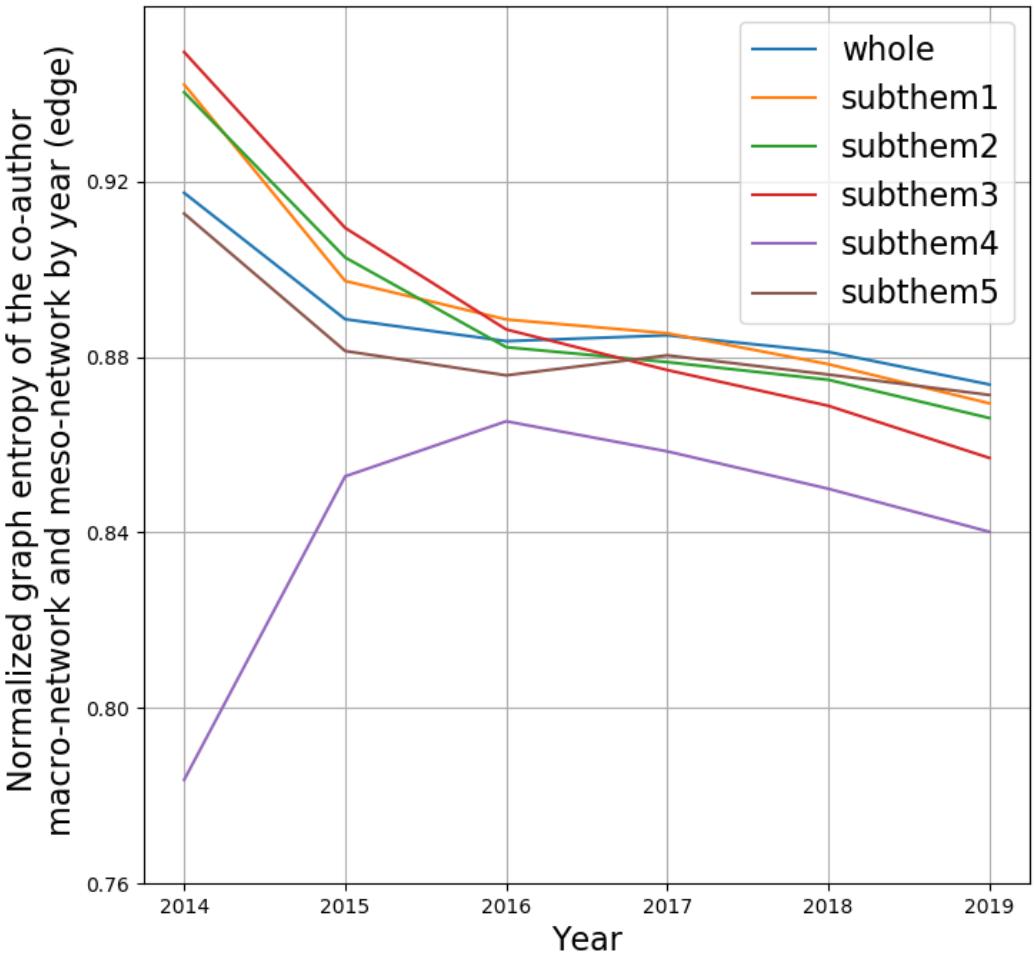

Figure 3.21 Time-course plot of the whole project and 5 subthemes (scale adjusted for easier to see). The horizontal axis represents the period of observation in the present study. The vertical axis represents normalized graph entropy over time for the co-author macro-network and meso-network.

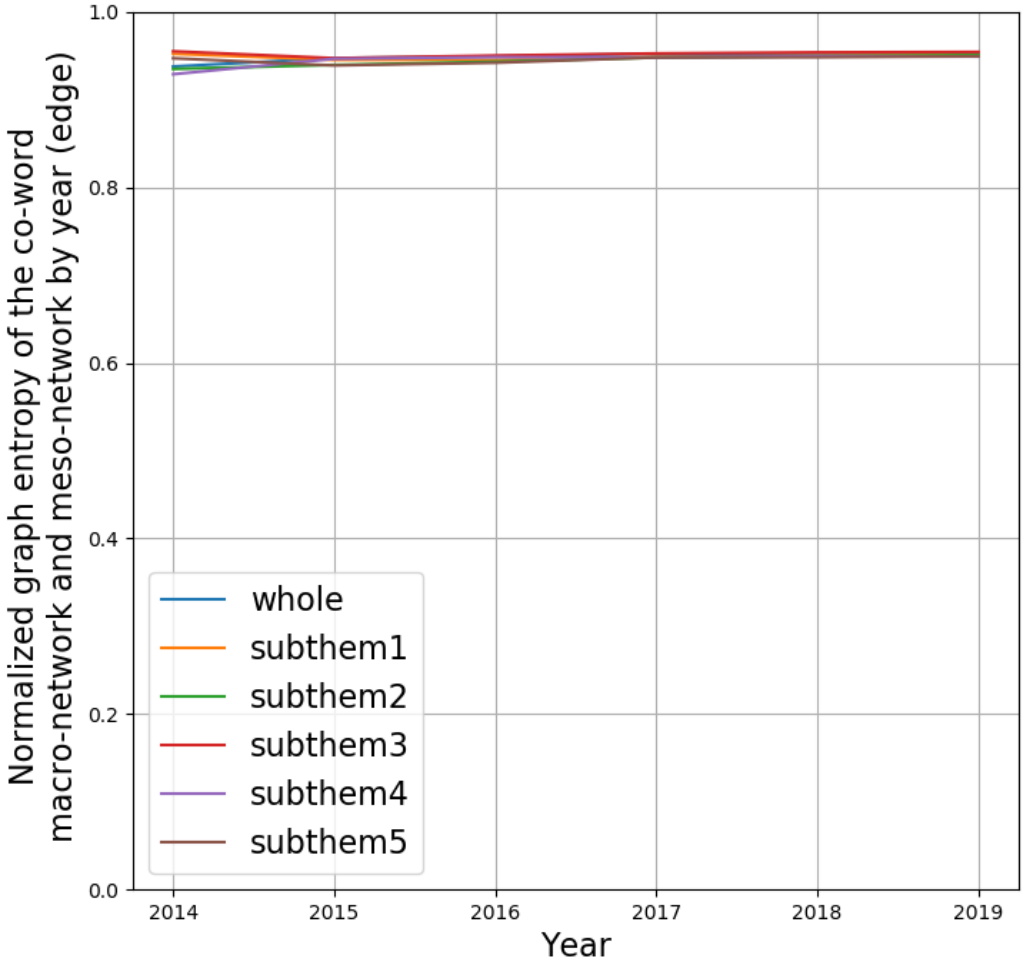

Figure 3.22 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents normalized graph entropy over time for the co-word macro-network and meso-network.

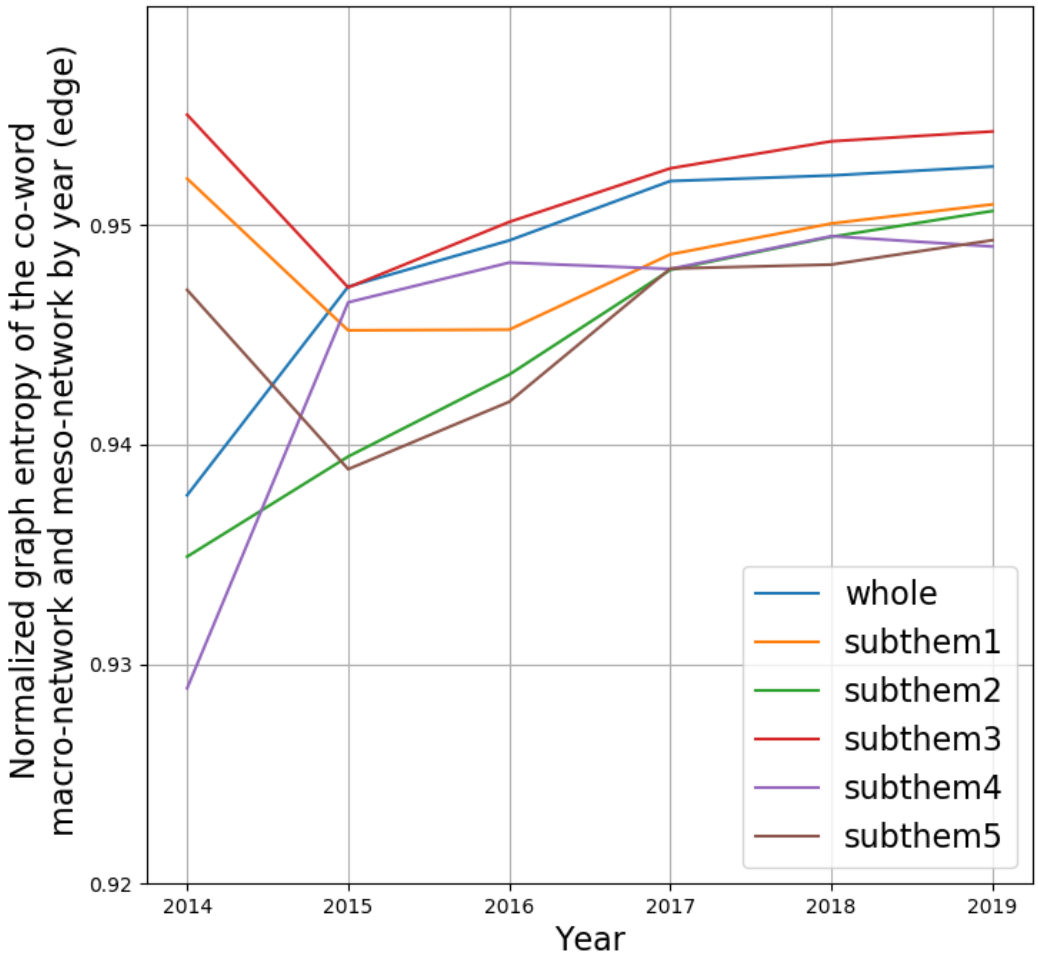

Figure 3.23 Time-course plot of the whole project and 5 subthemes (scale adjusted for easier to see). The horizontal axis represents the period of observation in the present study. The vertical axis represents normalized graph entropy over time for the co-word macro-network and meso-network.

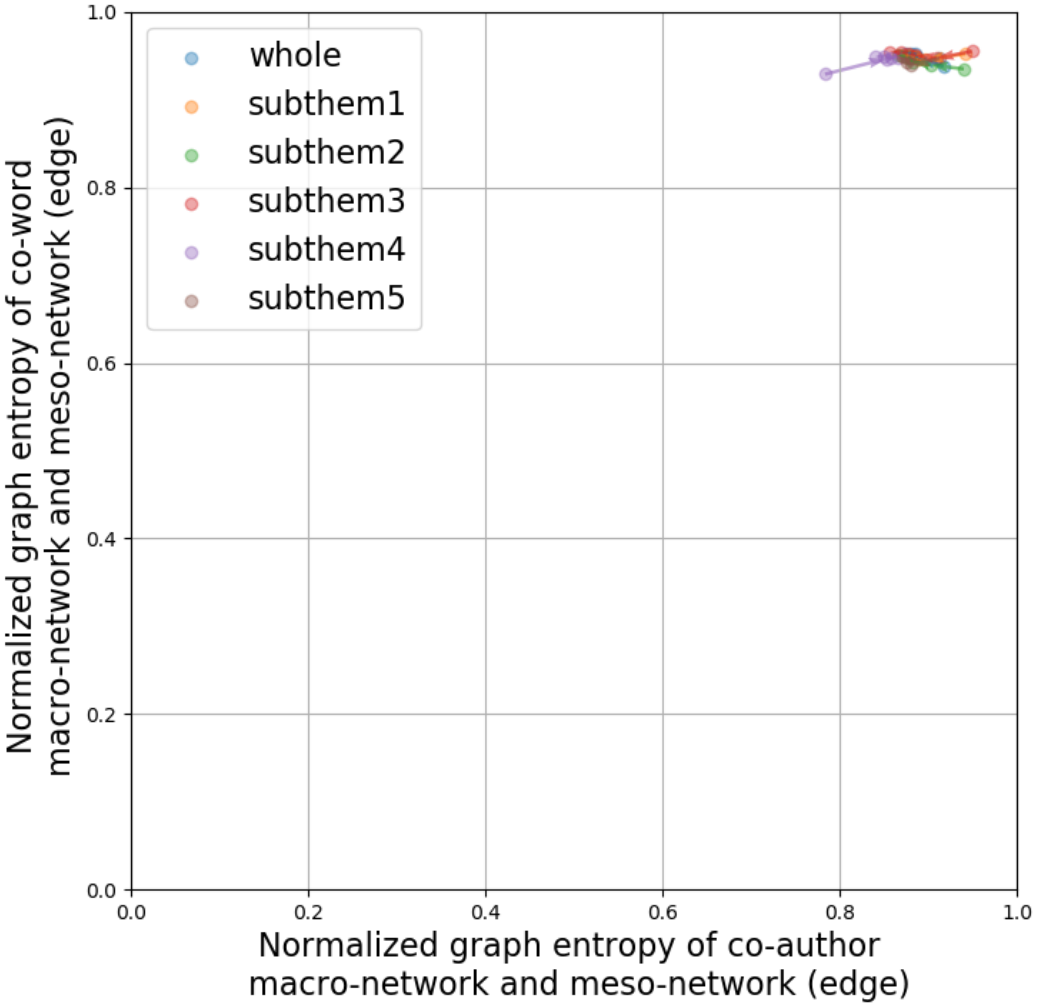

Figure 3.24 Trajectory of the normalized graph entropy of co-author and co-word networks. The horizontal axis represents the normalized graph entropy for the co-author macro-network and meso-network. The vertical axis represents the normalized graph entropy for the co-word macro-network and meso-network.

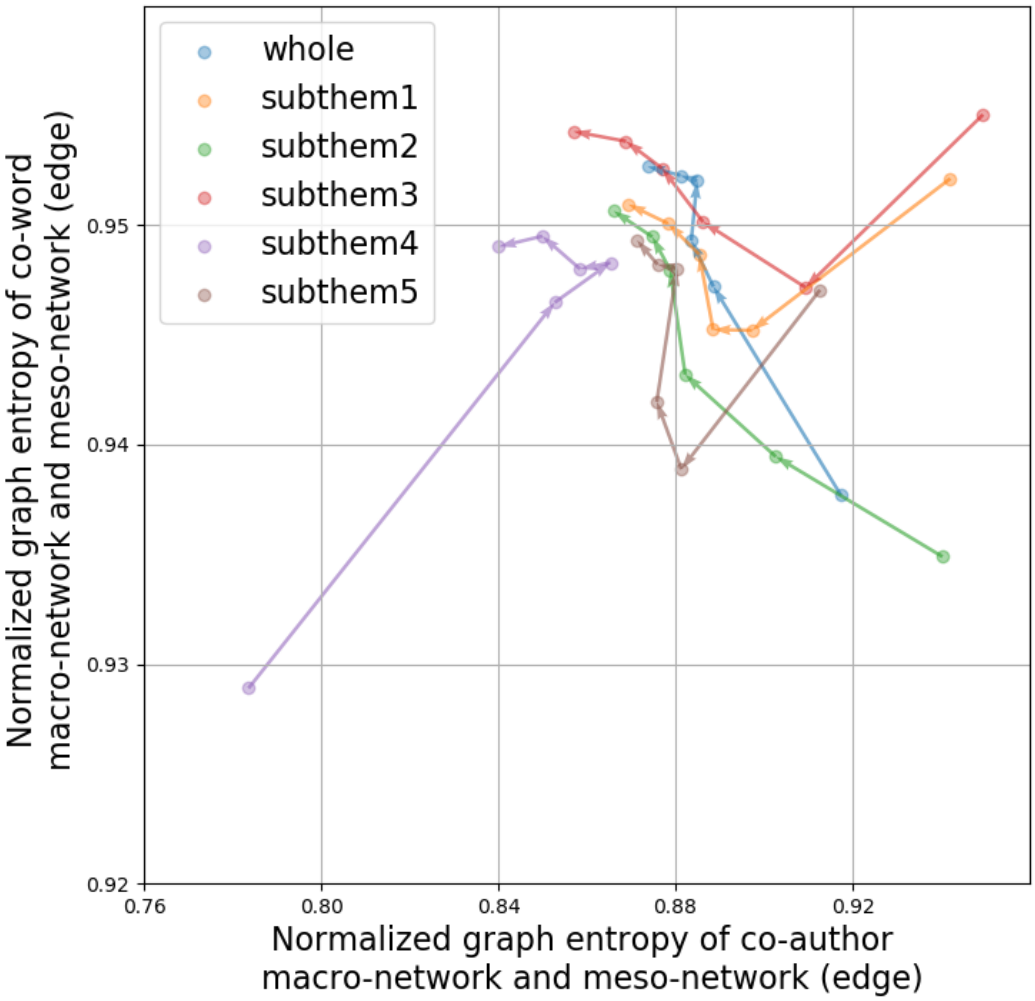

Figure 3.25 Trajectory of the normalized graph entropy of co-author and co-word networks (scale adjusted for easier to see). The horizontal axis represents the normalized graph entropy for the co-author macro-network and meso-network. The vertical axis represents the normalized graph entropy for the co-word macro-network and meso-network.

3.5 Discussion

3.5.1 Interpretation of the structure of co-author and co-word networks

Observed structural features can be interpreted more clearly by adding qualitative information about the target of case study[11]. The co-author meso-network (Fig. 3.3-3.7) seems to reflect the collaboration structure that can be compatible with both efficiency and diversity in transdisciplinary research activity. This network has the core focusing on basic research in nano-biotechnology and each specific cluster of subthemes 1, 2, 3, and 5 for applied and clinical research. The knowledge structure represented by the co-word meso-network (Fig. 3.8-3.12) also reveals a similar tendency. The core cluster at the center of the network represents the fundamental knowledge of nano-biotechnology, whereas the peripheral part is interpreted as applied knowledge.

Meanwhile, subtheme 4 focused not on nano-biotechnology itself, but on its application, that is, the development of diagnostic devices. The co-author and co-word meso-network of this subtheme reflect the characteristics of this research domain. A small number of people can conduct this research in an isolated environment due to relatively less labor-intensive processes (e.g., animal experiments, which are often used in basic nano-biotechnology research); the new concepts originated in device development are far from the established body of knowledge in basic science.

The phenomena observed in time evolution of dynamic networks can be interpreted as a reflection of the management policy. In the early phase of the project, each subtheme was assigned individual research topics and generated research outcomes from basic science. However, such a research strategy was inadequate to realize the vision of the project as a whole. After the project progressed, as the outcomes obtained by that time were expected to integrate and reconstruct from diverse interdisciplinary or transdisciplinary fields to an established monodisciplinary one, the collaboration structure consisting of many modules was gradually connected. Furthermore, the knowledge structure evolved to a higher level through the deepening of the knowledge basis of nano-biotechnology and expanding the applicable limits. The research leader concurred with the above interpretation, confirming that the results of our analysis accurately reflected the status of ongoing research projects. It was suggested that the visualization of the collaboration and knowledge structure can be used to plan future research strategies.

3.5.2 Changes caused by structural decomposition in the complexity of collaboration and knowledge structures

As is often observed in large-scale research projects, groupings to allot individual topics to small research groups are considered to reduce complexity by decomposing the collaboration and knowledge structures into their substructures. The cross-sectional analysis confirmed that the reduction in complexity of collaboration structure was correlated with that of the knowledge structure when these

[11] This information was obtained from internal meetings, symposia, and personal communication with researchers and staff working at the core organization (iCONM) through the author's participation as a member of the COINS.

were decomposed into their substructures. This suggests that the collaboration and knowledge structures in research projects are interdependent, and a change in one structure can induce a change in the other. If this implication is correct, the nature of such collaboration and knowledge structures can be beneficial for managing interdisciplinary research projects with scientific research, because organizing appropriate research groups enable us to facilitate the creation of the desired interdisciplinary knowledge.

The nature of the graph entropy, based on the distribution of edge weights, is such that its value depends not only on the distribution of edge weights but also on the network size. In other words, the more edges there are in the network or the more papers there are, the larger the value of graph entropy is likely to be[12]. Therefore, the difference in the graph entropy for each meso-network can be interpreted as representing the magnitude of interdisciplinarity as a quantification of the variety and balance of the relationships between the components of the collaboration and knowledge structures in the research activities for each subtheme[13] (Stirling 2007; Wagner et al. 2011).

Based on the aforementioned insights, the present study built a data-driven hypothesis that, the creation of interdisciplinary knowledge was intended to achieve change in the knowledge structure through deliberately manipulating the collaboration structure. However, this project emphasized emergence of interdisciplinary knowledge based on internal inductions to a shared vision, rather than external interventions such as organizational restructuring of management to achieve project goals[14]. The maintenance of a complex collaboration structure is considered a necessary condition for the creation of transdisciplinary knowledge. On the other hand, it is not considered to be a sufficient condition for inducing changes in knowledge structures.

3.5.3 Changes caused by time evolution in the complexity of collaboration and knowledge structures

Considering the objectives of interdisciplinary research projects, the aforementioned structures will evolve such that their complexity increases over time. In addition, as explained in the previous section, the graph entropy tends to become larger as the number of edges in the network increases with the increase in the number of papers. Therefore, the complexity of these structures in interdisciplinary research projects is expected to increase over time. Although the graph entropy of the co-author and co-word network tends to increase over time, the fact that these increments falter implies there is a

[12] If more detailed system complexity should be measured, we should consider not only the objective complexity that depends on the system size, such as the graph entropy introduced in present study, but also the subjective complexity that reflects the disparity (or distance) from the reference model (Efatmaneshnik and Ryan 2016).
[13] The present is intended to measure the structural complexity which reflects the strength of ties between the network components (i.e., authors or words). For example, the complexity can be different magnitude even if the number of edges is the same, because the strength of ties between researchers who have been collaborating with many co-authorships for a long time is different from the strength of ties between students and their supervisors who appear only in a single co-authored paper.
[14] For creation of interdisciplinary project outcomes, the research leader referred necessity of the mindset that individual scientists voluntarily set research agenda to realize the vision of the project.

factor that suppresses the increase in the complexity of the collaboration and knowledge structures, which can be interpreted as the result of project management preventing these structures from becoming complex. From the project management perspective, excessively complicated collaboration and knowledge structures are undesirable; hence, the design and implementation of external interventions are needed to suppress the increase in complexity. Put simply, interdisciplinary research projects require flexible management to control their complexity according to the project progresses.

The historical pattern of the complexity of these structures implies the research strategy and the management style taken in this case. As shown in Figure 3.13 and 3.14, the saturation of the complexity over time was caused by a gentle increase in the number of edges. In the early years, it can be interpreted that the adoption of strategies to promote interdisciplinary research offered an increase in complexity, but that the complexity of these structure was controlled and reduced as a result of the interventions to converge on accountable project outcomes for national investments. According to the research leader, because of the operations by an entity independent of established organizations, the project provides an environment in which scientists can voluntarily engage in open and agile inter-organizational collaboration, thereby promoting active changes in the collaboration structure. The fact that the complexity of the collaboration structure began to decrease at a certain time while that of the knowledge structure continued to increase over time can be interpreted the reflection of such project operations.

Moreover, the saturation of graph entropy over time might be caused by the saturation of the number of edges, based on a proportional relationship between the time variation of the graph entropy and of the logarithm of the number of edges (data not shown). This tendency was interpreted as a reflection of the observed management style where subthemes and subtheme leaders (core individuals) were fixed and had been unchanged. In other words, If the subthemes have been drastically reorganized and/or these core individuals have been replaced in a different management style, by the definition of the graph entropy, the increase in the number of edges would have not been saturated and the distributions of edge weights would have been changed over time, resulting in changes in the pattern of graph entropy values. However, because of a limitation to test this speculation it in this study that deals with a single case, it should be examined separately in future studies based on a comparison of multiple cases.

3.5.4 Contributions and limitations

The theoretical contribution of the present study is the proposal of a framework for the strategic and organizational management of scientific research, which reveals insights through the extension of scientometric methodology to real-time activity measurement. The proposed framework can provide evidence for the management of public scientific research, and may contribute to evidence-based management (Rousseau 2006) as well as the systematization of the theory for management of

interdisciplinary research project[15].

The practical implications support the practice of evidence-based management, especially in publicly funded scientific research projects. Incorporating the process of the hypothesis-driven data instrumentation can accelerate the hypothesis-testing cycle for data-driven hypotheses in management of interdisciplinary research project. Practitioners of management of interdisciplinary research project, scientific research project managers, and research administrators[16] should build and test their own hypotheses to form the basis of the managerial decision-making. Introducing the methodology for data-driven decision-making will be fruitful for the ongoing development of management of interdisciplinary research project (Figure 3.26).

Finally, the present study has several the limitations. First, the graph entropy of a co-word network depends on the selection of nodes. Some factors, such as the criteria of the word elimination in the preprocessing, how to extract key phrases from the text, and calculations of edge weights in the network edges which may cause some issues in the data. Second, the generalizability of the findings and discussions cannot be guaranteed as single-case analysis was employed. More in-depth evidence is desirable through comparative studies across projects with controlled conditions (e.g., multiple projects in a single funded program), systematic reviews, or meta-analyses. Third, the argument on the time evolution of dynamic networks is only qualitative descriptions. The present study avoided the deeper topic of quantitative aspects of network time evolution itself as this is outside the scope of this research. Future research can utilize quantitative analyses for understanding the dynamics of network evolution. Fourth, the graph entropy introduced in the present study depends on the size of the system (i.e., the number of edges) and does not consider a distance from a reference model, which is typically a network where the edge weights follow a uniform distribution (Efatmaneshnik and Ryan 2016). Incorporating the notion of this distance may provide a better measure of complexity.

[15] In the practice of management, the fact that only evidence-based knowledge is not a unique solution for decision-making is highlighted (Lengnick-Hall and Griffith 2011). Desirable decision-making requires balanced and appropriate use of accessible knowledge at the time.

[16] Not all of them are always familiar with individual research topics of the state-of-the-art science. Of course, due to desirability they also understand such research topics, scientists themselves often play the role of project managers for scientific research. Meanwhile, recent research and development projects increasingly emphasize transdisciplinary research. Managing scientific research involving various stakeholders by scientists alone has already been the limit.

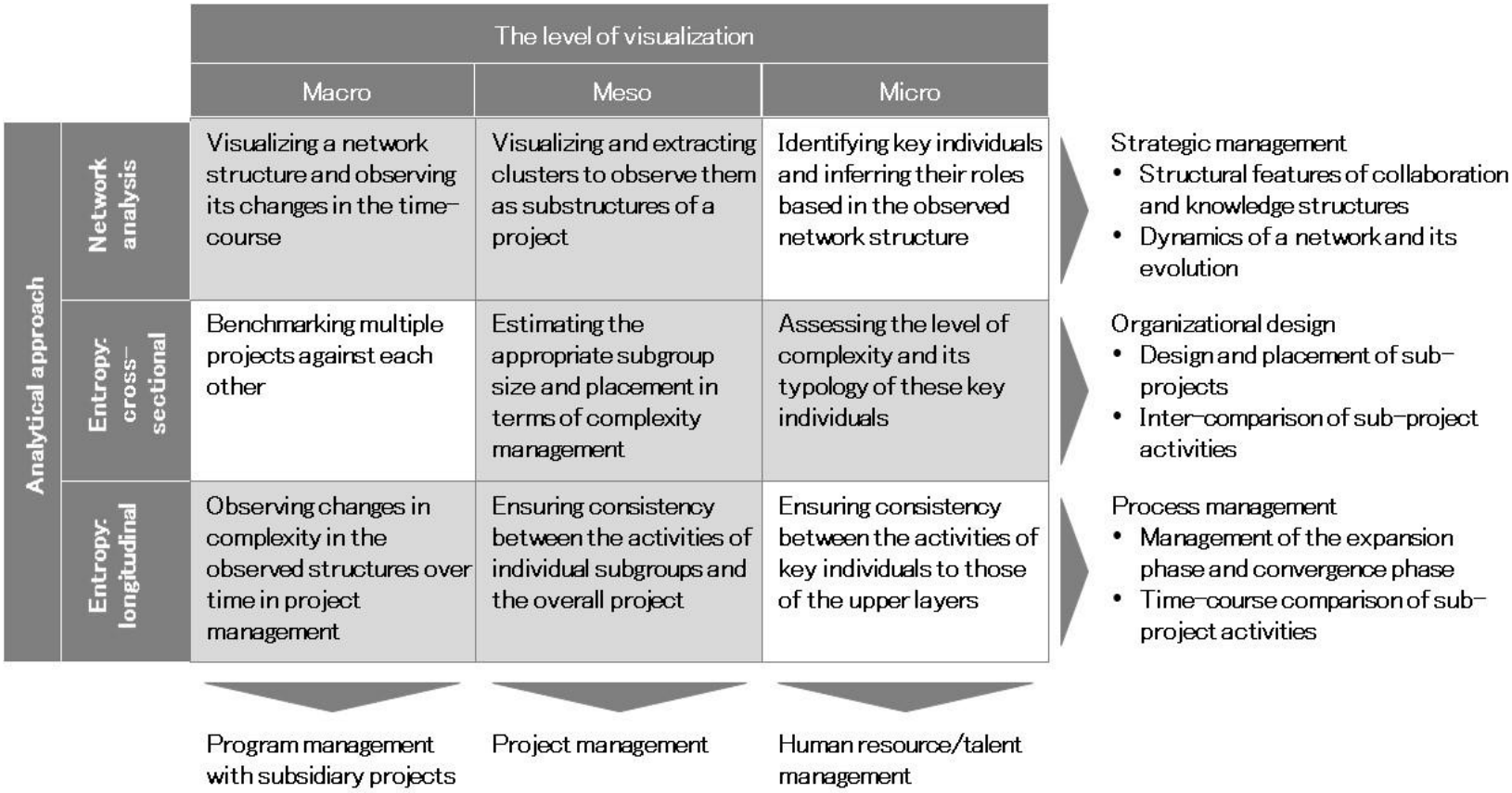

Figure 3.26 A proposed framework for the management of interdisciplinary project based on the present study. Each box represents an analytical outcome in accordance to the level of visualization (horizontal) and analytical approaches (vertical). The shaded areas are based on the empirical evidence of the present study whereas white ones are described hypothetically.

3.6 Conclusions

This study focused on the relationship between the collaboration and knowledge structures in an interdisciplinary research project and the static and dynamic characteristics of the co-author and co-word networks obtained from bibliographic information. Furthermore, the structural complexity of these networks was quantified by the graph entropy, which was used to perform some evaluations at the micro (individual), meso (subtheme) and macro (project as a whole) levels. The insights derived from measuring the structural complexity for the management of interdisciplinary research projects were verified through an interview with the research leader. The proposed framework for strategic and organizational management in scientific research can offer real-time measurement of the activities and broad coverage over many fields.

The observational characteristics of the aforementioned structures revealed the interdependence between these structures, reflection of the domain-specific properties at the meso level, and the different dynamics in time evolution. The cross-sectional analysis demonstrated a correlation between the changes in the complexity of the structures caused by structural decomposition. This result suggested that the complexity of such structures in an interdisciplinary research project was managed by organizing appropriate research groups. Moreover, changing the collaboration structure can induce the change in the knowledge structure. The longitudinal analysis confirmed that the variability of the complexity decreased over time in both the collaboration and knowledge structures. This reflects the strategy that this project pursued interdisciplinarity (or innovative research outcomes) in the early stages and the integration of obtained outcomes for realizing the vision in the later stages.

References

Anzai, T., Kusama R., Kodama, H., & Sengoku, S. (2012). Holistic observation and monitoring of the impact of interdisciplinary academic research projects: An empirical assessment in Japan. *Technovation* 32(6):345–357

Ávila-Robinson, A., & Sengoku, S. (2017a). Multilevel exploration of the realities of interdisciplinary research centers for the management of knowledge integration. *Technovation* 62-63:22–41

Ávila-Robinson, A., & Sengoku, S. (2017b). Tracing the knowledge-building dynamics in new stem cell technologies through techno-scientific networks. *Scientometrics*, 112(3), 1691-1720

Barre, R. (2019). Les indicateurs sont morts, vive les indicateurs! Towards a political economy of S&T indicators: A critical overview of the past 35 years. *Research Evaluation* 28(1):2–6

Börner, K., Contractor, N., Falk-Krzesinski H. J., Fiore ,S. M., Hall, K. L., Keyton, J., Spring, B., Stokols, D., Trochim, W., & Uzzi, B. (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine* 2(49):1–6, 49cm24-49cm24

Efatmaneshnik, M., & Ryan, M.J. (2016). A General Framework for Measuring System Complexity. *Complexity* 21(S1):533–546

Etzkowitz, H. (1998). The norms of entrepreneurial science: Cognitive effects of the new university-industry linkages. *Research Policy* 27(8):823–833

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy* 29:109–123

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A. L. (2018). Science of science. *Science* 359(6379)

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-directed Placement. *Software: Practice and experience* 21:1129–1164

Gomez-Pilar, J., Poza, J., Bachiller, A., Gómez, C., Núñez, P., Lubeiro, A., Molina, V., & Hornero, R. (2018). Quantification of Graph Complexity Based on the Edge Weight Distribution Balance: Application to Brain Networks. *International Journal of Neural Systems* 28(1):1–19

Henderson, R. M., & Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly* 35(1):9–30

Kapsali, M. (2011). Systems thinking in innovation project management: A match that works. International *Journal of Project Management* 29(4):396–407

Kapsali, M. (2013). Equifinality in Project Management Exploring Causal Complexity in Projects. *Systems Research and Behavioral Science* 30(1):2–14

Kell, D. B., & Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays* 26(1):99–105

Kodama, H., Watatani, K., & Sengoku, S. (2013). Competency-based assessment of academic interdisciplinary research and implication to university management. *Research Evaluation* 22(2):93–104

Lengnick-Hall, C. A., & Griffith, R. J. (2011). Evidence-based versus tinkerable knowledge as strategic assets: A new perspective on the interplay between innovation and application. *Journal of Engineering and Technology Management* 28(3):147–167

Leydesdorff, L. (2003). The mutual information of university-industry-government relations: An indicator of the Triple Helix dynamics. *Scientometrics* 58(2), 445-467.

Mackenzie, K. D. (1966). The information theoretic entropy function as a total expected participation index for communication network experiments. *Psychometrika* 31(2):249–254

Ràfols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation* 28(1):7–22

Rousseau, D. M. (2006). Is There Such a Thing As "Evidence-based Management"? *Academy of Management Review* 31(2):256–269

Sen, B., Chu, S. H., & Parhi, K. K. (2019). Ranking Regions, Edges and Classifying Tasks in Functional Brain Graphs by Sub-Graph Entropy. *Scientific Reports* 9(1):1–20

Sen, B., Bernstein, G. A., Mueller, B. A., Cullen, K. R., & Parhi, K. K. (2020). Sub-graph entropy based network approaches for classifying adolescent obsessive-compulsive disorder from resting-state functional MRI. *NeuroImage: Clinical* 26:102208

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal* 27(3):379–423

Silva, F. N., Rodrigues F.A., Oliveira Jr O.N., & Costa L. da F. (2013). Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics* 7(2):469–477

Stirling, A. (2007) A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface* 4(15):707–719

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008a). The Science of Team Science. Overview of the Field and Introduction to the Supplement. *American Journal of Preventive Medicine* 35(2) S77-S89

Stokols, D., Misra S., Moser, R. P., Hall, K. L., & Taylor, B. K. (2008b). The Ecology of Team Science. Understanding Contextual Influences on Transdisciplinary Collaboration. *American Journal of Preventive Medicine* 35(2) S96-S115

Toutanova, K., & Manning, C. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. *In Proceedings of the 2000 Joint SIGDAT Conference EMNLP/VLC*, 63-71, 2000.

Van Helden, P. (2013). Data-driven hypotheses. *EMBO Reports* 14(2):104

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols I. & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* 5(1):14–26

Wang, B., Wang, Z. (2018). Heterogeneity evaluation of China's provincial energy technology based on large-scale technical text data mining. Journal of Cleaner Production 202:946–958

Wang, B., Liu, S., Ding, K., Liu, Z., & Xu, J. (2014). Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology. *Scientometrics* 101(1):685–704

Wang, Q., Schneider, J. W. (2020). Consistency and validity of interdisciplinarity measures. *Quantitative Science Studies* 1(1):239–236

CHAPTER 4. Configurational approach to generate prescriptive knowledge based on qualitative description of contexts, interventions, mechanisms, and outcomes

4.1 Introduction

Science and technology innovation is a driver for economic growth and industrial development; it is also expected to create the means to solve social issues. In recent years, large-scale and complex R&D programs and projects have been adopted with the aim of addressing global societal issues such as the SDGs. As these projects are usually funded by governmental initiatives (i.e., public funds), responsibility for implementation, accountability to society, and transparency of the decision-making process have to be ensured. This means that R&D programs and projects subsidized by public funds are expected to practice evidence-based management by utilizing the knowledge of management science (Rousseau, 2006).

Promoting evidence-based management necessitates the accumulation of empirical evidence on managerial phenomena in the target domain and recognition of the differences between management science and other disciplines (especially, natural sciences). Management science is a discipline that pursues solution-oriented knowledge to construct new systems and interventions or to improve the performance of existing entities, unlike other disciplines whose mission is to describe natural and social phenomena, such as physics and sociology (Van Aken, 2004). The purpose of management science is to acquire prescriptive knowledge that enable us to design interventions for effective and pragmatic management through improvement of systems or their performances, not only providing knowledge to analyze the managerial phenomena but also clarifying the relationships among their elements and the hidden mechanisms.

To provide empirical evidence for acquiring such prescriptive knowledge, hypothesis-driven data instrumentation that enables observation of organizational and managerial phenomena in large-scale interdisciplinary R&D projects is proposed in the previous chapter and instrumentation for changes in the complexity of collaboration and knowledge structures represented by co-author and co-word networks is realized by introducing graph entropy. Instrumentation is a methodology for the quantification and measurement of concepts that are difficult to observe directly by conventional means, because the measurement object does not physically exist, for example, collaboration or knowledge. In other words, the object of instrumentation is a concept that is difficult to measure directly at least at the time. It can only be measured by achieving instrumentation, however, the measurable quantity that represents the concept should reflect the actual situation. Regarding the objects that are difficult to measure, executing the contradictory requirement to confirm the validity by other means is requested. Paradoxically, considering that instrumentation is a concept to encompass even its application to automatic control, the evidence obtained needs to realize instrumentation of objects that have not been measured hitherto yet, but can be verified by the implementation of the

measurement objects. To build a well-grounded theory for management of science, handling the paradox between instrumentation and implementation is necessary.

In this chapter, in the context of a large-scale interdisciplinary R&D project, identification of the mechanisms that generate interventions to achieve research collaboration that result in outcomes, such as the knowledge to achieve the mission, is conducted. The framework that arranges the configuration among contexts, interventions, mechanisms, and outcomes of a case is applied to R&D activities in large-scale projects and the mechanisms for generating interventions to achieve collaboration were conceptualized. In particular, the discussions are how the project as a whole and each subtheme leader implemented mechanisms to generate interventions to achieve research collaboration for creating outcomes in their respective contexts, how the entities for R&D activities in each subtheme (meso-level) are designed and implemented to achieve the entire project objectives (macro-level), and how the characteristics of individual researchers (micro-level) are blended to achieve these objectives (meso-level).

4.2 Framework to generate prescriptive knowledge through configurational approach

4.2.1 Problems related to the research synthesis of case studies on the management of a large-scale R&D program and project

To practice evidence-based management, it is not only necessary to accumulate empirical evidence related to the organizational and managerial phenomena of interest, but also to prescribe solutions to problems based on generalizable insights derived from the multiple evidence accumulated. Owing to this prescriptive aspect, implementation of the prescriptive solutions for managerial and organizational problems requires increased generalization through the integration of empirical evidence from existing research. This approach is called research synthesis (Barnett-Page and Thomas, 2009; Van Aken and Romme, 2009), and there are several established methods, such as meta-analyses (Gurevitchet al., 2018) and systematic reviews (Tranfield, Denyer, and Smart, 2003). However, the application of existing research synthesis methods to the management of R&D programs and projects is problematic for several reasons, including the uniqueness of the case-specific characteristics or the inherent nature of the research synthesis approach. For appropriate research synthesis, the research results and findings should be articulated so that case-specific information is not lost, that is, explicit presentations in each case study request a description of the case study, in what context, what interventions, by what mechanisms are generated, and what outcomes are produced.

In the research synthesis of case studies of the management of large-scale R&D programs and projects, four problems may be considered. First, the difficulty of case collection owing to the small number of available large budget cases or there are no cases with similar managerial and organizational phenomena. Second, the context-specificity of interpretations, which means the loss of superiority of each case through the process of research synthesis because the interpretation of the phenomenon in the original research depends on the context of the case. Third, the heterogeneity of

outcomes across different disciplines due to unsuitability of the same criteria (e.g., using article or patents as in the previous chapter) to evaluate the research outcomes as their contribution of the objectives of the program on project depends on their discipline. Fourth, the compatibility of generative mechanisms for intervention, which means that the mechanisms that resulted in research synthesis as a prescription for the actual managerial problems to be solved need to be compatible with the mechanisms that generate interventions to achieve the intended outcomes in the original case.

These problems can be partially resolved by using a bibliometric approach in individual case studies, as discoursed in the previous chapters. Representing the structural features of collaboration and knowledge in a quantitatively analyzable form like networks enable us to observe managerial and organizational phenomena in individual research activities, to discover related existing theories in management and organization sciences, and to interpret potential contexts based on observational facts that can be discovered by quantitative approaches. Constructing evaluation criteria that reflect the characteristics of the field by adopting multiple measures and expanding the number of documents to be measured can mitigate some of the above problems. However, to reveal how interventions were designed and implemented to achieve the objectives of the program/project in individual cases, and especially how the mechanisms functioned for generating interventions to achieve transdisciplinary collaboration, bibliometric methods are insufficient, and qualitative information based on field research, such as participant observation, is necessary.

4.2.2 Arrangement framework for configuration of contexts, interventions, mechanisms, and outcomes

The present study explores how research collaborations to generate the knowledge for the project's mission were designed and implemented using a framework to arrange the configurations of context, interventions, mechanisms, and outcomes in focused cases. This arrangement framework is inspired by CMO configuration, a framework for program evaluation used in evidence-based policy making (Pawson, 2002), and CIMO-logic, a framework for synthesis of existing research results and insights to derive design propositions for the prescription of management problems (Denyer, Tranfield, and Van Aken, 2008). This framework has two major parts: one part describing the interventions and their generative mechanisms for meso-level research collaboration to obtain macro-level outcomes in the context of overall project management and the other part describing the interventions and their generative mechanisms for micro-level research collaboration to obtain meso-level outcomes in the context of each subtheme management. The former focuses on the entire project management, which is the process of translating operations from macro-level objectives to meso-level activities, while the latter focuses on the characteristics of each subtheme leader, who translates meso-level activities into individual micro-level actions and discusses how these translations were integrated and implemented as an entire system.

4.3 Descriptions of contexts, interventions, mechanisms, and outcomes for the entire project and subtheme leaders based on qualitative observations

4.3.1 Context, interventions, mechanisms, and outcomes of the entire project

Context of the entire project

COINS is an R&D project funded by the COI Program, which aims to provide intensive support for high-risk but challenging interdisciplinary and cross-field integration based on a future vision of society and life realized by innovation that transcends existing disciplinary and organizational boundaries. This implies that this project sets research agenda and seeks their solutions based on societal demands, that is, it is a mission-oriented project.

Unlike existing projects that aim to improve industrial and medical value by developing existing technologies and treatment, this project ultimately aims to implement a new concept of a "smart life care society," in which people can autonomously gain better health in their daily lives. For the same, this project aims to achieve an "in-body hospital" by 2045 using smart nanomachines based on nanobiotechnology. This vision of an "in-body hospital" depicts that everyone will be liberated from the psychological, physical, and economic burdens of exiting medical care while spending daily lives (e.g., preventing the loss of quality of live (QOL) with hospitalization), using virus-size smart nanomachines that patrol in the human body 24 hours a day to diagnose and treat diseases.

A unique feature of this project is that the core organization is the Innovation Center of NanoMedicine (iCONM), which is operated by the Kawasaki Institute of Industrial Promotion, an affiliated organization of the local government of Kawasaki City. This organization has managed the project as an entity independent from existing organizations like universities, to mitigate their influence. In addition, the junior researchers here are employed by the Kawasaki Institute of Industrial Promotion, which provides R&D environments and ensures their positions to a certain extent.

Another feature of this project is its pseudo-continuity. The key members of this project also participated in NanoBio First, which is supported by the FIRST Program, and this project virtually functions as a successor to NanoBio First. This allows to shorten the processes needed to organize appropriate research teams because the key members have already created trust among them.

Interventions of the entire project

The core institution iCONM assumes the role of a hub for the creation of interdisciplinary knowledge by promoting the emergence of interaction among researchers who share facilities and collaboration among different organizations, taking advantage of the building integrating the facilities with different functions such as microfabrication, materials evaluation, organic synthesis, cell and gene experiments, and animal experiments. This interdisciplinary research institution, which functions to mediate the relationships among different kinds of knowledge and the people who deal with them, is a means to develop platforms for the continuous innovations that are difficult to realize by universities and companies alone, which is one of the objectives of the COI Program. COINS has promoted the

integration of research outcomes from different fields by utilizing such institution with distinctive characteristics.

Initially, this project was intended to achieve the vision of "in-body hospital" by setting up subthemes with different targets. However, as the R&D stage progressed, the integration of subthemes progressed, which is expected to create advanced knowledge by incorporating the research outcomes obtained from each subtheme. From the co-word network representing the knowledge structure, the tendency of the knowledge integration is also confirmed except for subtheme 4, which is a somewhat different research topic from nanobiotechnology.

Even when the R&D stage progressed, the same management scheme was continued without any organizational intervention such as changing the subtheme leader. This could be because COINS emphasizes the "formation of a place" to cultivate the mindset of individual researchers and to promote interaction among researchers and practitioners. In fact, general meetings to report the research progress of each subtheme and research camps to deepen interaction among participants are regularly held with the intent to facilitate the formation of such places.

Mechanisms of the entire project

In this project and its predecessor, NanoBio First, features that the network representing the collaboration structure has a core-periphery structure were observed. This structure allows the actualization of an efficient R&D cycle in the core and ensures diversity by conducting complementary research in the periphery. In fact, the importance of the research topics corresponding to the periphery was also recognized and incorporated as a component of the project while the core R&D activities were intensively conducted among actors who share expertise although the subthemes were separately set during the project's initiation.

Similar implications can be derived from the establishment of AccuRna and its subsequent merger into NanoCarrier. Initially, AccuRna was established as an entity for intensive R&D focused on nucleic acid drugs. Later, as its R&D stage progressed to clinical trials, this firm was merged to leverage the knowledge of NanoCarrier, which already had experience in clinical trials, by consolidating knowledge, especially tacit knowledge like know-how, into one organization so that more efficient social implementation of research outcomes would be achieved. Hence, mechanisms enabling desirable knowledge storage to divide and to merge organizations with core technologies and human resources in accordance with the R&D progress were implemented, by building organizations that specialize in specific domain as separate units apart from the existing one.

Outcomes of the entire project

In this project, three Ps are considered as important outcomes: paper, patent, and platform. The high-quality papers in the field of nanomedicine are obviously important and they endorse the scientific and technical validity. As patents are significantly important in this field than others, the research leader

said that high quality papers and strong patents are not contradictory, and that one of the project outcomes is to foster researchers who can create both to promote social implementation. Furthermore, as a part of platform, the continuous innovations that would be difficult for universities and companies alone would be upheld by startup firms such as Braizon Therapeutics and iXstream, whose basic technology corresponds with the research outcomes of this project.

4.3.2 Context, interventions, mechanisms, and outcomes of subtheme leader 1

Subtheme leader 1 conducted R&D to improve the performance of nanomicelles, which is a core technology of the COINS project. Subtheme leader 1 succeeded to the R&D conducted by Kazunori Kataoka, who is the research leader of COINS, at the University of Tokyo, and developed nanomachines to control in vivo and intracellular dynamics of nucleic acid drugs. In Phase 1, subtheme leader 1 conducted material engineering research to improve the physical properties of nanomicelles, specifically, downsizing nanomicelle particles and developing siRNA/mRNA-loaded nanomachines, along with collaborator Yu Matsumoto (Nomoto, et al, 2014; Kim, et al, 2014). Meanwhile, in Phase 2, the focus of the R&D of subtheme leader 1 shifted to the drug development such as nanodelivery systems for anticancer drugs. Although many co-authors of Phase 1 left the project, Horacio Cabral, who is an associate professor at the University of Tokyo, joined as a new collaborator. Based on the nanomicelles developed in Phase 1, the focus of the R&D was on the implementation of functions such as cancer cell directionality, intracellular visualization, and response to external stimuli (Cabral, et al, 2018; Yi, et al, 2019; Watanabe, et al, 2019).

Context of subtheme leader 1

Subtheme leader 1 is from the research leader's laboratory, who is currently an associate professor and has the laboratory at the Department of Materials Engineering, Graduate School of Engineering, the University of Tokyo, and also works on this project as a visiting scientist at iCONM. Subtheme leader 1's specialty is biomaterials, and engages in the development of nanomachines to improve the effectiveness of treatments for intractable diseases using polymeric materials that encapsulate nucleic acid drugs. The leader mainly targeted cancer and conducted R&D aimed at improving drug delivery system, which realizes the high accumulation rate in cancerous tissues, by developing nanomachines with higher performance than the polymeric nanomicelle encapsulating anticancer drugs that are already under clinical trials.

Interventions of subtheme leader 1

As cancer is a focus domain in NanoBio First, and the research collaboration with key members was already established. Subtheme leader 1 developed the knowledge base for drug delivery systems using biomaterials, mainly using existing relationships, that is, collaboration between research leaders or between existing companies such as NanoCarrier. In addition, efforts for the fusion of

subthemes seems to be relatively emphasized because the knowledge base of drug delivery systems created in the research of subtheme 1 has high potential for application in other subthemes.

Mechanisms of subtheme leader 1

The reason why subtheme leader 1 selected the collaborators who has already established relationships to address the improvement of the basic technology for the drug delivery system may be due to the mechanism that promotes the acceleration of knowledge reproduction through collaboration with actors who have shared expertise, such as the research leader. To enhance the performance of the system as a whole, core knowledge generated through research activities that require a high level of expertise would be iteratively produced because of the necessity of improving not only the elemental technologies but also the system architecture in an integrated manner. Hence, it seems that subtheme leader 1 has reduced communication costs such as those caused by differences of expertise and accelerated the reproduction of core knowledge by forming a transaction-free zone (Baldwin, 2008) with actors who had shared expertise, such as the research leader. In addition, the fact that the research topics covered in subtheme 1 are relatively mature and the increase in actors with shared expertise could promote the effective functioning of such a mechanism.

Outcomes of subtheme leader 1

Subtheme leader 1 developed nanomachines to treat cancers that have been considered difficult to treat with conventional nanomedicine such as metastatic cancer. The leader created a technology to improve the performance of the drug delivery system itself, in other words, the core knowledge as an outcome, which is the basis of drug delivery systems. For example, enhancing tissue permeability by downsizing the nanomachine because existing nanomedicines for these diseases are too large to pass through tissues in vivo.

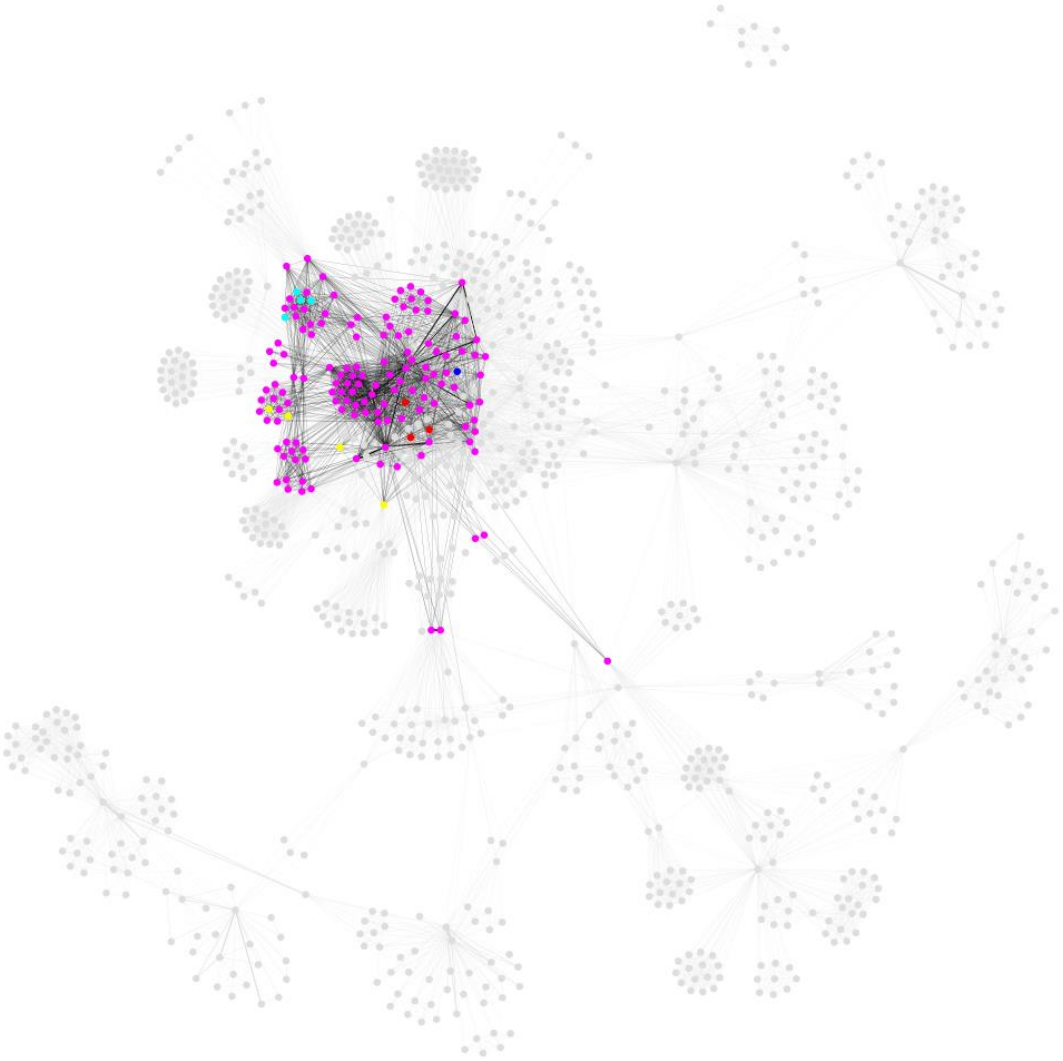

Figure 4.1 Co-author micro-network of the publications from subtheme leader 1 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

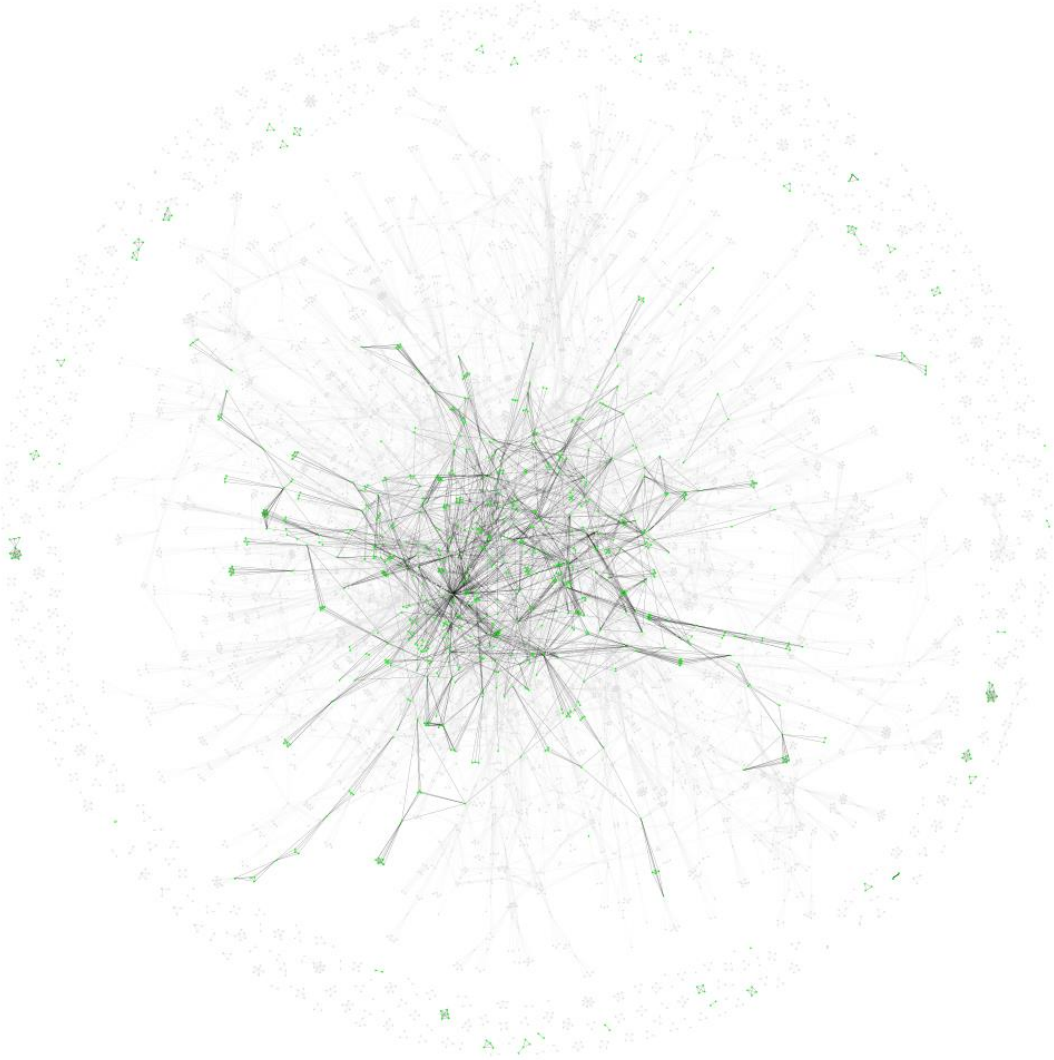

Figure 4.2 Co-word micro-network of the publications from subtheme leader 1 of COINS project.

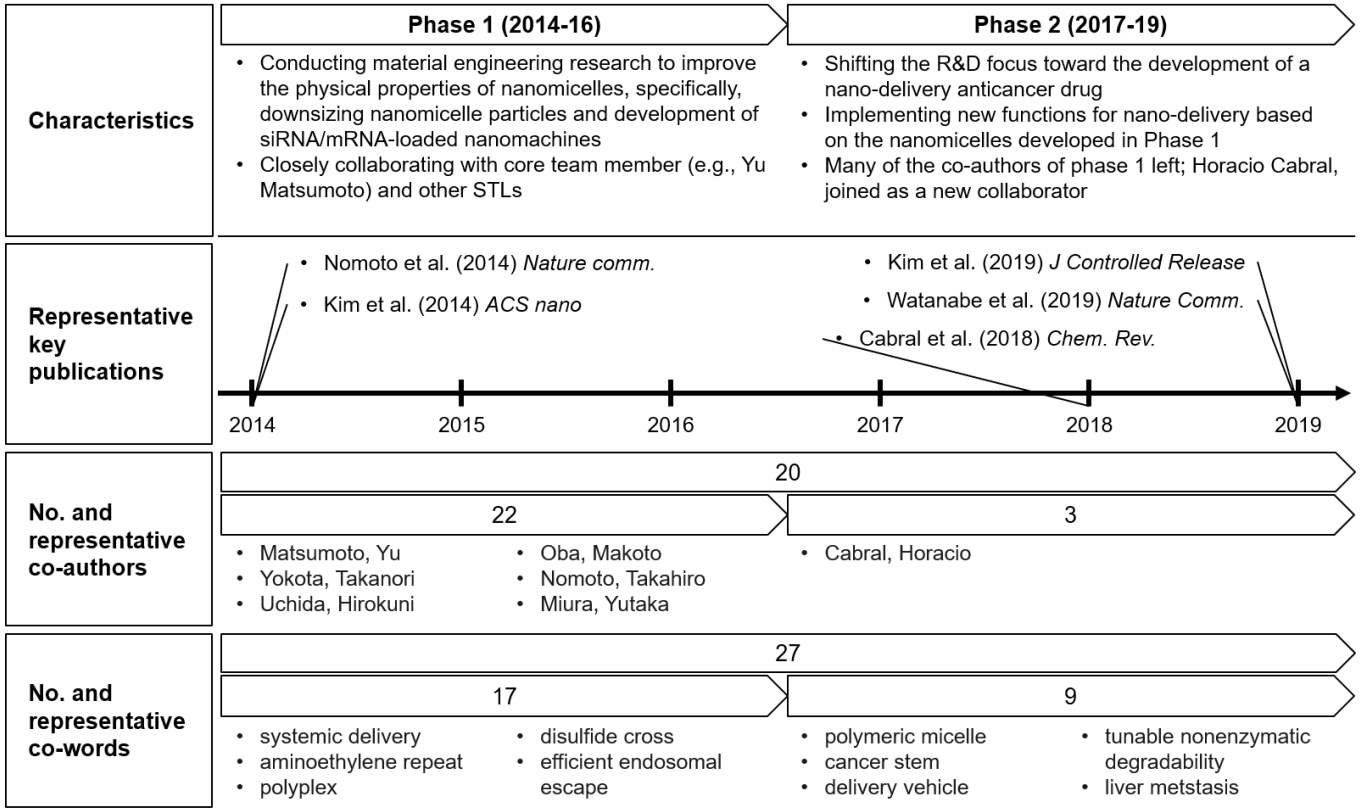

Figure 4.3 A summary of the research strategy for Phase 1 and Phase 2 of subtheme leader 1. Representative key papers were selected based on the judgement of the biology experts as remarkable research outcomes. Representative co-authors were tallied by dividing co-authors with two or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-authors illustrated in each phase were the researchers judged to have played an important role in that phase. Representative co-words were tallied by dividing co-words with three or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-words illustrated in each phase were words that were judged to be related to important outcomes in that phase.

4.3.3 Context, interventions, mechanisms, and outcomes of subtheme leader 2

Subtheme leader 2 is responsible for the R&D to improve the performance of nanomicelles, a core technology of the COINS project, in a different research unit from subtheme leader 1. In Phase 1, subtheme leader 2 succeeded in the research when his supervisor Kazunori Kataoka was at the University of Tokyo, and carried out fundamental R&D to improve the physical properties of nanomicelles (Chen, et al, 2014; Chuanoi, et al, 2014). Specifically, nanomicelles were developed and they performed some functions by transforming into polyion complex vesicles for photoinduced intracellular delivery with enhanced salt and temperature resistance. In Phase 2, subtheme leader 2 focused on R&D to break through the blood-brain barrier (Xie, et al, 2019), which is the biggest challenge for drug delivery to the brain, but there were no major changes in the organization, and a consistent team was established under Akihiro Kishimura and Kensuke Osada (e.g., Li, et al, 2017).

Context of subtheme leader 2

Being from the research leader's laboratory, project associate professor at the Department of Bioengineering, Graduate School of Engineering, the University of Tokyo, and visiting scientist at iCONM, subtheme leader 2 has a relatively similar background to subtheme leader 1. Subtheme leader 2, who was originally engaged in research for cancer too, mainly engaged in research activities related to treatment technologies for neurological diseases, such as the Alzheimer's disease.

Interventions of subtheme leader 2

Subtheme leader 2 was involved in the industrialization and commercialization as a scientific advisor to Braizon Therapeutics, a startup firm that was established using the research outcome. During the project initiation, a policy was formed wherein the outcomes of the subtheme 2 would be utilized to establish a startup for social implementation. This policy seemed to be decided due to the social appeal of the project, that is, the concept of drug delivery to the brain by establishing a firm specializing in the treatment of neurological diseases using nanobiotechnology, rather than licensing the patents of the basic technology to existing companies. After Braizon Therapeutics was established, collaboration with other subthemes has progressed.

Mechanisms of subtheme leader 2

Subtheme leader 2 is the youngest researcher among the subtheme leaders. The reason why subtheme leader 2 targeted another intractable disease, neurological disorder, rather than cancer (which is the mainstream of conventional drug delivery systems using nanomedicine), could be because subtheme leader 2 intended to establish a career path as a researcher by adopting a research strategy to avoid the risk of competition with existing research domain. The existence of a mechanism to generate such strategic behavior suggested that there were both a risk of competition and an opportunity to explore new domain, resulting in pioneering a drug delivery system for neurological

diseases and the creation of a startup that promoted the social implementation of the basic technology for the system.

Outcomes of subtheme leader 2

The major research outcome of subtheme leader 2 is a technology for polymeric nanomicelles which could pass through the blood-brain barrier. The human brain is protected by the blood-brain barrier that prevents the passage of anything except nutrients, like glucose. This barrier also makes it difficult for the drugs to reach the brain. Subtheme leader 2 succeeded in improving the amount of drug accumulation in the brain by binding glucose to the surface of polymeric nanomicelles. Subtheme leader 2 obtained this technology's patent, which is the basis for the launch of Braizon Therapeutics. This outcome was realized by the precise design technology for introducing glucose molecules to the surface of polymeric micelles in an appropriate form, which is an application of the core knowledge of the drug delivery system for cancer treatments. Therefore, the core technology, which was originally developed in the R&D of drug delivery systems for cancer diseases, was converted into peripheral knowledge for the treatment of neurological diseases, resulting in the creation of a high-impact outcome.

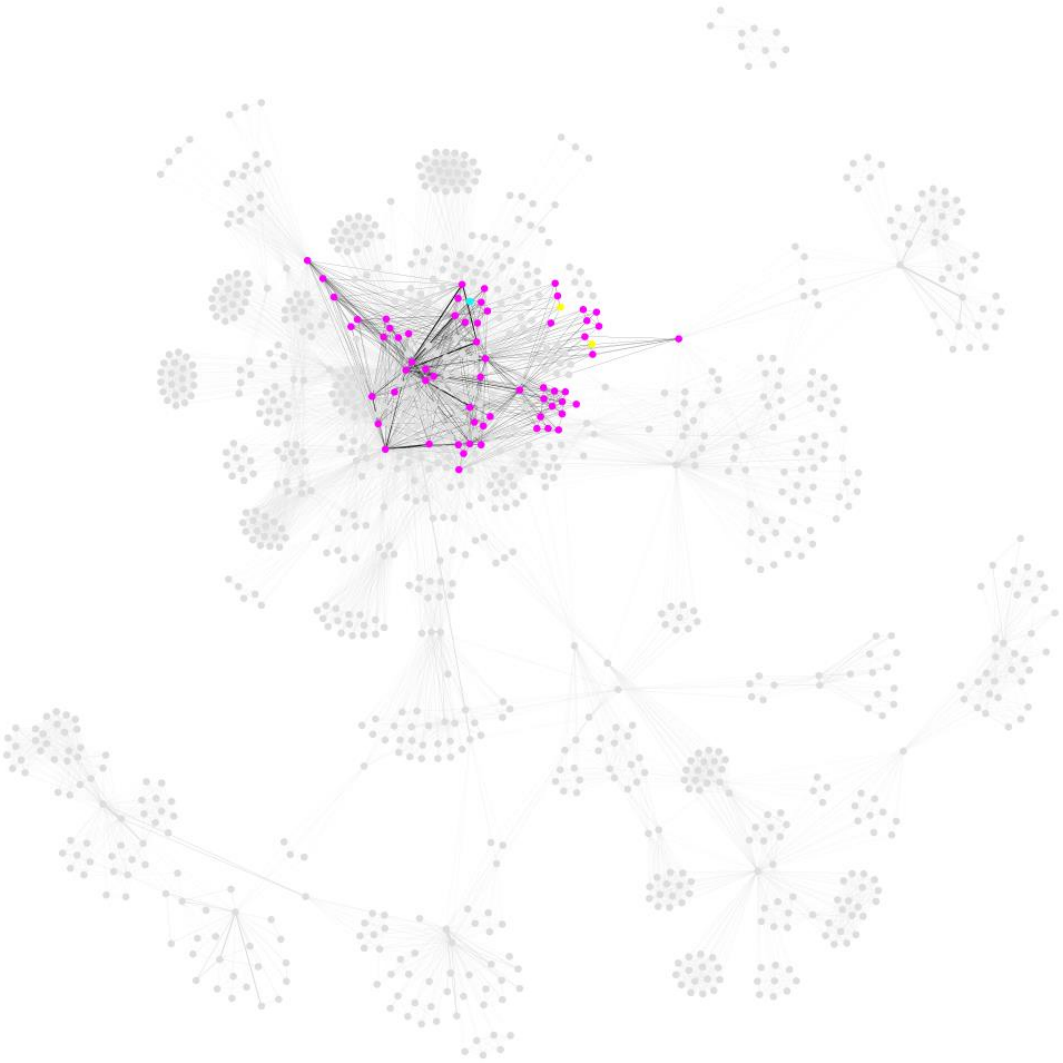

Figure 4.4 Co-author micro-network of the publications from subtheme leader 2 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

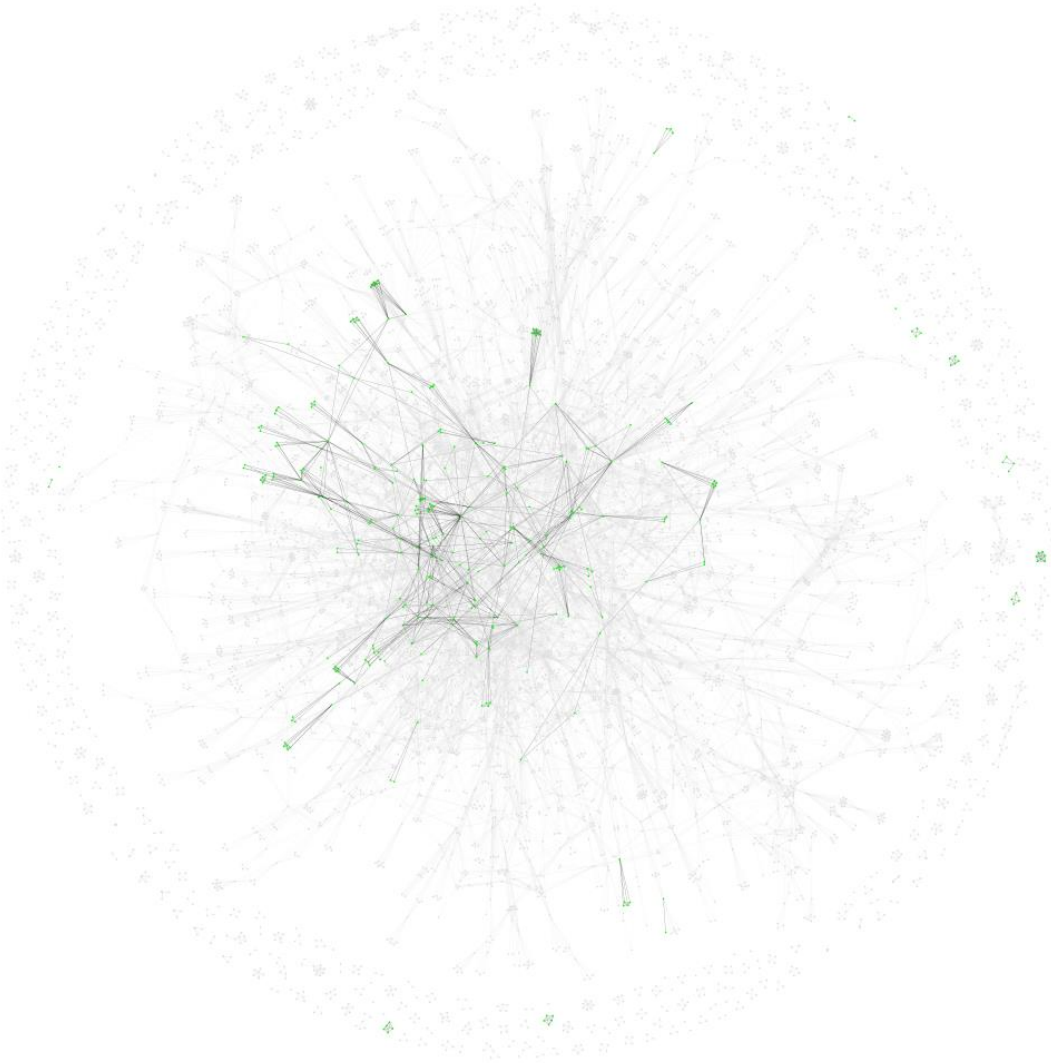

Figure 4.5 Co-word micro-network of the publications from subtheme leader 2 of COINS project.

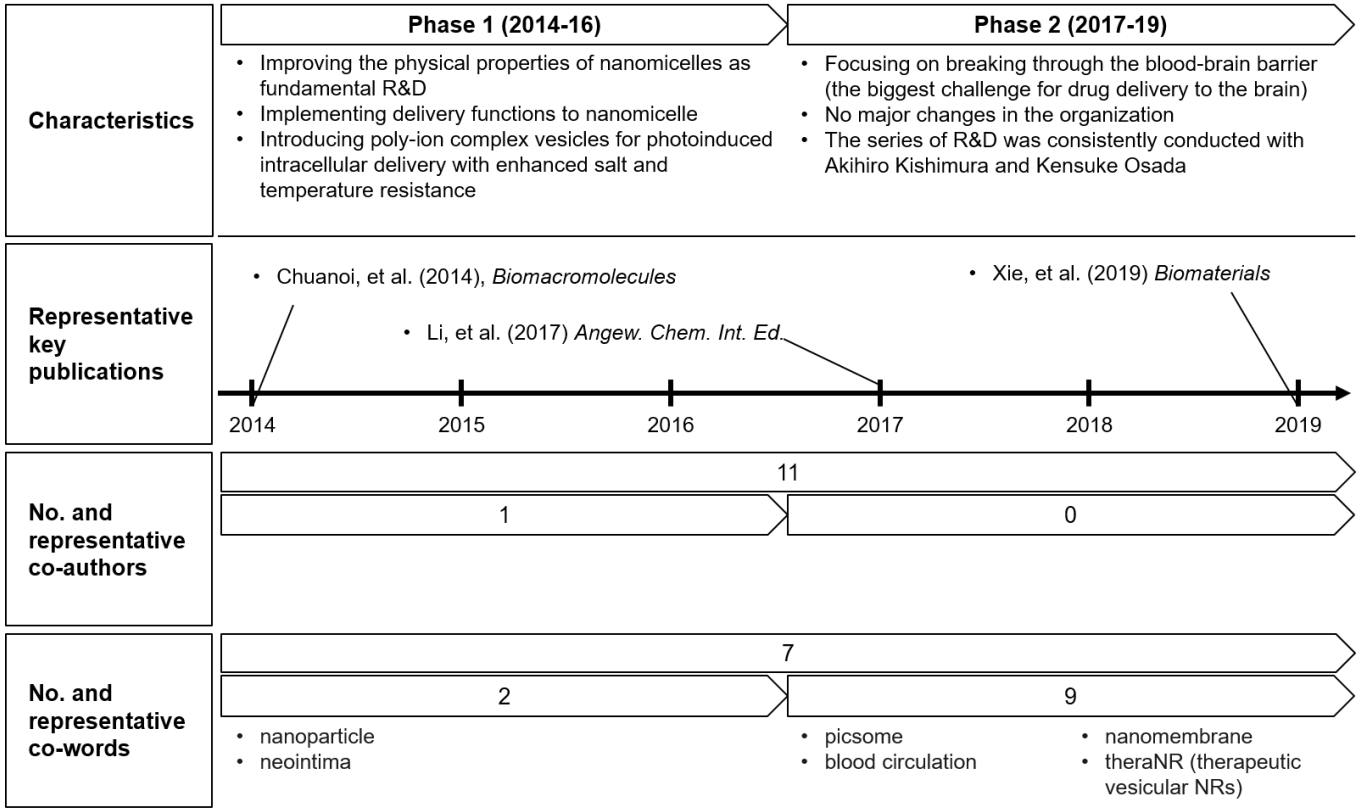

Figure 4.6 A summary of the research strategy for Phase 1 and Phase 2 of subtheme leader 2. Representative key papers were selected based on the judgement of the biology experts as remarkable research outcomes. Representative co-authors were tallied by dividing co-authors with two or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-authors illustrated in each phase were the researchers judged to have played an important role in that phase. Representative co-words were tallied by dividing co-words with three or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-words illustrated in each phase were words that were judged to be related to important outcomes in that phase.

4.3.4 Context, interventions, mechanisms, and outcomes of subtheme leader 3

In Phase 1, subtheme leader 3 narrowed down the research focus to drug delivery system development to focus on gene and nucleic acid delivery and motor-sensory organ therapy by collaborating with Hiroaki Kinoh, Takehiko Ishii, Hirokuni Uchida, and others. The focus was on the selection of polyplex micelles suitable for therapeutic purposes, the implementation of mRNA into nanomicelles, and the investigation of targeted gene delivery targeting pancreatic cancer (Ge, et al, 2014; Matsui, et al, 2015; Uchida, et al, 2016). In Phase 2, the collaborative R&D in Phase 1 was dissolved, but a team of core members led by Kazunori Kataoka, Satoshi Uchida, and Kayoko Yanagihara, was established to expand the scope of R&D to translational research, such as the design of double-stranded messenger RNA for effective vaccination and the development of treatment methods for bone defects by transplantation of spheroids (Uchida, et al, 2018; Yanagihara, et al, 2018).

Context of subtheme leader 3

Subtheme leader 3 differs from subtheme leader 1 and subtheme leader 2 in terms of the affiliation and position at iCONM. Subtheme leader 3 is a professor at the Department of Biofunction Research, Institute for Biomaterials and Bioengineering, Tokyo Medical and Dental University, and a principal research scientist at iCONM. Subtheme leader 3, who is originally a clinician of orthopedic surgery, conducted R&D in regenerative medicine in the field of orthopedics and drug discovery based on mRNA, focusing on substances to be delivered such as mRNA, rather than polymeric nanomicelles. The technology originated in drug delivery systems is used to deliver mRNA, which is an unstable substance, to its destination, the cell.

Interventions of subtheme leader 3

Subtheme leader 3 also works on the R&D of drug delivery systems for nucleic acid drugs with AccuRna, serving as the scientific advisor. In combination with other subthemes, the experience as an orthopedic surgeon and the characteristics of the affiliation are utilized in the research collaboration and R&D for clinical applications, such as regenerative medicine or mRNA drug discovery, from the viewpoint of a user of drug delivery systems.

Mechanisms of subtheme leader 3

The reason why the R&D was conducted from the viewpoint of a user of drug delivery systems could be because subtheme leader 3 originally treated patients as a clinician. Medical science (as well as management science) is oriented toward prescribing solutions to the problems currently faced and emphasizes to achieve the best outcomes using the means available at that time. Although distinguishing whether it was consciously or unconsciously done was slightly difficult, the research collaboration seemed to be conducted based on such a pattern of thinking.

Outcomes of subtheme leader 3

Subtheme leader 3 created the outcomes related to drug delivery systems using mRNA to treat intractable diseases of the cranial nerves and motor and sensory disorders. Owing to their design, mRNA drugs are also used in the field of preventive medicine such as vaccines for infectious diseases and they are expected to be used in the domain of drug delivery system in the future. Concomitantly, this new modality raises a new challenge to conventional therapeutic approaches. In other words, these outcomes were interpreted to actualize medical care beyond the existing framework, by combining new therapeutic approach using mRNA, which is the peripheral knowledge, and the drug delivery system, which is the core knowledge.

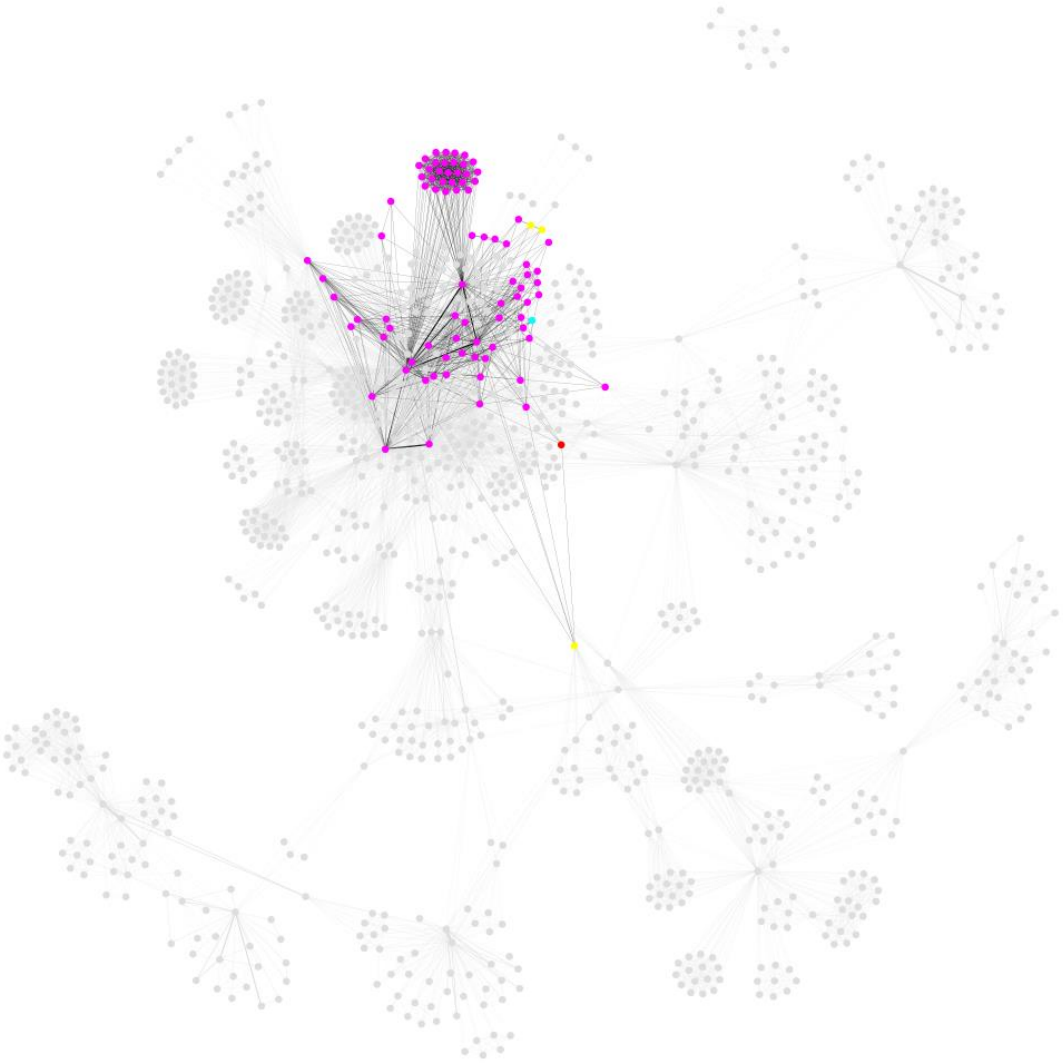

Figure 4.7 Co-author micro-network of the publications from subtheme leader 3 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

Figure 4.8 Co-word micro-network of the publications from subtheme leader 3 of COINS project.

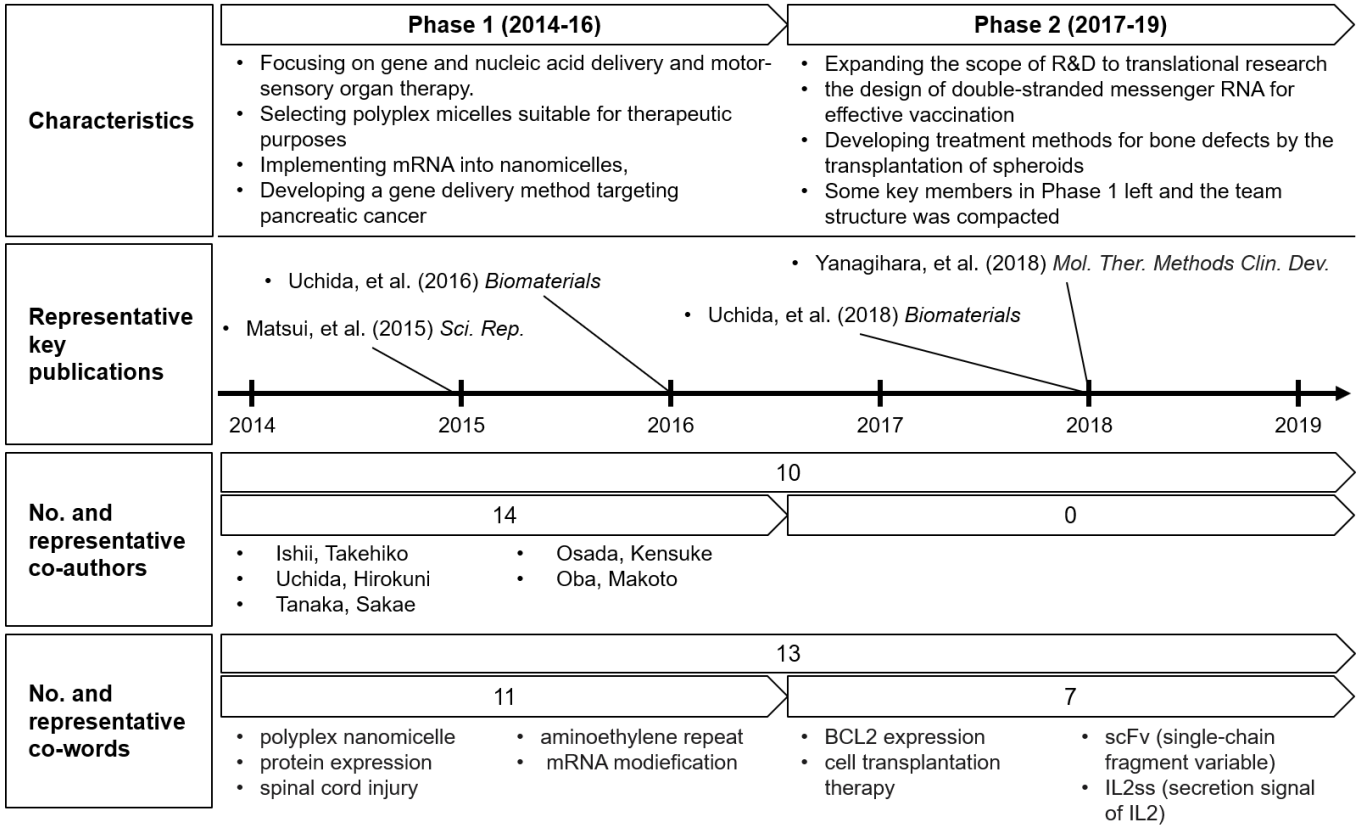

Figure 4.9 A summary of the research strategy for Phase 1 and Phase 2 of subtheme leader 3. Representative key papers were selected based on the judgement of the biology experts as remarkable research outcomes. Representative co-authors were tallied by dividing co-authors with two or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-authors illustrated in each phase were the researchers judged to have played an important role in that phase. Representative co-words were tallied by dividing co-words with three or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-words illustrated in each phase were words that were judged to be related to important outcomes in that phase.

4.3.5 Context, interventions, mechanisms, and outcomes of subtheme leader 4

In COINS, subtheme leader 4 is responsible for the peripheral area of diagnostic system development in a coherent manner. Specifically, based on microfabrication technology and heterogeneous material-integrated device engineering, the R&D of highly functional biodevices was conducted with Hiroaki Takehara, who was his laboratory staff member. In Phase 1, on-chip immunoelectrophoresis of extracellular vesicles and micro-optical fluidic device "lab-on-a-brain" were developed with Takahiro Ochiya of National Cancer Center Japan (Akagi, et al, 2015; Takehara, et al, 2014). In Phase 2, the R&D from Phase 1 was continued under a new structure with renewal of the unit and collaborative members: for example, a microfluidic model for optical detection of nanoparticles in blood (Takehara, Kanda, and Ichiki, 2018).

Context of subtheme leader 4

Subtheme leader 4 is a professor at the Department of Materials Engineering, Graduate School of Engineering, University of Tokyo, as well as a principal research scientist at iCONM. Subtheme leader 4 specializes in materials science, especially in microfabrication technology using semiconductor manufacturing processes and bioelectronics. Therefore, subtheme leader 4 has a different background from the other subtheme leaders; that is, the background is not drug delivery system using nanobiotechnology, but worked to develop diagnostic devices for early detection of cancer by measuring miRNA and exosomes in blood.

Interventions of subtheme leader 4

Unlike the other subtheme leaders, subtheme leader 4 is not from the research leader's laboratory, but has hitherto conducted collaborative research with the research leader. However, they seem to have complementary roles rather than fusion of their expertise as far as the co-author and co-word networks are observed. In addition, subtheme leader 4 work together on R&D with Nikon for a long time, but eventually new startup firm, iXstream was established as spinout. At that time, a collaborator who had conducted R&D with subtheme leader 4 was appointed as the CEO.

Mechanisms of subtheme leader 4

Subtheme leader 4 adopted a different research strategy from the other subtheme leaders. It aimed to realize "smart life care society" and "in-body hospital" vision through diagnosis for early detection rather than treatment of diseases with drug delivery systems. Subtheme leader 4 selected research strategy that creates complementary knowledge, did not adopt the direction to develop core knowledge by merging the expertise with drug delivery systems, that is, a research strategy similar to that of subtheme leader 3. The reason why subtheme leader 4 selected this research strategy could be because there are mechanisms to adjust interest, that is, which entities would address social implementation of the created outcomes. The company (Nikon) with which subtheme leader 4 initially

collaborated had the strength of peripheral knowledge rather than drug delivery systems. Had the company proceeded with the industrialization and commercialization of the research outcomes, focusing on the creation of peripheral knowledge would have been a rational choice for this company. In fact, the research topic of subtheme leader 4 got closer to core knowledge after the company dropped the industrialization and commercialization of the research outcomes and established iXstream, a new entity for social implementation.

Outcomes of subtheme leader 4

Subtheme leader 4 created outcomes of automation on the processes for miRNA analysis, such as the isolation, purification, and detection of miRNA in exosomes contained in bodily fluids like blood, by developing integrated biodevice for both early detection and minimally invasive diagnosis of cancer using microfluidic device technology. These outcomes did not come under the core knowledge of nanobiotechnology related to drug delivery systems, and to achieve the peripheral knowledge to realize the “smart life care society” and “in-body hospital” vision, not only treatment (main focus of other subthemes) but also diagnosis would be required. Subtheme leader 4 established a startup firm called iXstream using the outcomes. This startup is also oriented to medical applications of exosome evaluation technology, which is equivalent to peripheral knowledge, rather than drug delivery systems.

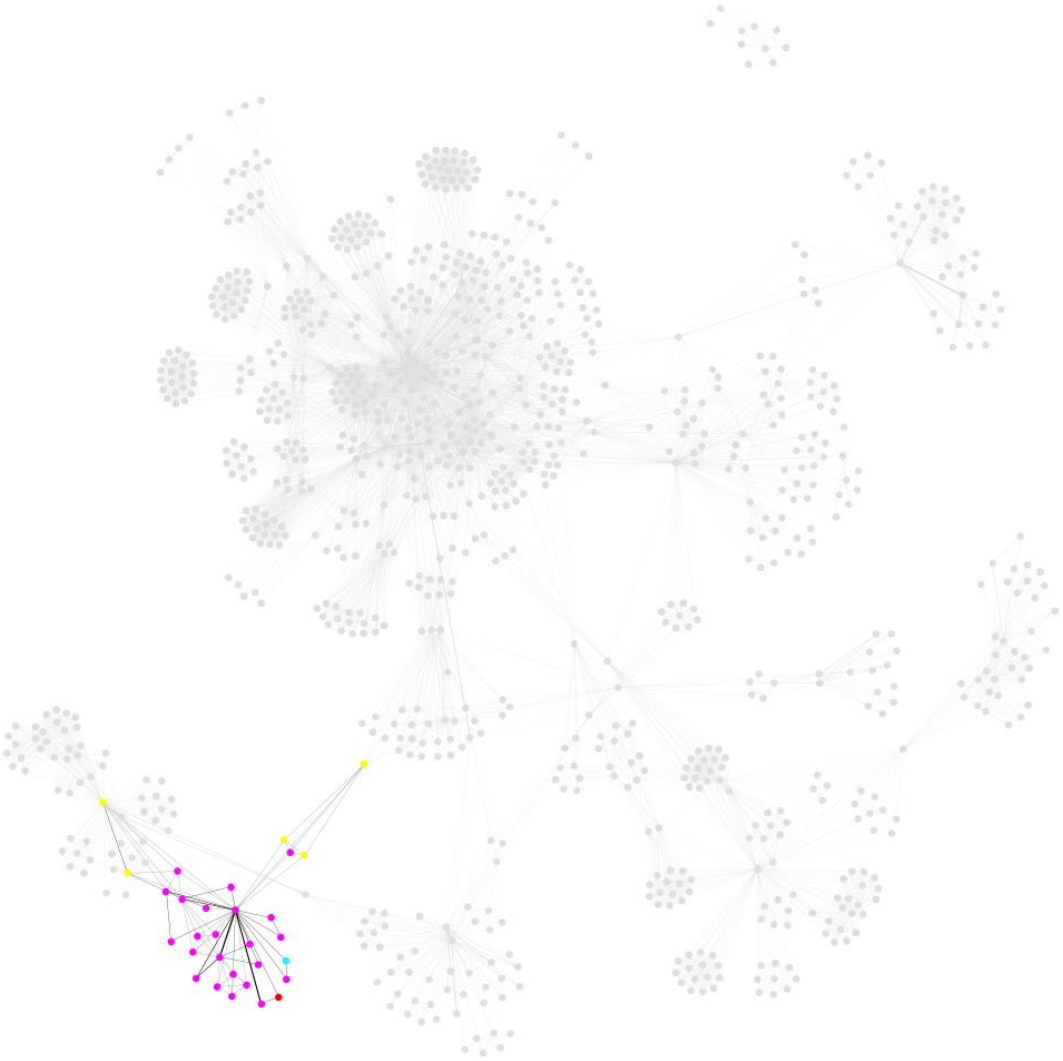

Figure 4.10 Co-author micro-network of the publications from subtheme leader 4 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

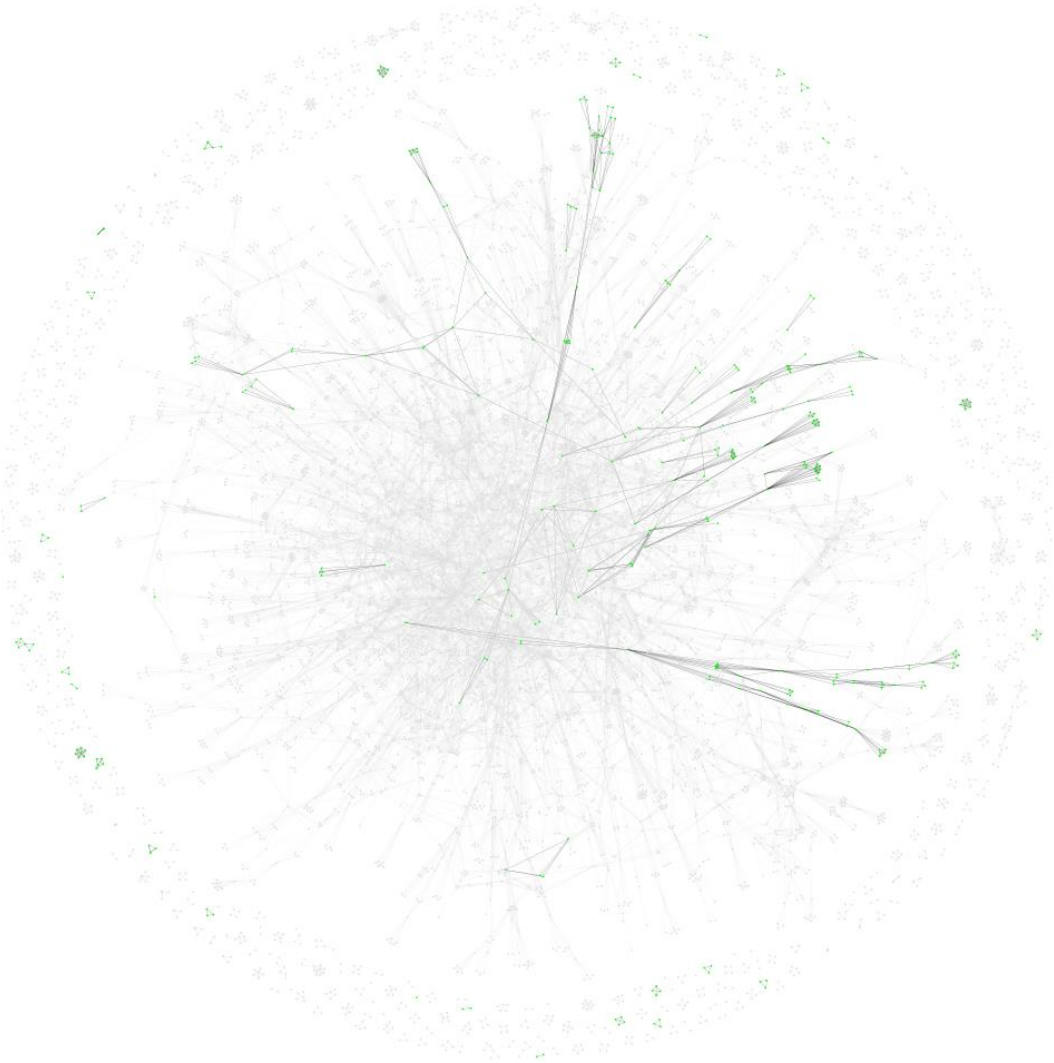

Figure 4.11 Co-word micro-network of the publications from subtheme leader 4 of COINS project.

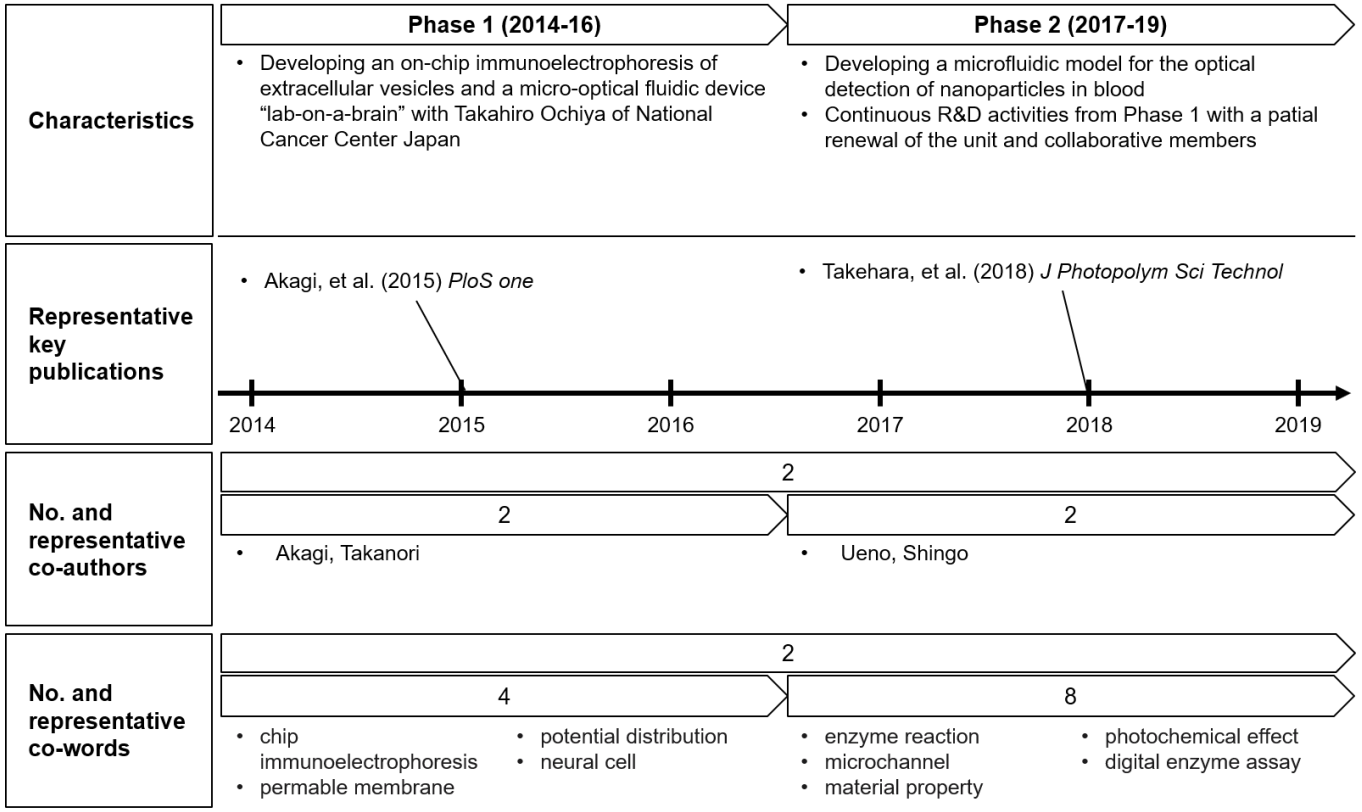

Figure 4.12 A summary of the research strategy for Phase 1 and Phase 2 of subtheme leader 4. Representative key papers were selected based on the judgement of the biology experts as remarkable research outcomes. Representative co-authors were tallied by dividing co-authors with two or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-authors illustrated in each phase were the researchers judged to have played an important role in that phase. Representative co-words were tallied by dividing co-words with three or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-words illustrated in each phase were words that were judged to be related to important outcomes in that phase.

4.3.6 Context, interventions, mechanisms, and outcomes of subtheme leader 5

Subtheme leader 5 led the largest research unit, consisting of core members such as Kazunori Kataoka (research leader), Horacio Cabral (University of Tokyo), and Takahiro Nomoto (Assistant Professor of his own laboratory), to develop pinpoint diagnosis and treatment technology by integrating the nanomachines that deliver drugs to lesions, a core technology of the COINS project, with medical devices such as MRI and array transducers. In Phase 1, subtheme leader 5 developed a new technology for the delivery of siRNA to target solid tumors and lymph node metastasis using poly-ion complex micelles and a technology for the activation of drug-loaded polymeric micelles by light response and pH change with Kanjiro Miyata (subtheme leaders 1), Yu Matsumoto, and Yutaka Miura in addition to the core members mentioned above (Oe, et al, 2014; Yen, et al, 2014; Cabral, et al, 2015; Tangsangasaksri, et al, 2016). However, in Phase 2, many of the collaborative R&D activities in Phase 1 were ended or reduced, subtheme leader 5 and the core members expanded the scope of R&D to translational research. Specifically, they conducted multifaceted R&D toward clinical applications, such as signal amplification for non-invasive imaging of tumor malignancy, vascular burst technology to increase the permeability of tumor blood vessels, boron neutron capture therapy for solid tumors, and the development of in vivo rendezvous methods to target cancer cells (Mi, et al, 2016; Matsumoto, et al, 2016; Mi, et al, 2017; Watanabe, et al2019). Moreover, a new collaborative R&D group was initiated under Yoshihiro Muragaki at Tokyo Women's Medical University to develop sonodynamic therapy using nanomicelle technology and high-intensity focused ultrasound in cancer treatment (Horise, et al, 2019).

Context of subtheme leader 5

Subtheme leader 5 had been conducting research on polymeric nanomicelles under the research leader, and currently has a laboratory as a professor at the Laboratory for Chemical and Life Science, Institute of Innovative Research, Tokyo Institute of Technology. Like subtheme leader 3 or subtheme leader 4, subtheme leader 5 is also a principal research scientist at iCONM. Subtheme leader 5 aims to realize minimally invasive treatment systems that enable one-day treatment with less burden on patients, and imaging systems that can visualize the behavior of nanomachines in the body, by combining drug delivery systems using polymeric nanomicelles with medical devices. Although this research topic is partially overlapped with subtheme leader 1 in the context of the R&D on drug delivery systems using polymeric nanomicelles encapsulated anticancer drugs, subtheme leader 5 conducted the R&D aiming to develop the knowledge base in a different direction—the application as an alternative to the existing contrast agents for imaging.

Interventions of subtheme leader 5

As the research topics of subtheme leader 5 are related to all nanomachine research in other subthemes, the collaboration with other subthemes was mainly focused on, as well as the case of

subtheme leader 1. A unique feature of subtheme leader 5 is the collaboration with the National Institute of Radiological Sciences and the National Institutes for Quantum and Radiological Science and Technology, which are engaged in R&D on imaging. In addition, research collaboration for clinical application is also conducted.

Mechanisms of subtheme leader 5

The research topics handled by subtheme leader 5 are closely related to the other subthemes of this project. Therefore, subtheme leader 5 plays a significant collaborative role in accelerating the process of producing and improving the core knowledge that may become the fundamental technology for drug delivery systems, by creating a transaction-free zone (Baldwin, 2008) with actors who share expertise, such as research leader or subtheme leader 1. The selection of such a form of collaboration suggested that there are hidden mechanisms that influence the form of collaboration by the type of produced knowledge. In this case, as the knowledge reproduced by using the produced core knowledge also becomes core knowledge, a transaction-free zone is formed which accelerates the knowledge production cycle and continues the incremental improvement of the knowledge base of the drug delivery system.

Outcomes of subtheme leader 5

Subtheme leader 5 created outcomes that enhance the safety and effectiveness of treatment by combining the drug delivery system with medical equipment such as MRI to realize one-day treatment of cancer, which is a key component of the vision of an "in-body hospital." In particular, the outcomes related to imaging not only contribute to the realization of precision medicine, but also to improvement of the drug delivery system itself, by clarifying nanomachines' behavior in the body to deepen the understanding of the dynamics of nanomachines. In other words, subtheme leader 5 created core knowledge as outcomes to reproduce novel core knowledge.

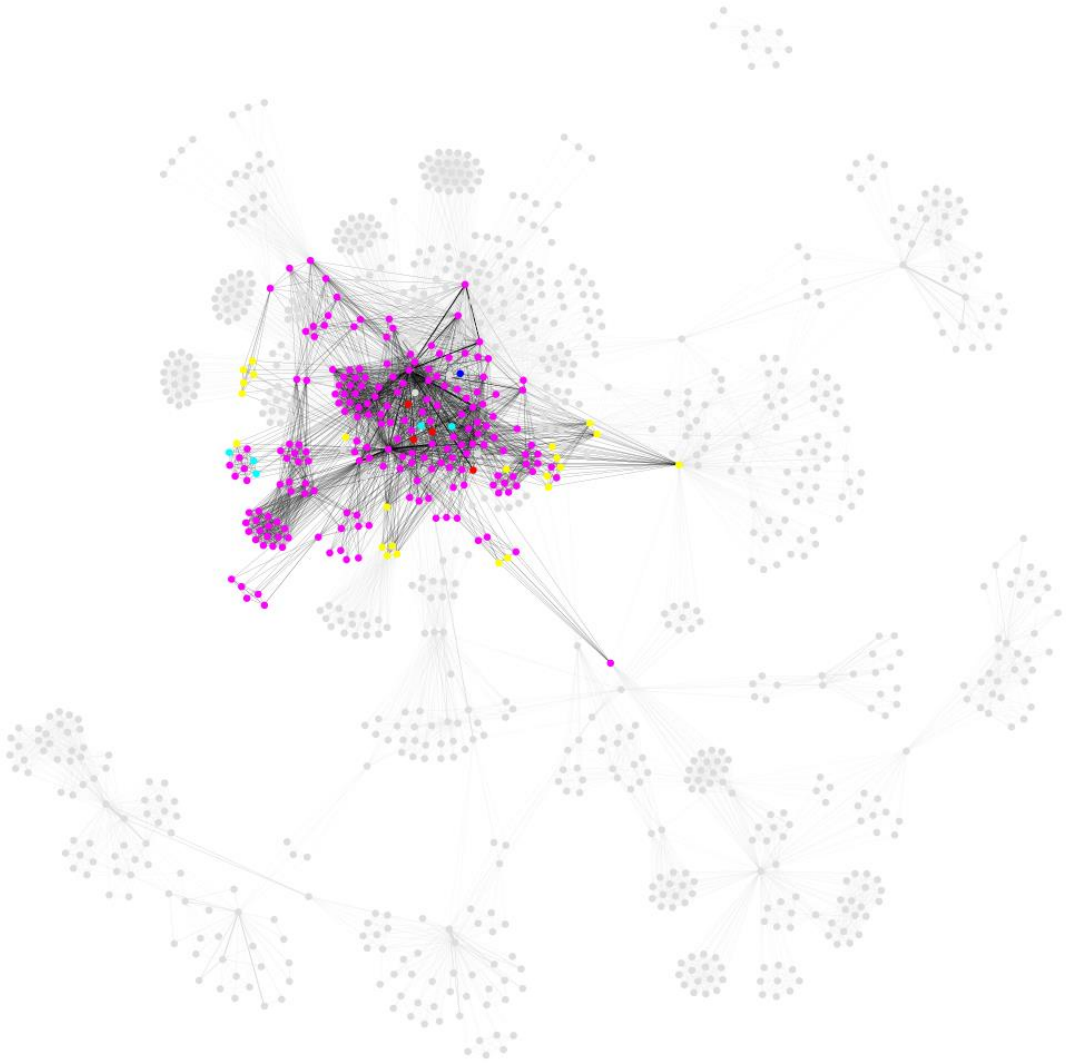

Figure 4.13 Co-author micro-network of the publications from subtheme leader 5 of COINS project. The color of a node represents its attributes: purple, university; red, core organization (iCONM); light blue, industry; yellow, public institute; blue, no data.

Figure 4.14 Co-word micro-network of the publications from subtheme leader 5 of COINS project.

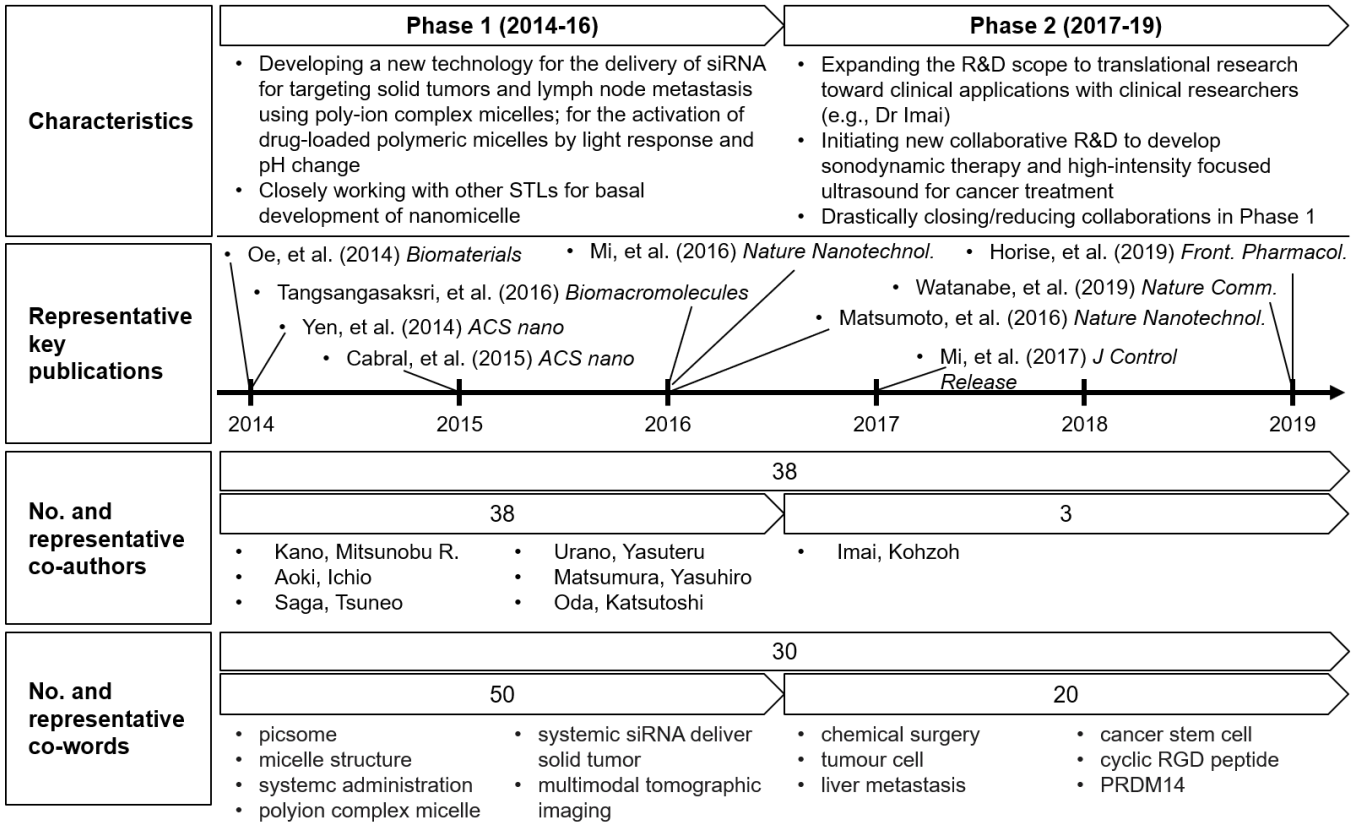

Figure 4.15 A summary of the research strategy for Phase 1 and Phase 2 of subtheme leader 5. Representative key papers were selected based on the judgement of the biology experts as remarkable research outcomes. Representative co-authors were tallied by dividing co-authors with two or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-authors illustrated in each phase were the researchers judged to have played an important role in that phase. Representative co-words were tallied by dividing co-words with three or more occurrences into three categories: only in Phase 1, only in Phase 2, and in both phases. The co-words illustrated in each phase were words that were judged to be related to important outcomes in that phase.

4.4 Validation based on the R&D Strategy

4.4.1 Classification based on type of knowledge of knowledge producers and knowledge consumers

The results suggest that the COINS subtheme leaders are likely to adopt different research strategies while conducting R&D to achieve the common vision "in-body hospital." Different research strategies may have been chosen owing to the different relationship between the production and consumption of knowledge of each subtheme leader. That is, they are considered to be classified based on whether it is core knowledge or peripheral knowledge that is produced or consumed.

4.4.2 Validation based on real-world R&D strategies: Quadrant of core knowledge producers and core knowledge consumers

The fact that both produced and consumed knowledge are core knowledge implies that information asymmetry caused by expertise between knowledge producers and knowledge consumers hardly exist, and smooth communication on advanced research topics allows rapid rotation of the R&D cycle, that is, increased efficiency of R&D. In other words, in this quadrant, utilizations of collaboration among actors with shared expertise, especially the alumni network, as a mechanism to accelerate the reproduction of core knowledge was implied to establish the knowledge base necessary to realize the vision of "in-body hospital" using nanomachines. This quadrant includes subtheme leaders 1 and 5, who were from research leader's laboratory, and the specific research activities conducted by these subtheme leaders are described below.

The subthemes of the subtheme leaders included in this quadrant are characterized by the fact that the normalized graph entropy of the co-author network continues to decrease, but the normalized graph entropy of the co-word network increases after decreasing once. As the subtheme leaders in this quadrant have implemented core knowledge enhancement such as the development of nanomicelles in Phase 1 and then conducted R&D for clinical applications in Phase 2, the complexity of the knowledge structure has decreased as a result of focusing on specific research topics in Phase 1; however, it is considered that the relative complexity of the knowledge structure increased due to the expansion of research topics in Phase 2. As for the collaboration structure, the decrease in relative complexity slowed down due to the entry of some new actors, but the overall trend is considered to have continued to decrease.

4.4.3 Validation based on real-world R&D strategies: Quadrant of core knowledge producers and peripheral knowledge consumers

This quadrant has a feature that core knowledge is produced while peripheral knowledge is consumed. To realize the COINS project's vision of "in-body hospital," not only the existing core knowledge of nanomachines, especially nanomicelles for the treatment of cancer, but also the core knowledge base that will extend the function of treatment for other diseases, such as Alzheimer's disease, should be

produced. Therefore, this quadrant included subtheme leader 2, who conducted R&D specific to nanomicelle delivery systems for drug delivery to the brain, a knowledge base for the treatment of non-cancer diseases that is consumed as peripheral knowledge.

The subtheme of the subtheme leader included in this quadrant is as follows: the normalized graph entropy of the co-author network keeps decreasing and the normalized graph entropy of the co-word network keeps increasing. If the summary of the research strategy of this subtheme leader was checked, he continues with the challenging research topic of drug delivery to the brain by breaking through the blood-brain barrier. On the contrary, there has been little change in the collaboration structure and the ties with specific members have become stronger. Therefore, it is thought that the relative clutter of the collaboration structure continues to decrease while the clutter of the knowledge structure continues to increase.

4.4.4 Validation based on real-world R&D strategies: Quadrant of peripheral knowledge producers and core knowledge consumers

In this quadrant, contrary to the case of previous subsections, peripheral knowledge is produced while core knowledge is consumed. That is, it corresponds to addressing the question of how to utilize core knowledge to produce peripheral knowledge to achieve social implementation of the medical concept "in-body hospital." This quadrant includes subtheme leader 3, who conducted R&D of nanoreconstruction technology for the maintenance and regeneration of biological functions in the COINS and has a different background from the other subtheme leaders as an orthopedic surgeon.

The subtheme of the subtheme leader included in this quadrant is characterized by the fact that the normalized graph entropy of the co-author network continues to decrease but the normalized graph entropy of the co-word network increases after decreasing once, as in the case of the quadrant of core knowledge producers and core knowledge consumers. Hence, the complexity of the relative knowledge structure decreased in Phase 1 but increased in Phase 2 as a result of the implementation of the reinforcement of core knowledge in Phase 1 and then the R&D for clinical application in Phase 2. The interpretation of the collaboration structure is similar, and it is considered that the complexity of the relative collaboration structure continued to decrease in Phase 2 because there was no entry of new actors in Phase 2.

4.4.5 Validation based on real-world R&D strategies: Quadrant of peripheral knowledge producers and peripheral knowledge consumers

The fact that there are cases where none of the knowledge produced and consumed is core knowledge, that is, both are peripheral knowledge may seem strange at first glance. However, to achieve social implementation of systems with diverse functions, considering only the production and consumption of core knowledge can be insufficient. In this context, COINS' vision is "in-body hospital," but the functions of a hospital are not limited to the treatment of diseases. Furthermore, to realize the vision

of not only an “in-body hospital” but also a “smart life care society,” proposed in COINS, incorporating knowledge bases of the existing core technologies of nanomachines or drug delivery systems, that are not related (at present), is necessary. In fact, subtheme leader 4, who is in this quadrant, focused on the development of complementary knowledge bases that will be needed to achieve the “in-body hospital”—early diagnosis of cancer using exosomes and miRNAs, rather than the therapies that focus on other subthemes.

The trajectory of the normalized graph entropy of the subtheme leaders in this quadrant shows that the normalized graph entropy of the co-word tends to increase, while but the normalized graph entropy of the co-author network decreases after increasing. This change in the complexity of the knowledge structure and collaboration structure can be explained by the fact that the research topic itself is consistent between Phase 1 and Phase 2, but the collaboration style has changed significantly. This may reflect the fact that the research and development system has changed significantly.

Core-periphery profile		Knowledge consumer (Exploitation)	
		Core knowledge	Peripheral knowledge
Knowledge producer (Exploration)	Core knowledge	Acceleration of R&D to establish knowledge base for project vision “in-body hospital” by nanomachine Subtheme leader 1 Subtheme leader 5	Utilization of core knowledge base to expand functions of "in-body hospital" Subtheme leader 2
	Peripheral knowledge	Novel medical care approach by "in-body hospital" concept using peripheral knowledge base Subtheme leader 3	Another approach to achieve vision "in-body hospital" by development of complementary knowledge base Subtheme leader 4

Figure 4.16. Classification based on type of knowledge of knowledge producers and knowledge consumers on the core-periphery profile.

References

Akagi, T., Kato, K., Kobayashi, M., Kosaka, N., Ochiya, T., & Ichiki, T. (2015). On-chip immunoelectrophoresis of extracellular vesicles released from human breast cancer cells. *PloS one* 10(4):e0123603.

Baldwin, C. Y. (2008). "Where do transactions come from? Modularity, transactions, and the boundaries of firms". *Industrial and Corporate Change* 17(1):155–195.

Barnett-Page, E., & Thomas J. (2009). "Methods for the synthesis of qualitative research: A critical review". *BMC Medical Research Methodology* 9(1): 1-11.

Cabral, H., Makino, J., Matsumoto, Y., Mi, P., Wu, H., Nomoto, T., Toh, K., Yamada, N., Higuchi, Y., Konishi, S., Kano, M. R., Nishihara, H., Miura, Y., Nishiyama, N., & Kataoka, K. (2015). Systemic targeting of lymph node metastasis through the blood vascular system by using size-controlled nanocarriers. *ACS nano* 9(5):4957-4967.

Cabral, H., Miyata, K., Osada, K., & Kataoka, K. (2018). Block copolymer micelles in nanomedicine applications. Chemical reviews, 118(14), 6844-6892.

Chen, H., Xiao, L., Anraku, Y., Mi, P., Liu, X., Cabral, H., Inoue, A., Nomoto, T., Kishimura, A., Nishiyama, N., & Kataoka, K. (2014). Polyion complex vesicles for photoinduced intracellular delivery of amphiphilic photosensitizer. *Journal of the American Chemical Society* 136(1):157-163.

Chuanoi, S., Anraku, Y., Hori, M., Kishimura, A., & Kataoka, K. (2014). Fabrication of polyion complex vesicles with enhanced salt and temperature resistance and their potential applications as enzymatic nanoreactors. *Biomacromolecules* 15(7):2389-2397.

Denyer, D., Tranfield, D., & Van Aken, J. E. (2008). "Developing design propositions through research synthesis". *Organization Studies* 29(3):393–413.

Ge, Z., Chen, Q., Osada, K., Liu, X., Tockary, T. A., Uchida, S., Dirisala, A., Ishii, T., Nomoto, T., Toh, K., Matsumoto, Y., Oba, M., Kano, M. R., Itaka, K., & Kataoka, K. (2014). Targeted gene delivery by polyplex micelles with crowded PEG palisade and cRGD moiety for systemic treatment of pancreatic tumors. *Biomaterials* 35(10):3416-3426.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). "Meta-analysis and the science of research syn-thesis". *Nature* 555(7695):175–182.

Horise, Y., Maeda, M., Konishi, Y., Okamoto, J., Ikuta, S., Okamoto, Y., Ishii, H., Yoshizawa, S., Umemura, S., Ueyama, T., Tamano, S., Sofuni, A., Takemae, K., Masamune, K., Iseki, H., Nishiyama, N., Kataoka, K., & Muragaki, Y. (2019). Sonodynamic therapy with anticancer micelles and high-intensity focused ultrasound in treatment of canine cancer. *Frontiers in pharmacology* 10:545.

Kim, H. J., Takemoto, H., Yi, Y., Zheng, M., Maeda, Y., Chaya, H., Hayashi, K., Mi, P., Pittella, F., Christie, R. J., Toh, K., Matsumoto, Y., Nishiyama, N., Miyata, K., & Kataoka, K. (2014). Precise engineering of siRNA delivery vehicles to tumors using polyion complexes and gold

nanoparticles. *ACS nano* 8(9):8979-8991.

Li, J., Dirisala, A., Ge, Z., Wang, Y., Yin, W., Ke, W., Toh, K., Xie, J., Matsumoto, Y., Anraku, Y., Osada, K., & Kataoka, K. (2017). Therapeutic Vesicular Nanoreactors with Tumor-Specific Activation and Self-Destruction for Synergistic Tumor Ablation. *Angewandte Chemie*, 129(45):14213-14218.

Matsui, A., Uchida, S., Ishii, T., Itaka, K., & Kataoka, K. (2015). Messenger RNA-based therapeutics for the treatment of apoptosis-associated diseases. *Scientific reports* 5(1):1-10.

Matsumoto, Y., Nichols, J. W., Toh, K., Nomoto, T., Cabral, H., Miura, Y., Christie, R. J., Yamada, N., Ogura, T., Kano, M. R., Matsumura, Y., Nishiyama, N., Yamasoba, T., Bae,Y . H., & Kataoka, K. (2016). Vascular bursts enhance permeability of tumour blood vessels and improve nanoparticle delivery. *Nature nanotechnology* 11(6):533-538.

Mi, P., Kokuryo, D., Cabral, H., Wu, H., Terada, Y., Saga, T., Aoki, I., Nishiyama, N., & Kataoka, K. (2016) A pH-activatable nanoparticle with signal-amplification capabilities for non-invasive imaging of tumour malignancy. *Nature nanotechnology* 11(8):724-730.

Mi, P., Yanagie, H., Dewi, N., Yen, H. C., Liu, X., Suzuki, M., Sakurai, Y., Ono, K., Hiroyuki T., Cabral, H., Kataoka, K., & Nishiyama, N. (2017). Block copolymer-boron cluster conjugate for effective boron neutron capture therapy of solid tumors. *Journal of Controlled Release* 254:1-9.

Nomoto, T., Fukushima, S., Kumagai, M., Machitani, K., Matsumoto, Y., Oba, M., Miyata, K., Osada, K., Nishiyama, N., & Kataoka, K. (2014). Three-layered polyplex micelle as a multifunctional nanocarrier platform for light-induced systemic gene transfer. *Nature communications* 5(1):1-10.

Oe, Y., Christie, R. J., Naito, M., Low, S. A., Fukushima, S., Toh, K., Miura, Y., Matsumoto, Y., Nishiyama, N., Miyata, K., & Kataoka, K. (2014). Actively-targeted polyion complex micelles stabilized by cholesterol and disulfide cross-linking for systemic delivery of siRNA to solid tumors. *Biomaterials* 35(27):7887-7895.

Pawson, R. (2002). "Evidence-based Policy: The Promise of 'Realist Synthesis'". *Evaluation* 8(3):340–358.19

Rousseau, D. M. (2006). "Is There Such a Thing As "Evidence-based Management"?" *Academy of Management Review* 31(2):256–269.

Takehara, H., Kanda, Y., & Ichiki, T. (2018). Microfluidic model for optical detection of nanoparticles in whole blood. *Journal of Photopolymer Science and Technology* 31(1):59-63.

Takehara, H., Nagaoka, A., Noguchi, J., Akagi, T., Kasai, H., & Ichiki, T. (2014). Lab-on-a-brain: Implantable micro-optical fluidic devices for neural cell analysis in vivo. *Scientific reports* 4(1):1-6.

Tangsangasaksri, M., Takemoto, H., Naito, M., Maeda, Y., Sueyoshi, D., Kim, H. J., Miura, Y., Ahn, J., Azuma, R., Nishiyama, N., Miyata, K., & Kataoka, K. (2016). siRNA-loaded polyion complex micelle decorated with charge-conversional polymer tuned to undergo stepwise response to intra-tumoral and intra-endosomal pHs for exerting enhanced RNAi efficacy. *Biomacromolecules*

17(1):246-255.

Tranfield, D., Denyer, D., & Smart, P. (2003). “Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review”. *British Journal of Management* 14:207–222.

Uchida, S., Kinoh, H., Ishii, T., Matsui, A., Tockary, T. A., Takeda Machitani, K., Uchida, H., Osada, K., Itaka, K., & Kataoka, K. (2016). Systemic delivery of messenger RNA for the treatment of pancreatic cancer using polyplex nanomicelles with a cholesterol moiety. *Biomaterials* 82:221-228.

Uchida, S., Yoshinaga, N., Yanagihara, K., Yuba, E., Kataoka, K., & Itaka, K. (2018). Designing immunostimulatory double stranded messenger RNA with maintained translational activity through hybridization with poly A sequences for effective vaccination. *Biomaterials* 150:162-170.

Van Aken, J. E. (2004). “Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules”. *Journal of Management Studies* 41(2):219–246.

Van Aken, J. E. & Romme G. (2009). “Reinventing the future: Adding design science to the repertoire of organization and management studies”. *Organisation Management Journal* 6(1):5–12.

Watanabe, S., Hayashi, K., Toh, K., Kim, H. J., Liu, X., Chaya, H., Fukushima, S., Katsushima, K., Kondo, Y., Uchida, S., Ogura, S., Nomoto, T., Takemoto, H., Cabral, H., Kinoh, H. Tanaka, H. Y., Kano, M. R., Matsumoto, Y., Fukuhara, H., Uchida, S., Nangaku, M., Osada, K., Nishiyama, N., Miyata, K., & Kataoka, K. (2019). In vivo rendezvous of small nucleic acid drugs with charge-matched block catiomers to target cancers. *Nature communications* 10(1):1-13.

Xie, J., Shen, Z., Anraku, Y., Kataoka, K., & Chen, X. (2019). Nanomaterial-based blood-brain-barrier (BBB) crossing strategies. *Biomaterials* 224:119491.

Yanagihara, K., Uchida, S., Ohba, S., Kataoka, K., & Itaka, K. (2018). Treatment of bone defects by transplantation of genetically modified mesenchymal stem cell spheroids. *Molecular Therapy-Methods & Clinical Development* 9:358-366.

Yen, H. C., Cabral, H., Mi, P., Toh, K., Matsumoto, Y., Liu, X., Koori, H., Kim, A., Miyazaki, K., Miura, Y., Nishiyama, N., & Kataoka, K. (2014). Light-induced cytosolic activation of reduction-sensitive camptothecin-loaded polymeric micelles for spatiotemporally controlled in vivo chemotherapy. *ACS nano* 8(11):11591-11602.

Yi, Y., Kim, H. J., Zheng, M., Mi, P., Naito, M., Kim, B. S., Min, H. S., Hayashi, K., Perche, F., Toh, K., Liu, X., Mochida, Y., Kinoh, H., Cabral, H., Miyata, K., & Kataoka, K. (2019). Glucose-linked sub-50-nm unimer polyion complex-assembled gold nanoparticles for targeted siRNA delivery to glucose transporter 1-overexpressing breast cancer stem-like cells. *Journal of Controlled Release* 295:268-277.

CHAPTER 5. Conclusions

5.1 Summary of key findings

The study centers on the developments of methodology to realize the instrumentation of structural features on research activities in basic science research. In particular, two system structures in large-scale interdisciplinary R&D programs and projects that include basic science research are focused on: collaboration structure, which represents the cooperative relationship among actors involved in scientific research, and the knowledge structure, which represents the co-occurrence relationship among ideas used in scientific research. Realizing the instrumentation of organizational and managerial phenomena in scientific research activities enable us to measure data that can be used for strategic and organizational management of scientific research, by constructing ex ante activity indicators for scientific research rather than ex post performance indicators. Therefore, the main focus of this study is to support the prescription of solutions to organizational and managerial problems in the scientific research activity through the development of methods for the characterization and quantification of organizational and managerial phenomena, which would allow the data measurement for the management of scientific research.

In Chapter 2, program-level analysis for the FIRST Program and project-level analysis for NanoBio First to investigate inter-organizational collaboration in large-scale R&D projects are discussed in detail. In the program-level analysis, statistical modeling using data from 30 sites in the FIRST Program yield two major findings: First, an increase in research funding contributes to an increase in academic outcomes on a quantitative basis, owing to the adequacy of financial resources to hire researchers such as post-doctoral fellows who are actually engaged in activities to produce academic outcomes, but the contribution to the creation of intellectual property is unclear. Second, financial support alone will not lead to the creation of intellectual property, as different mechanisms might work in patent application and registration. This iterates the need for more explicit measures to reward creation of intellectual property, such as the design of incentive mechanisms, considering the results of the project-level analysis. For the project-level analysis, by adopting mixed-methods approach for the case study, that is, combining quantitative analysis based on the centrality indices calculated from co-author network of papers and co-inventor network of patents created in NanoBio First and a semi-structured interview with project leader, the factors to promote or impede the creation of R&D outcomes were investigated. The results of the quantitative analysis suggest that the co-author network of papers and co-inventor network of patents had a core/periphery structure, forming a transaction-free zone inside the core, achieving both efficiency and diversity in R&D. In addition, the social implementation of intellectual property suggests that it was promoted through joint applications with a startup firm that has a complementary role to universities. On the contrary, the results of the qualitative analysis suggest that the creation of intellectual property was promoted by building a knowledge logistics system that enables the acquisition of tacit knowledge (reciprocal mindset)

through experiential learning by collaborating with a startup firm. This would help cultivate a dual-skilled researcher with a reciprocal mindset who can bridge the gap between academia and industry.

In Chapter 3, the co-author and co-word networks are constructed based on the bibliographic information of papers generated until phase 2 in the COINS as a network that can approximate the collaboration and knowledge structure in a large-scale interdisciplinary R&D project. Graph entropy, which is a measure of structural complexity of the network, is introduced as a method to realize the instrumentation of static and dynamic features of the collaboration or knowledge structure. Qualitative observations of these network structures and analyses on the changes in the structural complexity of these networks, are conducted in terms of the changes in complexity due to the decomposition of the structure into project, team, and individual levels as well as the changes in complexity due to the time evolution of the network over time. The results of qualitative observations on the structure of the co-author and co-word networks confirm that domain-specific characteristics at the subtheme level of COINS are reflected in these networks. The cross-sectional analysis demonstrates correlations between the changes in complexity are induced by structural decomposition. The results suggest that such structural complexity in interdisciplinary research projects can be managed by organizing appropriate sub research groups. Furthermore, this correlation suggests that knowledge structures can be induced through collaboration structures, as shown by the qualitative observation between co-author and co-word networks. The longitudinal analysis confirms that the change in complexity decreased over time for both collaboration and knowledge structures. This fact may reflect the project's strategy of pursuing high interdisciplinarity (or innovative research outcomes) in the early stages and the integration of the outcomes obtained to realize the vision in the later stages. These insights into the management of interdisciplinary research projects derived from these measures of structural complexity are validated by an interview with research leader.

In Chapter 4, based on the insights obtained from the previous chapters, investigations are made into how the research strategy was implemented to achieve the project's vision. In the context of a large-scale interdisciplinary R&D project, the generative mechanisms for interventions of research collaboration to create knowledge that contributes to the achievement of the COINS project's vision are discussed. This is done by using the information accumulated through the participant observation of the target cases and arranging the context, interventions, mechanisms, and outcomes for the COINS as a whole and for each subtheme leader. After summarizing the research strategies of each subtheme leader, the possibility that they can be categorized from two perspectives is discovered: production and consumption of knowledge and whether it is core knowledge or peripheral knowledge. This classification suggested that COINS subtheme leaders were adopting different research strategies while conducting R&D to achieve the common vision of an "in-body hospital."

5.2 Implications

5.2.1 Theoretical implications

The theoretical implication of this study is that introducing the concept of instrumentation of organizational and managerial phenomena in scientific research activities is an important step toward systematizing management of science. Extending the scientometric methodology to strategic and organizational management of scientific research can provide quantitative basis for the interpretation of the organizational and managerial phenomena observed in scientific research, thus contributing to the systematization of the theory of management of science.

5.2.2 Practical implications

The practical implication of this study is that it provides an instrument for evidence-based management to practitioners of management of scientific research. The methodology for instrumentation that enable the data acquisition which reflects the organizational and managerial phenomena observed in scientific research can contribute to data-driven decision-making by project managers and research administrators.

5.3 Limitations and future perspectives

The major limitation of this study is that generalizability cannot be guaranteed because it is essentially an analysis of a single case study. This study mainly focuses on two projects, NanoBio First and COINS; however, COINS is essentially a case study based on a single case because it is a successor to NanoBio First. To increase generalizability, comparative analysis with other cases is required. However, as mentioned in the previous chapter, due to problems such as difficulties in case collection, context-specificity of interpretation, heterogeneity of outcomes among different fields, and compatibility of the generative mechanism for interventions, further theoretical development to increase generalizability requires normative condition setting such as what are desirable observation that allow comparison under the same conditions.

In terms of instrumentation of organizational and managerial phenomena, another aspect of limitation is what theory should be used to measure the object that will be wanted to measure, namely the problem of the theory-ladenness of observations. For example, the graph entropy introduced in this thesis depends on the system size, the number of nodes and edges. Since the words or phrases are extracted from the text as nodes of the co-word network, the value of the graph entropy of the co-word network also necessarily depends on some factors, such as the pre-processing of the text, the method of key phrase extraction, and how to give the weights of the edges in the network. Therefore, the theory of determining what kind of measurement is desirable for the appropriate implementation of instrumentation for organizational and managerial phenomena is also an important topic for future developmental studies.

Acknowledgement

I am grateful to Prof Dr Shintaro Sengoku, my thesis supervisor, who gave me essential guidance and helpful feedback. I am equally grateful to my co-supervisor, Prof Dr Yuya Kajikawa, for helpful discussion. I appreciate the feedback offered by Dr Kazutoshi Sasahara, Prof Dr Masahiro Hashimoto, Prof Dr Nobuhiro Nishiyama, and Prof Dr Shuzo Fujimura. Thanks must also go to Prof Dr Kazunori Kataoka, Prof Dr Hiromichi Kimura, Dr Tomohiro Anzai and other COINS members who gave me essential information and suggestion, and a scholarly literature database Web of Science provided by Clarivate Analytics for data provision. I would like to acknowledge the contribution by the members of Sengoku laboratory. Discussions with members were stimulating, and helpful. Thank you also to my family for your endless support.

Appendix A

There are some processes to convert raw bibliographic data into suitable form for analyses after downloading from the Web of Science. Here, the detailed procedures to construct the following datasets for analyses are described:

- An aggregated dataset of bibliographic information
- A text corpus of titles and abstracts by publication
- Panel data of the summarized lists of institutions and publications by author
- Panel data of the adjacency matrices of co-author networks
- Panel data of the adjacency matrices of co-word networks

In the process of constructing these datasets, the present study mainly used pandas, which is a library to conduct data manipulation and analysis for Python. In addition, the present study describes the generation of network objects from the created adjacency matrix and the parameters used for drawing the network diagram by Gephi.

A.1 Preprocessing of collected data: Converting BibTex files into JSON format

To analyze the complexity of collaboration and knowledge structure, the present study used BibTex files downloaded from the Web of Science database. Since raw BibTex files are not suitable for dealing with Python, these files were converted into JSON format. The bibliographic information was extracted from raw BibTex files using the BibtexParser, which is a library to parse such files for Python. Then, files that could not be opened using BibtexParser were excluded from the dataset. Finally, bibliographic data extracted from 281 publications were aggregated to a dataset including information by publication as follows: document type, list of authors, published year, title, text of the abstract, and so on.

A.2 How to build a text corpus

Building a co-word network requires a corpus of key phrases extracted from the titles and abstracts in publications. In the present study, the corpus contained lists of key phrases per sentence extracted from the title and abstract was built using Natural Language Tool Kit (NLTK), which is a library for natural language processing in Python. The procedure to build the text corpus is as follows:

1. Text data of the title and abstract per publications were extracted from the bibliographic dataset.
2. To apply the key phrase extraction smoothly, text data were divided by sentences and pre-processed.
3. Sentences were tokenized and symbolic string was removed from some obtained tokens.
4. Part-of-speech (POS) tagging was performed on the obtained tokens in tokenized sentences.
5. Key phrases extraction using rule-based methods were applied to the tokenized sentences and the words that consisted of the extracted key phrases were lemmatized.
6. Common words were removed from the extracted key phrases as stop words.

To ensure the reproducibility of the analysis, that is, to eliminate the subjective judgments as much as possible, regular expressions were employed in the above processes. Applying the above processes to the dataset of bibliographic information resulted in a corpus containing the key phrases.

First, the text data of the title and abstract were divided by sentence, using "sent_tokenize" function of NLTK, because of the co-occurrence of the words and phrases that are defined by the sentence. To eliminate the influence of words and phrases that were not relevant to the content of the research topics, the following descriptions were removed from each sentence after dividing by sentence:

- Descriptions about copyright or publisher, such as "(C) 2015 Elsevier," "Published by Elsevier," and so on.
- Descriptions about statistical analysis, such as "($p<0.05$)" and "($N = 30$)."
- Abbreviations, such as "(PEG)" in "poly(ethylene glycol) (PEG)."

Next, the sentences containing unnecessary descriptions were tokenized, using the "word_tokenize" function of NLTK as the tokenizer. Some tokens contained symbolic strings such as "/" or "-," which would create noise in the analysis. Thus, the following processing was applied to suppress noise in analysis:

- The tokens joined by symbolic string "/" were split.
- The tokens joined by symbolic string "-" were split.
- Symbolic string "-" at the head of the tokens were removed.
- Symbolic string "/" at the tail of the tokens were removed.

After tokenization, tokens (or words) were executed via POS tagging using Stanford POS Tagger (Toutanova and Manning 2000), which can be performed using the NLTK wrapper. A pilot study about Stanford POS Tagger performed almost correctly, however, several problems were also confirmed: some symbolic strings were recognized as nouns, and some nouns were recognized as foreign words. To avoid these problems, the following process was applied after POS tagging.

- POS tags of the symbolic string such as ":" and "%" were set into another POS tag (SYMBOL).
- The strings of the brackets such as "(", ")" were removed.
- POS tags for words recognized as foreign words were converted into nouns.

A rule-based key phrase extraction method was adopted and was accomplished by chunking using the pattern of occurrence of POS in a sentence based on regular expression, provided by the "RegexpParser" function of NLTK. The rules for key phrase extraction were as follows:

1. Gerund + noun + noun +...
2. Past participle + adjective + adjective +...+ noun +noun +...
3. Adjective + gerund + noun
4. Noun + gerund + noun
5. Noun + past participle + noun
6. Adjective + adjective +... + Noun + Noun +...

7. Noun + noun +...
8. Nouns

Although the above rules imitated the previous study (Wang et al. 2014; Wang and Wang 2018), they were not strictly the same. Words composed of the extracted key phrases according to these rules were lemmatized using the WordNetLemmatizer, a lemmatizer available from NLTK. When a word was lemmatized, it was converted into lowercase to eliminate case sensitivity.

Finally, common words, such as "study" and "cell," were set to stop words and removed from the extracted key phrases. Words or phrases were set to stop words if they were in WordNet, a dictionary that can be easily called from NLTK, and removed from the corpus of key phrases.

A.3 How to build a panel data of the information about authors

Building a co-author network requires information about the node attributes (i.e., information about the authors themselves) as well as the edges (i.e., connections between authors). Panel data containing the following information for each author were built from the aggregated dataset of bibliographic information:

- Author's name
- Count of author's affiliation
- The main affiliation of an author
- A list of publications published the author
- Number of publications published the author
- Author's ORCID
- Author's Researcher-ID

An affiliation with the highest count of affiliations was selected as the main affiliation.

After building the panel data of the author information, authors' names were collated and author information was updated because some names were duplicated. In the case of an author having an identifier such as ORCID, the name was collated based on the identifier; otherwise, the names were collated manually.

A.4 How to build a panel data of adjacency matrices of co-author networks

The procedure to build the panel data for the adjacency matrices of the co-author network was as follows:

1. All authors, which corresponded to the nodes in the co-author network, were listed from the dataset of aggregated bibliographic information.
2. Collections of bibliographic data of publications by year were extracted to build the panel data of adjacency matrices from the dataset.
3. A list of authors per publication in that collection was obtained and the edge weights for all combinations in the list were computed.

4. The edge weights in the co-author network as a whole were calculated as the sums of each edge weight per publication for all publications.
5. The edge weights in the co-author network were normalized by dividing by the number of publications so that the sum of the edge weights is 1.

The unit of analysis for the co-author network was set to one publication because co-authorship is defined per publication. The edge weights in the co-author network were calculated as the sum of the inverse of the number of edges in the unit of analysis (each unit of analysis forms a complete graph and if such a graph has N vertices, then the number of edges is $N(N-1)/2$).

A.5 How to build the panel data of adjacency matrices of co-word networks

The procedure to build the panel data for the adjacency matrices of the co-word network was as follows:

1. All key phrases, which corresponded to the nodes of the co-word network, were listed from the text corpus.
2. Collections of corpora consisting of key phrases per publications by year were extracted for building the panel data of adjacency matrices from the corpus.
3. A list of key phrases per sentence in that collection was obtained and the edge weights for all combinations in the list were computed.
4. The edge weights in the co-word network as a whole were calculated as the sums of each edge weight per sentence for all sentences.
5. The edge weights in the co-word network were normalized by dividing by the number of sentences so that the sum of the edge weights is 1.

The unit of analysis for the co-word network was set to one sentence because the co-occurrence of the words or phrases is defined per sentence. Similar to the co-author network, the edge weights in the co-word network were calculated as the sum of the inverse of the number of edges in the unit of analysis.

A.6 Generation of network objects

In present study, to deal with the co-author and co-word networks, the present study uses NetworkX, which is a library for network analysis in Python. Adjacency matrices of the co-author and co-word networks were converted into network objects, using the “from_pandas_adjacency” function of NetworkX. Network objects have the advantage that it allows to set attributes of the nodes, such as information about the sector of the author’s affiliation in a co-author network.

In the cross-sectional analysis, the coauthor and co-word meso-network and micro-network are extracted from these macro-networks by using the “subgraph” function of NetworkX. This function generates a network object of the sub-network by passing a list of nodes, which is a subset of the nodes in an original network, as an argument. The subset of nodes (i.e., authors and words) used to extract the co-author/word meso/networks is defined by listing the nodes that represent authors or words in

the title and abstract that appear in publications published by a group of individuals belonging to a subtheme or by an individual, respectively.

To build a list of meso-level nodes, an article was considered to belong to a subtheme when we could confirm that an author belongs to that subtheme from the internal documents of the COINS[17]. The list of authors and words in an article belonging to asubtheme becomes a subset of nodes for extracting the co-author/word meso/macro-network. For example, if an author belonging to subtheme 1 and an author belonging tosubtheme 2 publish a co-authored paper, this paper was considered to belong to both subtheme 1 and 2 thus these authors appear in the co-author meso-network for both subthemes, and the words in the title and abstract of this paper also appear in the co-word network of both subthemes.

Similarly, in the case of the micro level, a list of authors and words in publications published by an author becomes a subset of nodes for extracting the co-author/word micro-network from the macro-network. For example, co-authors in a series of publications published by an individual appear in the co-author micro-network of the individual and words in the title and abstract of them appear in the co-word micro-network of the individual.

A.7 Parameters for drawing network diagram by Gephi

The parameters of the Fruchterman-Reingold algorithm for drawing the co-author network and co-word network are shown in the table A.1. The parameter of "Speed" was change manually.

Table A.1 Parameters of Fruchterman-Reingold algorithm

Parameter	Co-author network	Co-word network
Area	200000	100000
Gravity	0.25	0.5
Speed (initial)	100	100
Speed (after spread)	10	10

[17] The present study assigns a subtheme to each publication to determine the meso-level attributes of all authors because this document does not cover all authors. In addition, some members varied their subthemes belonged to depending on the year. Then, these members were considered to be included in that subtheme if they belonged to a certain subtheme in any year.

Appendix B

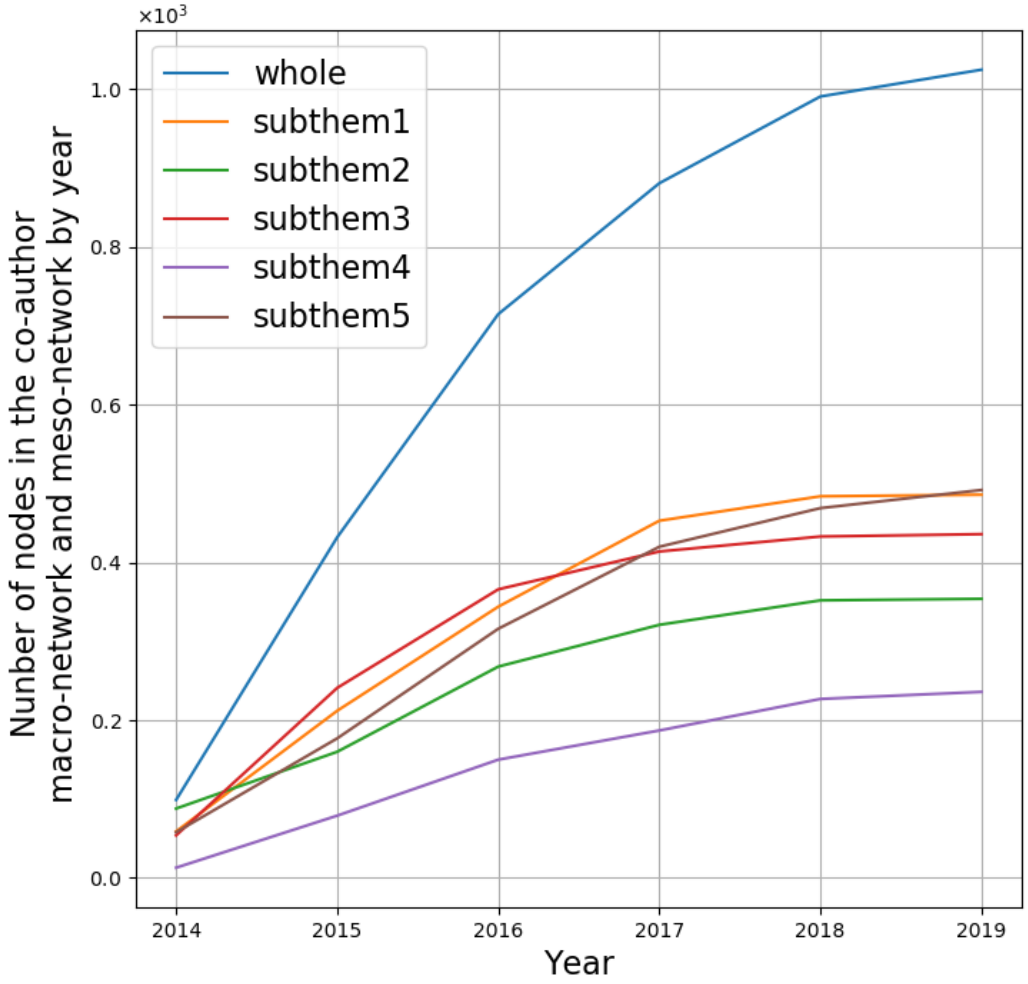

Figure B.1 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the number of nodes over time for the co-author macro-network and meso-network.

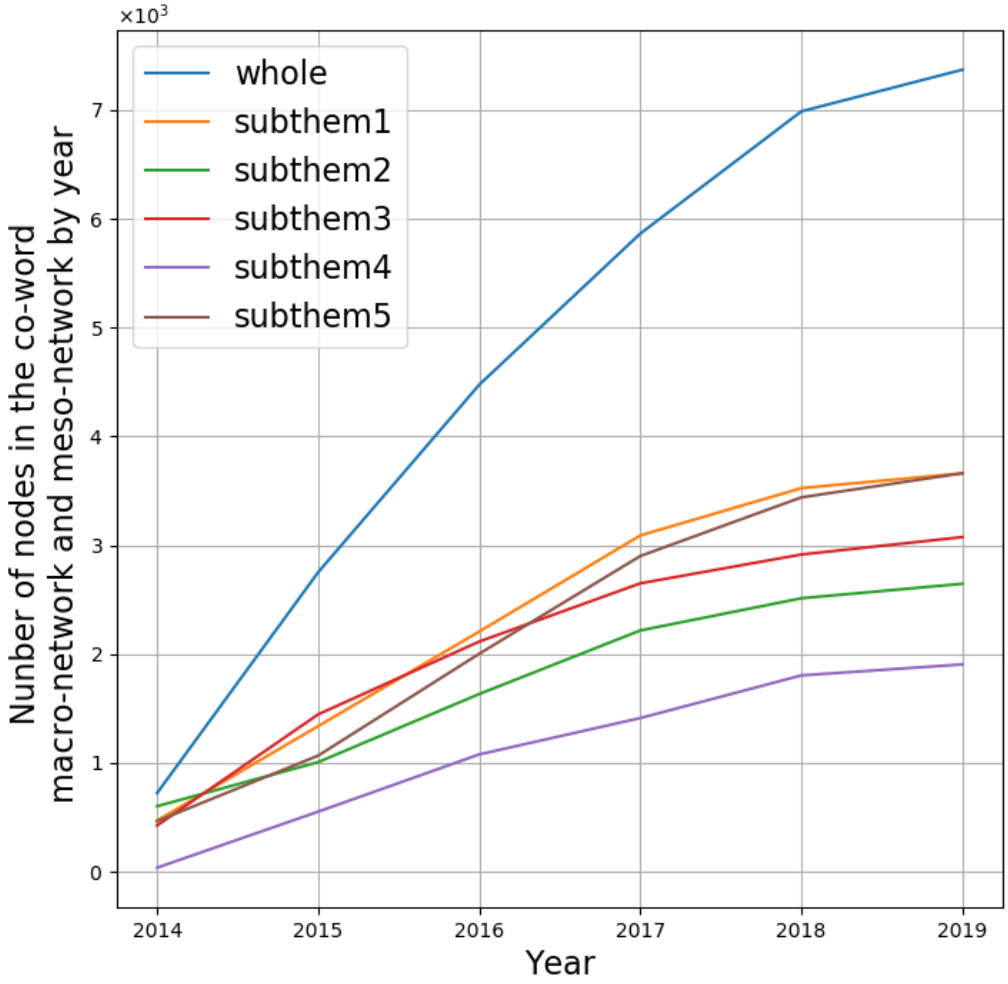

Figure B.2 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the number of nodes over time for the co-word macro-network and meso-network.

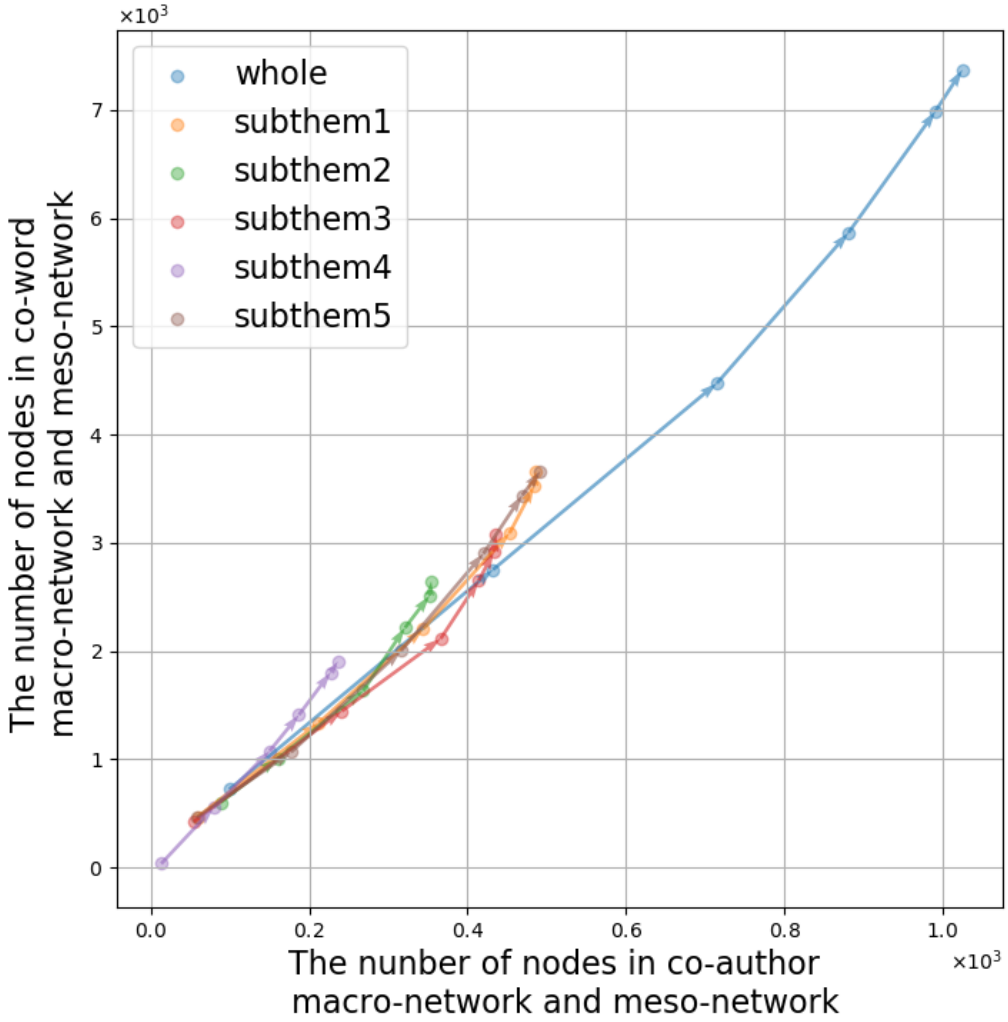

Figure B.3 Trajectory of the number of nodes in co-author and co-word networks. The horizontal axis represents the number of nodes for the co-author macro-network and meso-network. The vertical axis represents the number of nodes for the co-word macro-network and meso-network.

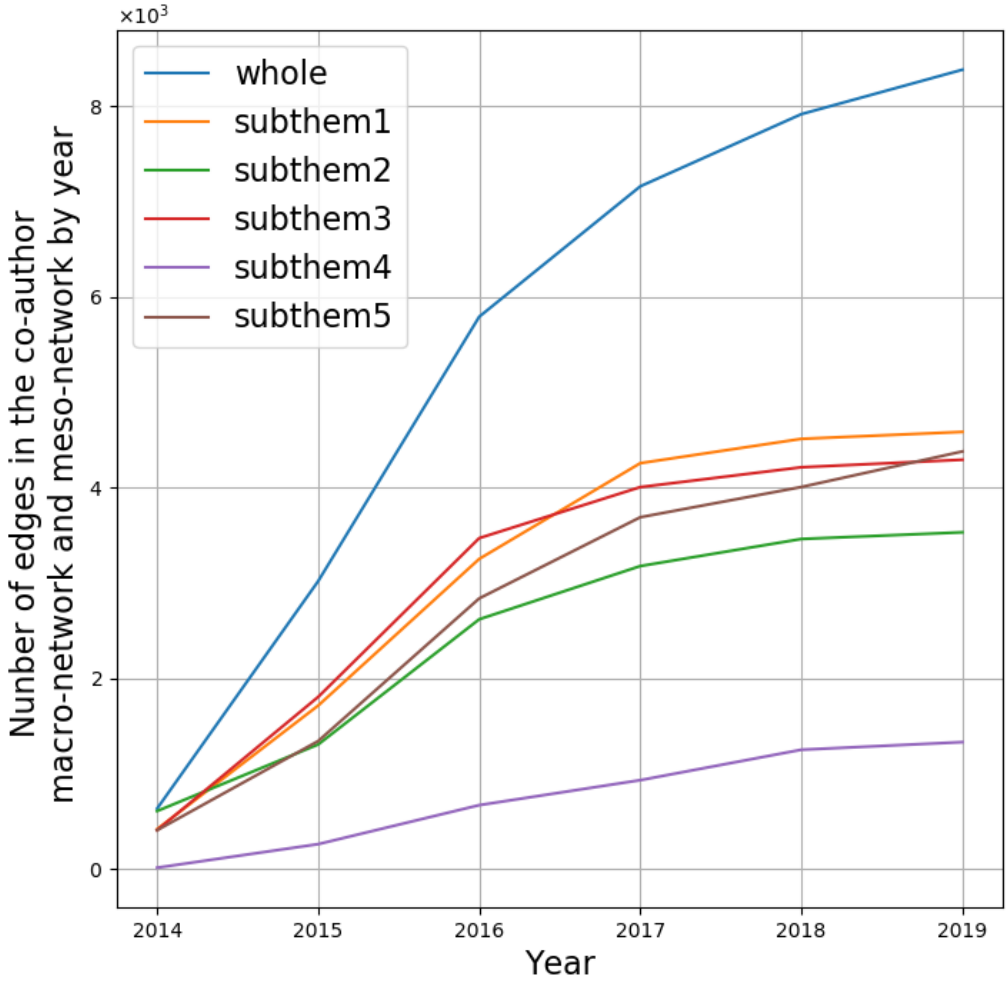

Figure B.4 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the number of edges over time for the co-author macro-network and meso-network.

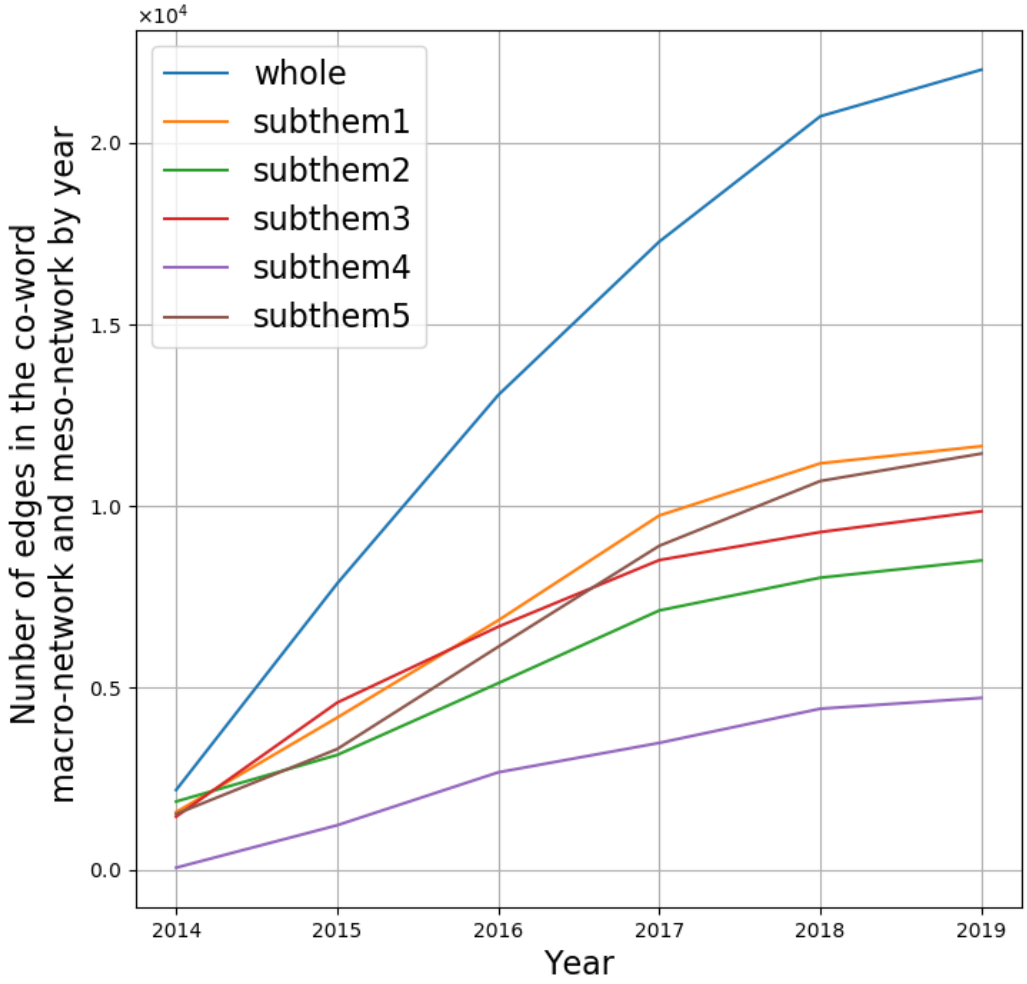

Figure B.5 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the number of edges over time for the co-word macro-network and meso-network.

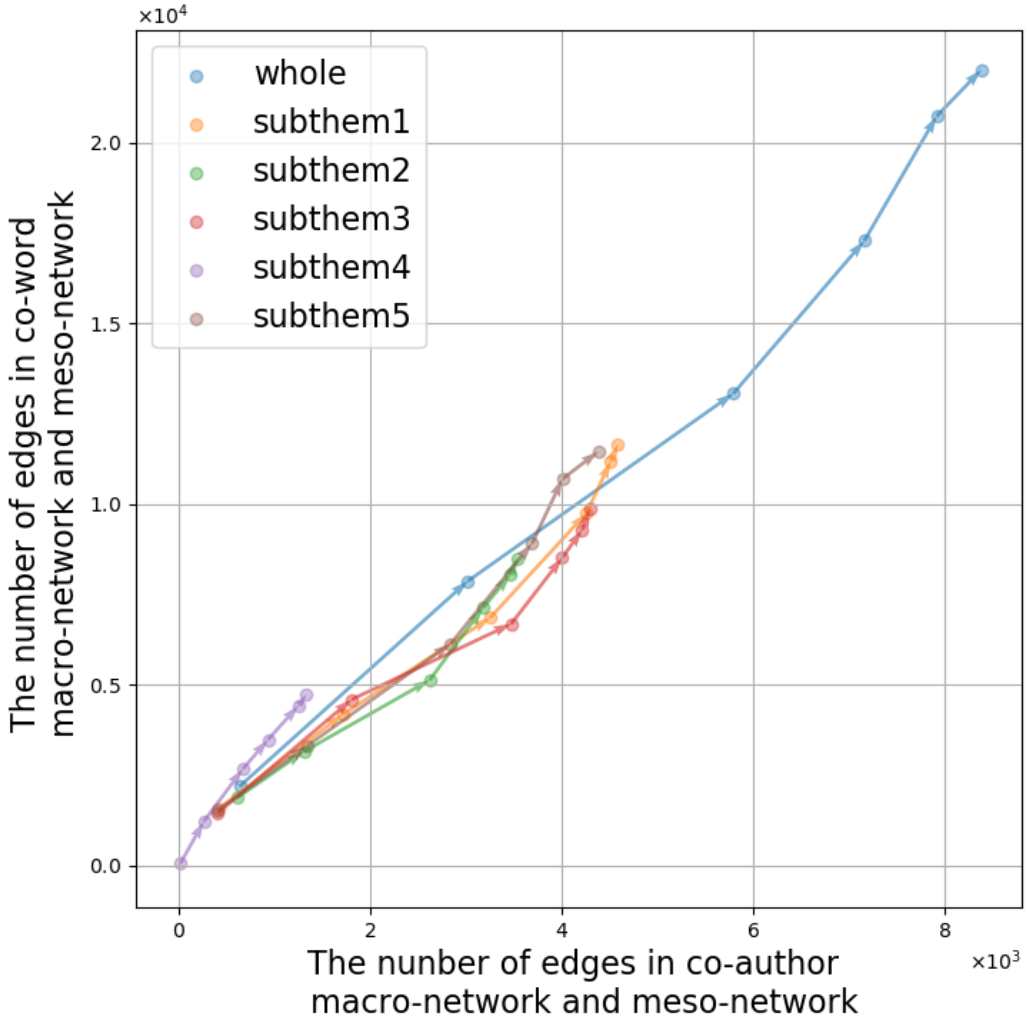

Figure B.6 Trajectory of the number of edges in co-author and co-word networks. The horizontal axis represents the number of edges for the co-author macro-network and meso-network. The vertical axis represents the number of edges for the co-word macro-network and meso-network.

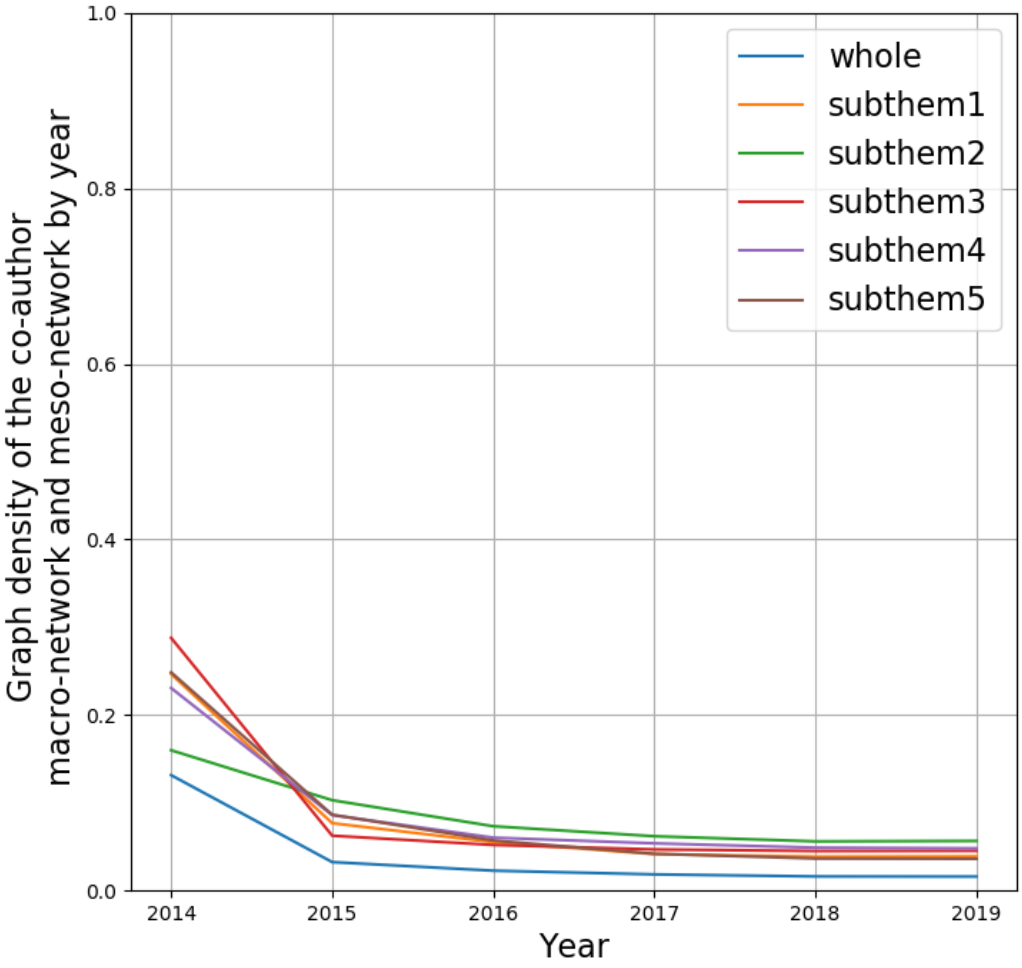

Figure B.7 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the density over time for the co-author macro-network and meso-network.

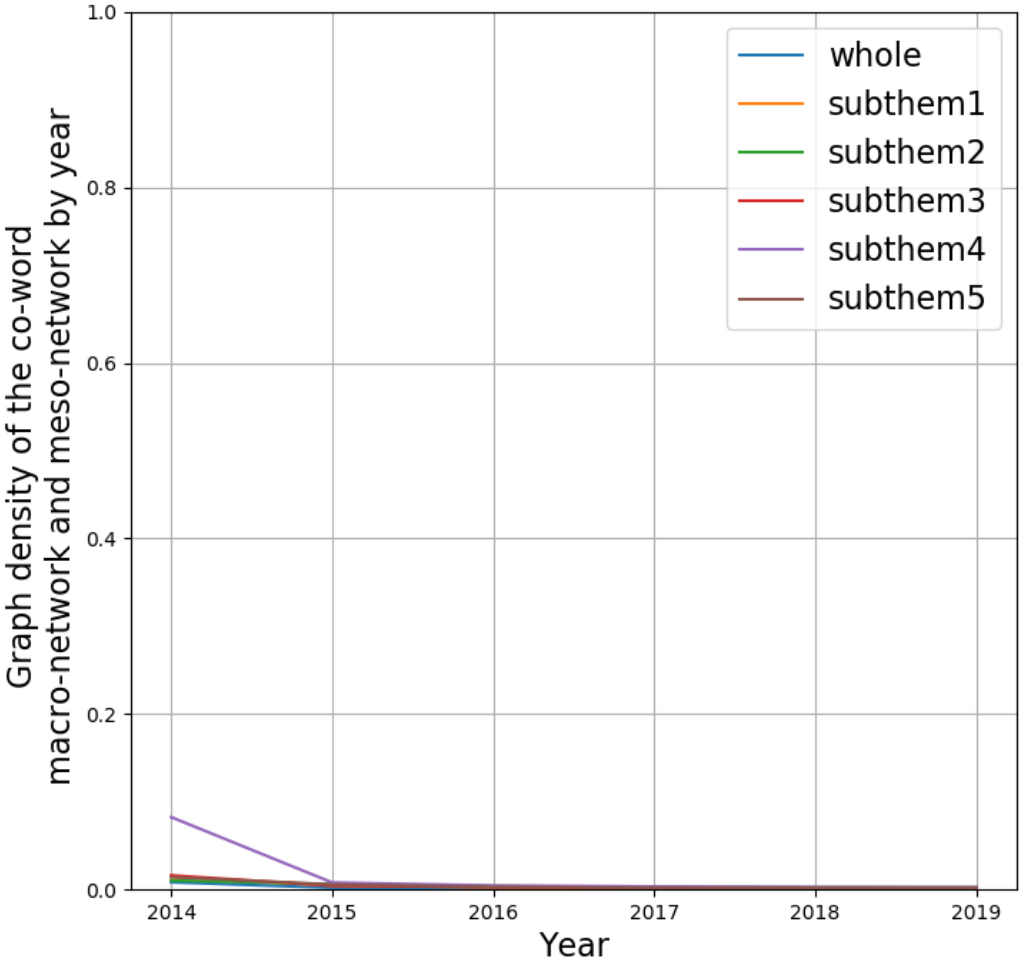

Figure B.8 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the density over time for the co-word macro-network and meso-network.

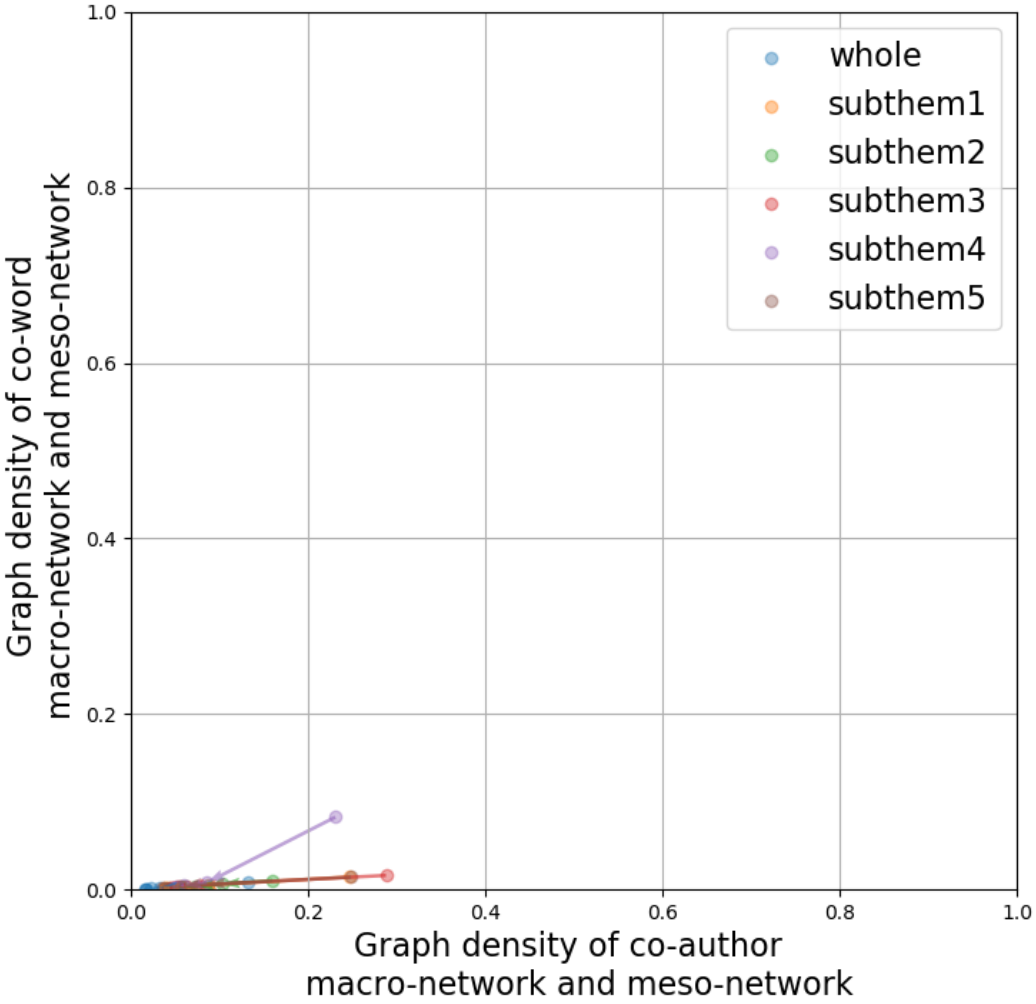

Figure B.9 Trajectory of the density in co-author and co-word networks. The horizontal axis represents the density for the co-author macro-network and meso-network. The vertical axis represents the average clustering coefficient for the co-word macro-network and meso-network.

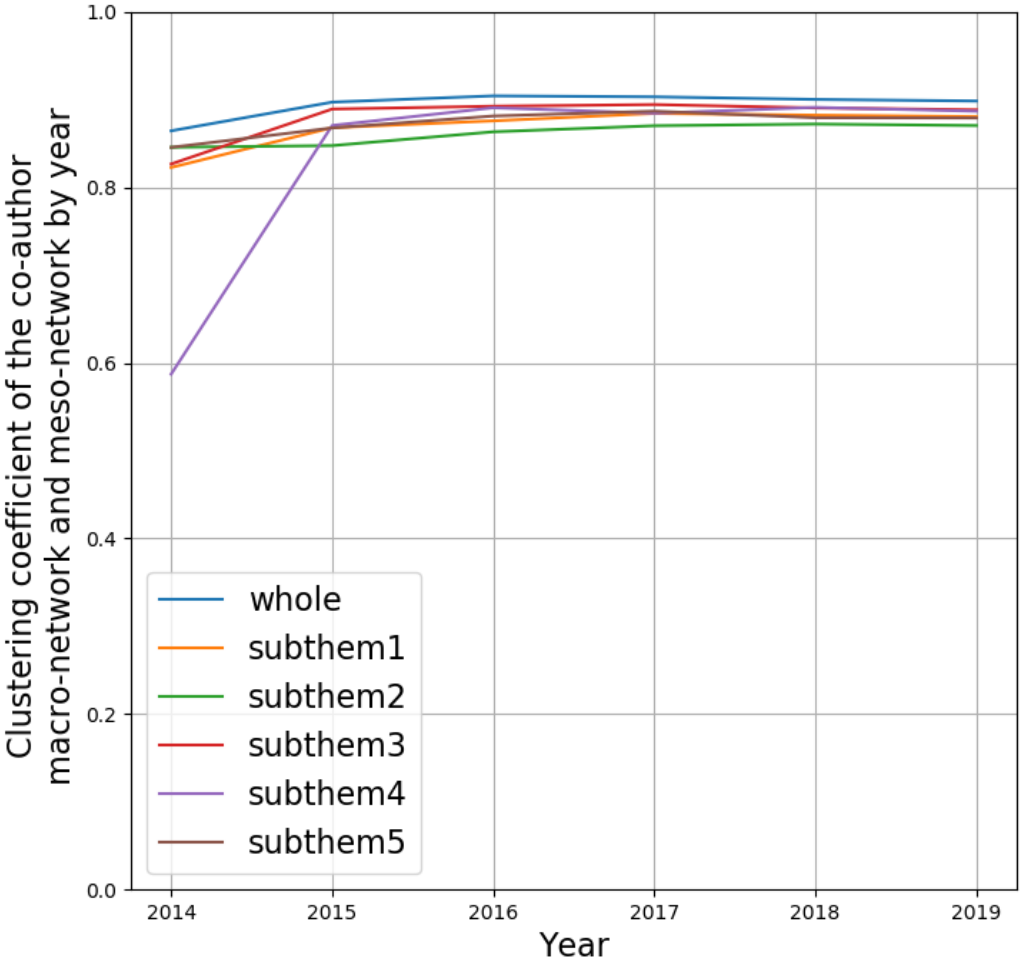

Figure B.10 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the average clustering coefficient over time for the co-author macro-network and meso-network.

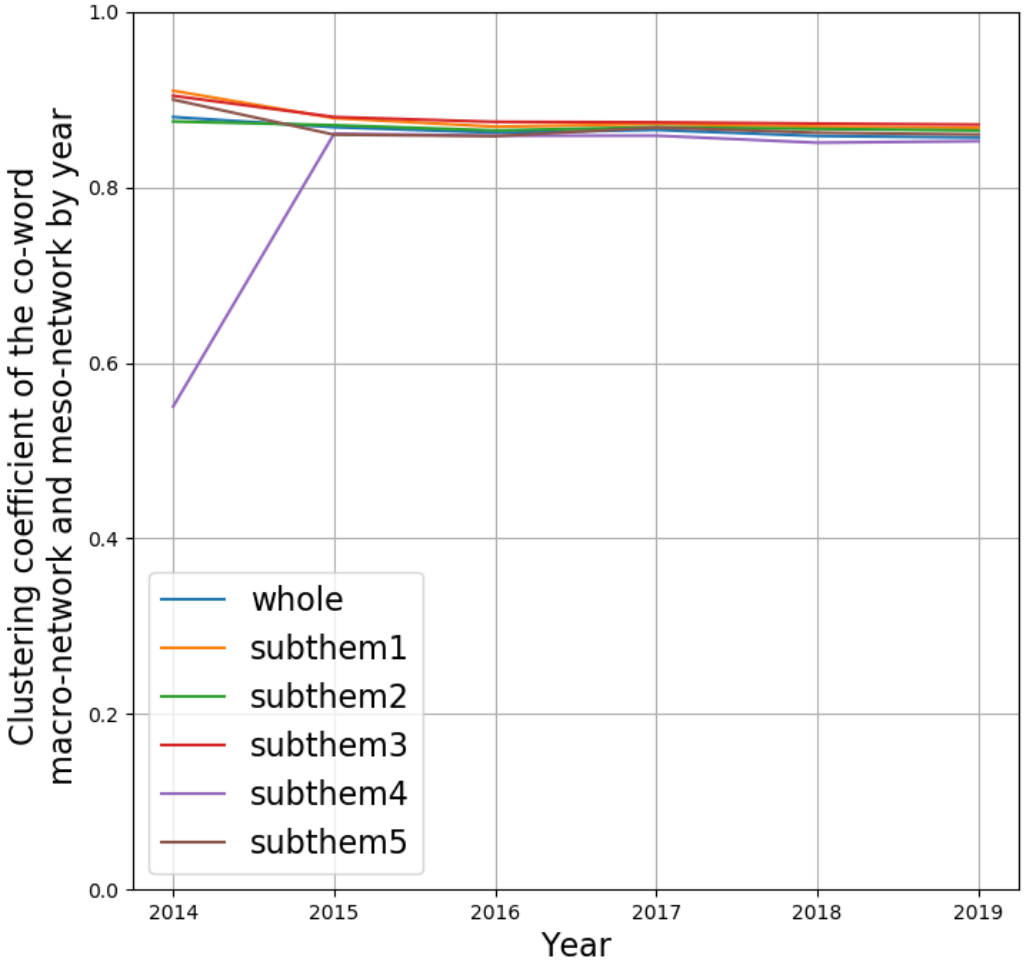

Figure B.11 Time-course plot of the whole project and 5 subthemes. The horizontal axis represents the period of observation in the present study. The vertical axis represents the average clustering coefficient over time for the co-word macro-network and meso-network.

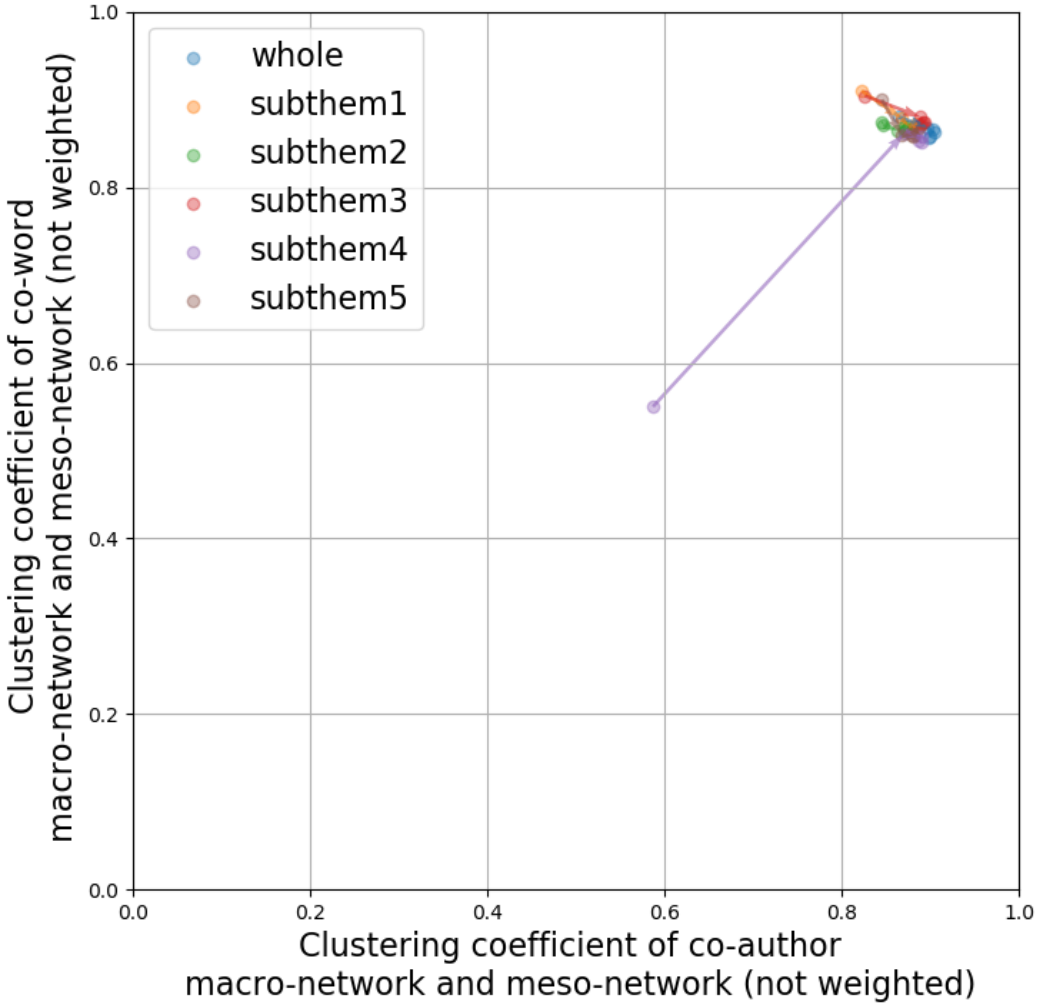

Figure B.12 Trajectory of the average clustering coefficient in co-author and co-word networks. The horizontal axis represents the average clustering coefficient for the co-author macro-network and meso-network. The vertical axis represents the average clustering coefficient for the co-word macro-network and meso-network.