

RESEARCH

Open Access



Multilayer network analysis of the drugs development cycle in the global pharmaceutical industry

Hiromitsu Goto^{1†}, Mari Jibu^{2†}, Wataru Souma^{1†} and Yuichi Ikeda^{3*†} 

*Correspondence:

ikedayuichi.2w@kyoto-u.ac.jp

¹Hiromitsu Goto, Mari Jibu, Wataru Souma and Yuichi Ikeda have contributed equally to this work

³Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Kyoto 6068306, Japan
Full list of author information is available at the end of the article

Abstract

Drug development is a time-consuming process from the start of research to obtaining approval, and the probability of success with a candidate compound is extremely low. We aim to understand the characteristics of the flow and localization of knowledge during drug development in the global pharmaceutical industry. We analyze the multilayer network constructed with the drug pipeline layer, global supply chain layer, and global ownership layer. First, we identify the bow-tie structure and the community structure of each network layer. The obtained bow-tie structure shows the large strongly connected component and suggests that the knowledge flow in drug pipelines has similar characteristics as the supply chain network. The communities in each layer are characterized by country, category of the company, and bow tie component. We then study the multilayer network's knowledge flow, conduct a statistical test, and verify the significance of the overlapping links between the drug pipeline and supply chain layers. Our results suggest a strong connection between open innovation in the pharmaceutical industry and firms' economic activities in the supply chain.

Keywords: Multilayer network, Pharmaceutical industry, Drug pipeline, Supply chain, Ownership

Introduction

The research and development (R&D) of a pharmaceutical product is time-consuming from the start of research to obtaining approval, and the probability of success with a candidate compound is extremely low. The probability of success from discovery and pre-clinical initiation of a candidate compound to launch is approximately 0.1 %, the cost of development to launch one product is in the hundreds of millions of dollars, and R&D requires more than ten years. Recently, the pharmaceutical industry's research costs have increased at a rate higher than the average for all industries.

In the pharmaceutical industry, leading companies undergo shifts in the processes of R&D, production, and sales (Comanor and Scherer 2013; Scherer 2010; Munos 2009). The R&D process is called the drug pipeline, starting with new drug discovery. The development of new drugs requires pre-clinical testing, three stages of clinical trials, and approval in each country.

The knowledge flow in the drug pipeline is defined as the shift of license to deal with drug candidates. Osabe and Jibu (2013a, b, 2014a, b, c, d, e, f) pointed out that the knowledge flow is critical in academic collaboration for the development of new drugs and is measured as the shift of leading companies observed in the drug pipeline data. However, the knowledge flow dynamics between the pharmaceutical companies in the drug pipeline remain a concern (Narayana et al. 2014; Mazzola et al. 2015).

The shift of license, i.e., the knowledge flow, in drug development, is a complex phenomenon arising from the relationships between pharmaceutical companies, government institutions, and educational institutions. The multilayer networks (MLNs) are significant concepts for understanding complex systems because many systems in the real world are constructed from various relations between components. For example, firms are connected by supply chain, ownership, a concurrent post of board members, and co-application of patents Aoyama et al. (2010). Further examples of MLNs include transportation networks (Kurant and Thiran 2006; Zou et al. 2010), climatic systems Donges et al. (2011), economic markets Yang et al. (2009), and energy-supply networks Buldyrev et al. (2010). The statistical mechanics in single networks are expanded into MLNs Bianconi (2013).

Therefore, our primary research goal is to understand the characteristics of the flow and localization of knowledge during drug development. Here, localization refers to knowledge accumulation within the same company, observed as incoming links and self-loops in the drug pipeline network. We analyze the drug pipeline data based on the MLN constructed with the drug pipeline network layer, global supply-chain network layer, and the global ownership network layer to achieve this research goal.

This paper is organized as follows. Section 2 describes the datasets used in this paper. We specifically explain the detailed characteristics of the drug pipeline data and drug pipelines. Section 3 explains and investigates the methods of how to construct the MLN. The analysis and results are explained in Sect. 4, showing a statistical test to verify the significance of the overlapping nodes and edges between a pair of layers. Finally, Sect. 5 is devoted to conclusions and discussion. Appendix 1 summarizes the analysis of a single-layer network, and Table 1 depicts a list of abbreviations used in this paper. We use ISO codes for countries in most of the figures and tables in this paper.

Data

Drug pipeline data

We constructed a drug database, extracting all drug data from the Cortellis Competitive Intelligence database provided by Clarivate Analytics Cortellis, on December 11, 2013. The Cortellis Competitive Intelligence database includes pipeline data from various drug candidates from the R&D stage to the clinical trial stages and deals reports from approximately 7000 pharmaceutical companies. The drug development cycle consists of discovery, launched, clinical trial, pre-registration, and registered. Furthermore, this clinical trial can be categorized into three phases: Phase I (examinations for drug safety conducted with healthy adults), Phase II (examinations for safety and effectiveness of a drug compared with existing drugs, conducted with a few patients), and Phase III (examinations for safety and effectiveness of a drug compared with existing drugs conducted with many patients). Deals reports include originator

Table 1 List of abbreviations

Abbreviation	Definition
R&D	Research and development
MLN	Multilayer network
SCC	Strongly connected component
GWCC	Giant weakly connected component
GSCC	Giant strongly connected component
TE	Tendril
M&A	Mergers and acquisition
AI	Artificial intelligence

companies (i.e., the licensor) and active companies (i.e., the licensee). We indexed these companies' country and type. Indeed, there exist various types of business entities dealing with drug pipelines. The types are indexed as small- and medium-sized business firms classified as a private company, large-sized businesses classified as a public company, universities classified as educational institutions and government institutions classified as government institutions. The extracted drugs amount to 38,295, excluding those being marked as "no development reported" in the highest status, 18- months behind the extracting date. Notably, the drug pipeline data are a snapshot and do not contain historical information regarding the drug development cycle. For example, Phase III data, whose licensee is different from the licensor, can show the shift of license of the drug candidate between them until Phase III but cannot show when it happened.

The drug pipeline starts with discovering a new drug candidate, which requires pre-clinical testing, three stages of clinical trials, and approval to launch. Generally, venture firms or educational institutions provide grants to the discovery of drug seeds and the licensee. For example, leading firms advance the development to launch, incurring considerable R&D expenses. These characteristics vary from country to country. Drug pipeline data help us to understand the knowledge flow in the drug development cycle as the shift of license to deal with drug candidates, i.e., the flow from the licensor to the licensee.

Figure 1 shows the number of drug pipelines at the launched status by country, where we counted the license of the launched drugs by country. Furthermore, we compared the licensee with the licensor in drug pipelines to measure the self-sufficiency of individual countries and companies. For example, in the United States (US), own country in the legend means that drug pipelines are started in the US, other country means all countries except the US, own company means that the US company as the licensee started the drug pipeline by own company, and "other company" indicates all companies except for the licensee. The US has launched several drug pipelines. We found that most of the top 20 countries have discovered and launched more than half of their drug pipelines in their countries, and the drug pipelines in China (CHN) tend to be developed in her country compared with other countries. We anticipate that the licensee of the drug pipeline frequently changes because of R&D expense. To characterize the changes from the viewpoint of the type of business entity, we compared the licensor or licensee type, namely, government institution, educational institution, a private company, and public company.

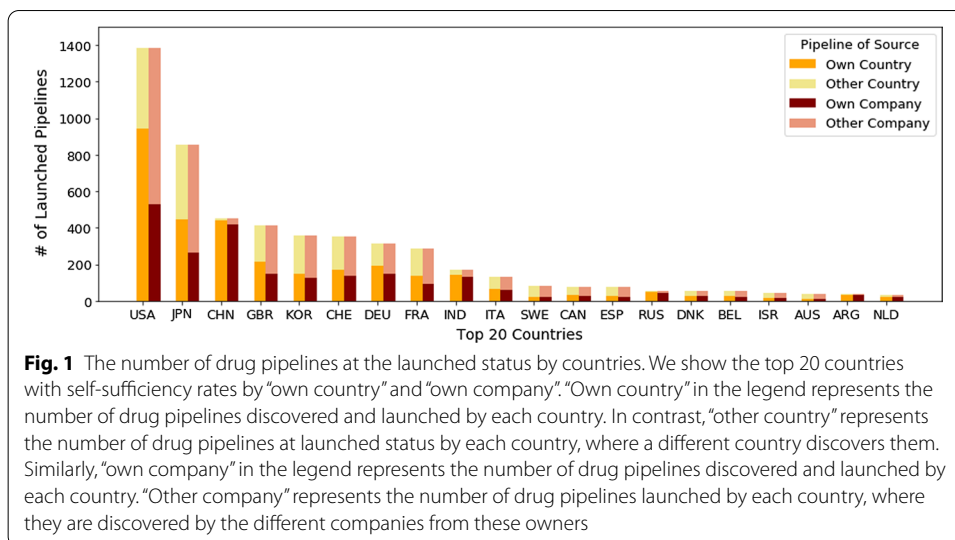
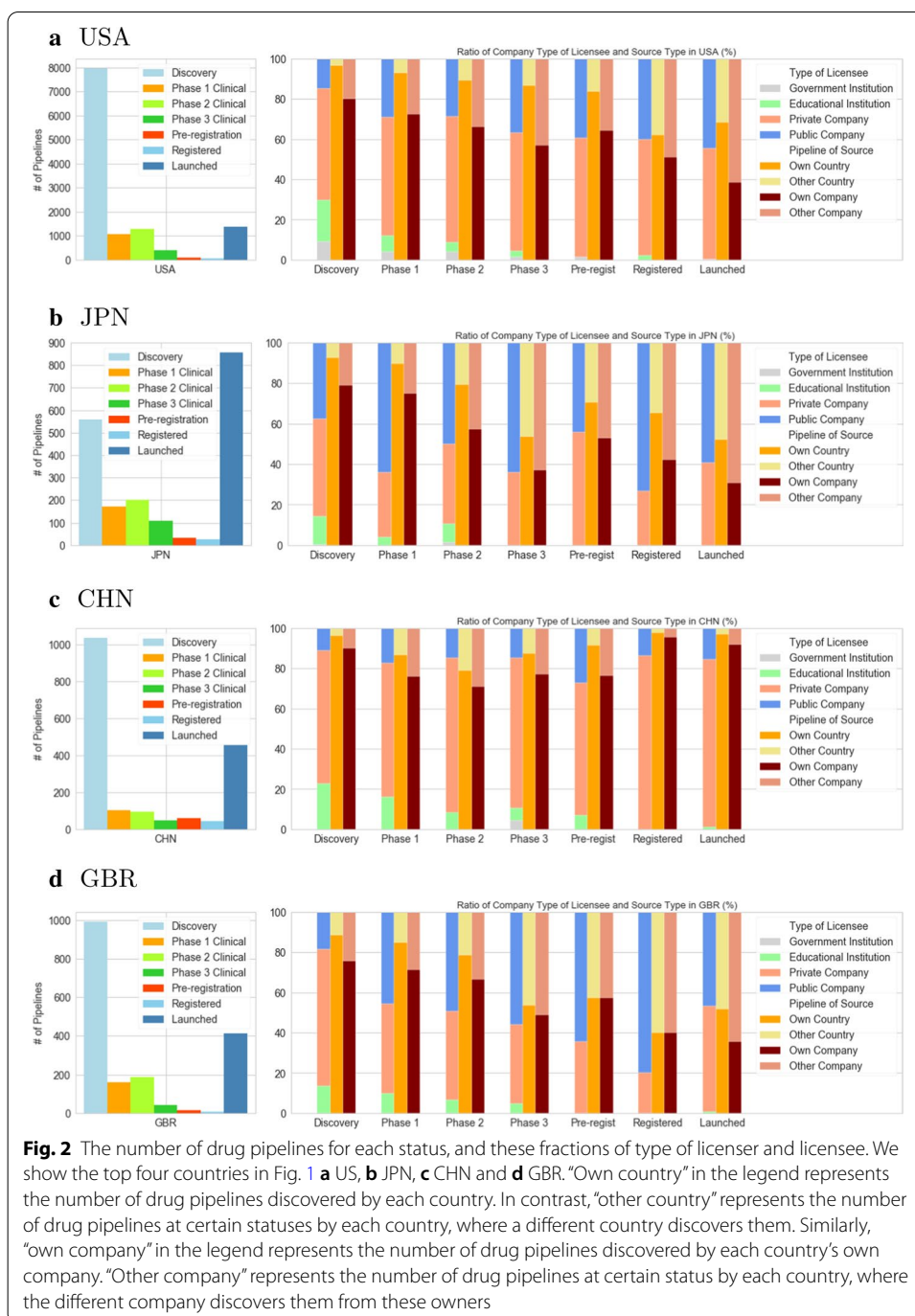


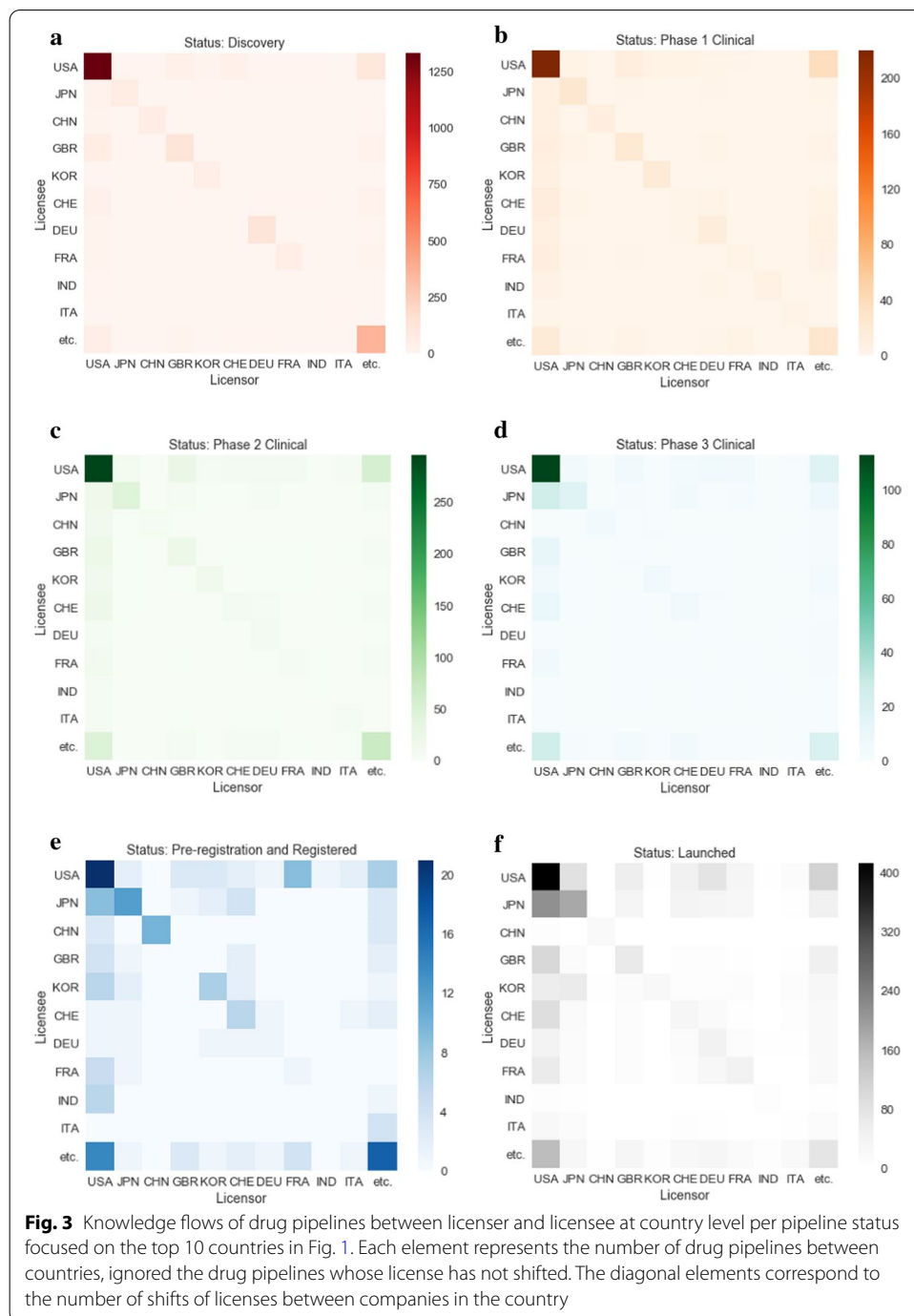
Figure 2 shows the number of drug pipelines for each status for the top four countries in Fig. 1 and the ratio of components regarding business entity and self-sufficiency. "Own country" in the legend represents the number of drug pipelines discovered by each country. In contrast, "other country" represents the number of drug pipelines at individual status by each country, where a different country discovers them. Similarly, "own company" in the legend represents the number of drug pipelines discovered by each country's own company. "Other company" represents the number of drug pipelines at individual status by each country, where the different company discovers them from these owners. As shown in Fig. 2a, the US has approximately 8,000 drug discovery seeds, which is more than eight times compared with that of the other countries. As expected, the government and educational institutions tend to release the license at an early stage of the R & D process and the public company's license acquisition rate; for example, the number of leading firms increases as the process reaches the launched status. Furthermore, only in Japan (JPN), the number of drug pipelines at the discovery status is smaller than that at the launched status. Because of the lack of drug discovery seeds nowadays will decline the country's pharmaceutical industry in the future, countries such as JPN must obtain drug pipelines from other countries.

Figure 3 shows the knowledge flows in the drug pipelines between the licensor and licensee at the country level per pipeline status. Here we focus on the top 10 countries in Fig. 1 and aggregate the pre-registration and registered statuses because of the availability of a few pipelines. We ignore the drug pipelines whose license has not shifted to investigate the flows between different business entities. The diagonal elements correspond to the knowledge flows between companies in the country; thus, the knowledge flows from the firms in the US to the other countries in the drug pipelines. Furthermore, the knowledge of drug pipelines seems to flow not only in the direction from firms in the US to the others but also in the opposite direction.



Supply chain and ownership data

The global supply chain and ownership data in the year 2017 were constructed by collecting various companies' data from the Standard & Poor's (S&P) Capital IQ platform website CapitalIQ. The Capital IQ dataset covers more than 500,000 companies with information on business relations in 217 countries in 159 industrial sectors defined by the S&P, including all listed companies in the world. The data include company ID, company name, GBRN, location of the company, company type, and the primary industry



as node information. Industrial classification is based on the Global Industry Classification Standard, developed by Morgan Stanley Capital International and the S&P. The supply chain data also include examples of the business relationship between the supplier and customer as the edge information. Although various business relations that fall under suppliers are supplier, creditor, franchiser, licensor, landlord, lessor, auditor, transfer agent, investor relations firm, and vendor, most are supplier and creditor. Here, the supplier is a company providing the products or services, whereas the creditor is a

private, public, or institutional entity availing funds to others to borrow. Notably, the links in the dataset are dominated by supply chain business relationships. Therefore, the characteristics of the dataset reflect the global supply chain network.

The ownership data include a list of shareholding companies and individuals for each company as the link information. The list comprises the top 100 owner companies and individuals with the ownership ratio data. The link’s weight is the ownership ratio owned by the owner between 0% and 100%. The listed firms dominate firms in the ownership data for each country. Therefore, the coverage of ownership data is narrower than that of the supply chain data (Table 2).

Notably, the collection years for the two datasets, namely, the drug pipeline data and the company’s supply chain and ownership data, are different because of data availability. The former data are for 2013, whereas the latter is for 2017. The four-year difference in the data on the rapidly changing pharmaceutical industry might be a weakness. In this paper, we clarify the relationship among the knowledge flows, the flow of goods, and capital flow in the global pharmaceutical industry by analyzing the data as the MLN. The study is the first attempt and therefore considered significant enough.

Methodology

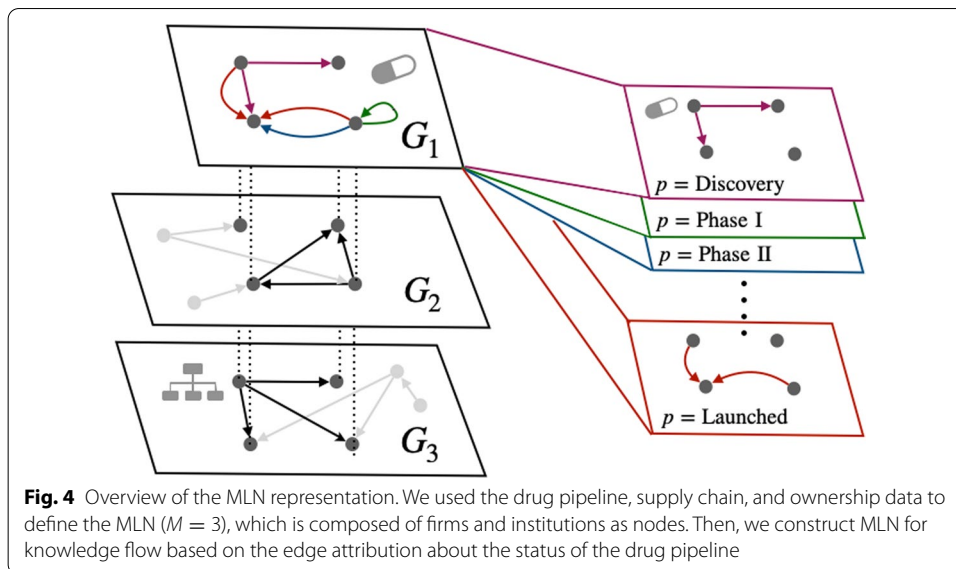
This study’s primary purpose is to reveal the dynamics of propagation and the localization of knowledge in the drug development cycle by analyzing the drug pipeline and supply chain and ownership data. For this research, we constructed an MLN, simultaneously representing three types of relationships between companies and institutions regarding the knowledge flow in the drug pipeline. Before we move to the MLN analysis, we explain the definition of the networks of each layer.

Drug pipeline network

We represent the drug pipeline data acquired from Clarivate Analytics as a drug pipeline network G_1 in Fig. 4. In the graph $G_1 = (V_1, E_1)$, V_1 is a node-set constructed from companies, government, and educational institutions listed in the drug pipeline data, and E_1 is an edge set constructed from licensor–licensee relations on drug pipelines. Let us denote that a directional edge is present as $i \rightarrow j$ in the drug pipeline network, where node i is the licensor, and node j is the drug candidate’s licensee in the drug pipeline data. As mentioned in Sect. 2, nodes are various types of business entities categorized as a government institution, educational institution, private company, and public company. One drug pipeline also does not always define one edge in the drug pipeline network because there is the case that the originator of the drug candidate shares the license with two companies. We denote the adjacency matrix of G_1 as $A^{[1]} = (a_{ij}^{[1]})$, where the element $a_{ij}^{[1]}$ corresponds to the number of edges from node i

Table 2 The numbers of nodes, edges, and self-loops in supply chain and ownership data

	Supply chain	Ownership
# of nodes	503,840	69,552
# of edges	1,379,344	1,147,937
# of self-loops	5	24



to j . The superscript represents the layer index. The in- and out-degrees of node i are defined by $k_{in,i}^{[1]} = \sum_j a_{ji}^{[1]}$ and $k_{out,i}^{[1]} = \sum_j a_{ij}^{[1]}$. In the graph G_1 , $k_{in,i}^{[1]}$ and $k_{out,i}^{[1]}$ represents the total number of drug pipeline that the company i is dealing with, and the total number of drug pipeline that the company i has discovered, respectively. Moreover, the drug pipeline data contains the case that the drug pipeline’s licensee corresponds to the licensor. In other words, the graph G_1 contains self-loops, and the number of self-loops of node i are computed as $\ell_i^{[1]} = a_{ii}^{[1]}$. In order to distinguish the shifts of license with the other, we denote the number of edges of node i from/to different nodes as in-/out-degree removed self-loops,

$$m_{in,i}^{[1]} = k_{in,i}^{[1]} - \ell_i^{[1]} \quad \text{and} \quad m_{out,i}^{[1]} = k_{out,i}^{[1]} - \ell_i^{[1]}. \tag{1}$$

Thus, $m_{in,i}^{[1]}$ and $m_{out,i}^{[1]}$ represents the number of license of drug candidates that the company i has transferred to the others, and that the company i has obtained from the others, respectively. Therefore, in the graph G_1 , the edges ($i \neq j$) are knowledge flows in the drug pipelines. However, the self-loops ($i = j$) represents the localization of the knowledge in one company.

Furthermore, the drug pipeline data is a snapshot data and a record at a particular development stage. Thus, we use the status of the drug development cycle as edge attribution, $E_1 = \bigcup_p E_1^p$, and E_1^p represents the drug pipeline at the status $p \in \{\text{Discovery, Phase I–III Clinical, Pre-registration, Registered, Launched}\}$. In response to the extension, we add the status index p to the definitions of the characteristics such as in-degree of the company i , $k_{in,i}^{[1,p]}$. For example, the launched drug without transfer of license is represented as a self-loop with the launched attribution in the network, and $\ell_i^{[1,p]}$ at $p = \{\text{Launched}\}$ represents the number of launched drugs for which the company i discovered and has launched. Although we cannot determine when drug candidates’ license is transferred because of data property, we can observe the tendency of knowledge flows in the drug pipeline between business entities in the pharmaceutical area.

Supply chain network

The supplier-customer relationship is most important at the company level in the real economy. We represent the supply chain data acquired from the S&P Capital IQ dataset as a supply-chain network G_2 in Fig. 4. In the graph $G_2 = (V_2, E_2)$, V_2 is a node-set constructed from companies listed in the supply chain data, and E_2 is a edge set constructed from these supplier-customer relations. The supply chain network is an unweighted directed network, representing the supply chain business relationships. We denote a directional edge as $i \rightarrow j$ when company i is a supplier to company j . The in- and out-degrees of node i are defined by $k_{in,i}^{[2]} = \sum_j a_{ji}^{[2]}$ and $k_{out,i}^{[2]} = \sum_j a_{ij}^{[2]}$. Moreover, we denote the number of edges of node i from/to different nodes as in-/out-degree removed self-loops, $m_{in,i}^{[2]} = k_{in,i}^{[2]} - \ell_i^{[2]}$ and $m_{out,i}^{[2]} = k_{out,i}^{[2]} - \ell_i^{[2]}$. In the graph G_2 , $m_{in,i}^{[2]}$ and $m_{out,i}^{[2]}$ represents the number of suppliers and customers of company i , respectively.

Ownership network

We define ownership networks G_3 in Fig. 4 based on the S&P Capital IQ dataset. In the graph $G_3 = (V_3, E_3)$, V_3 is a node-set constructed from companies listed in the ownership data, and E_3 is an edge set constructed from these ownership relations. We denote an edge from company i to j when company j has a stake in the company i . The in- and out-degrees of node i are defined by $k_{in,i}^{[3]} = \sum_j a_{ji}^{[3]}$ and $k_{out,i}^{[3]} = \sum_j a_{ij}^{[3]}$. Moreover, we denote the number of edges of node i from/to different nodes as in-/out-degree removed self-loops, $m_{in,i}^{[3]} = k_{in,i}^{[3]} - \ell_i^{[3]}$ and $m_{out,i}^{[3]} = k_{out,i}^{[3]} - \ell_i^{[3]}$. In the graph G_3 , $m_{in,i}^{[3]}$ and $m_{out,i}^{[3]}$ represents the number of companies which company i holds shares, and the number of shareholders of company i , respectively. Furthermore, the number of owners of each company was limited to 100 in the S&P Capital IQ dataset. Although the S&P Capital IQ dataset includes the ownership ratio, we define the ownership network as an unweighted directed network to increase the number of duplicated nodes between layers. Note that the ownership network in this paper represents the dependency flow, which is in the opposite direction to that typically used because we assume that the firm knowledge tends to flow to these owners.

Combining datasets

To construct the MLN in Fig. 4, we must combine the drug pipeline data and the company's supply chain and ownership data. Because the company's name in the two datasets is not always the same, we performed name identification between the two datasets by using information about the country and industry after removing abbreviations such as Inc and Corp. When multiple candidates arose, we checked them manually.

MLN representation

In this subsection, we generalize the definitions of our networks using an MLN framework. Generally, the MLN is a pair defined as $\mathcal{M} = (\mathcal{G}, \mathcal{C})$, where $\mathcal{G} = \{G_\alpha; \alpha \in \{1, \dots, M\}\}$ of the family of graphs $G_\alpha = (V_\alpha, E_\alpha)$, where the set of nodes of layer G_α is denoted as V_α , and M is the number of layers.

$\mathcal{C} = \{E_{\alpha\beta} \subseteq V_\alpha \times V_\beta; \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta\}$ is a set of interconnections between the nodes of different layers G_α and G_β with $\alpha \neq \beta$. We define the MLN ($M = 3$), composed of companies and institutions as nodes and three types of interactions between them, by using the drug pipeline, supply chain, and ownership data. Because each layer of our MLN is defined by the types of interactions, there are overlapped nodes but no interconnection between the nodes of different layers: $\mathcal{C} = \{\emptyset\}$. The graphs G_α for each layer are defined as

- G_1 : Drug pipeline network
- G_2 : Supply-chain network
- G_3 : Ownership network

where the definitions for each layer are explained in previous subsections. Figure 4 displays the conceptual representation of MLN.

The adjacency matrix of each layer G_α is denoted by $A^{[\alpha]} = (a_{ij}^{[\alpha]})$, where the element $a_{ij}^{[\alpha]}$ corresponds to the number of edges from node i to j in the α -th layer. The in- and out-degrees of a node i of the MLN are defined as vectors:

$$\mathbf{k}_{in,i} = (k_{in,i}^{[1]}, k_{in,i}^{[2]}, k_{in,i}^{[3]}) \quad \text{and} \quad \mathbf{k}_{out,i} = (k_{out,i}^{[1]}, k_{out,i}^{[2]}, k_{out,i}^{[3]}), \tag{2}$$

where $k_{in,i}^{[\alpha]}$ and $k_{out,i}^{[\alpha]}$ are the in- and out-degrees of node i in the α -th layer, i.e., $k_{in,i}^{[\alpha]} = \sum_j a_{ji}^{[\alpha]}$ and $k_{out,i}^{[\alpha]} = \sum_j a_{ij}^{[\alpha]}$. Similarly, we denote the number of self-loops of node i in the α -th layer as, $\ell_i^{[\alpha]} = a_{ii}^{[\alpha]}$. In this paper, we must recognize the flows as self-loops because self-loops in the drug pipeline network G_1 correspond to the accumulation of knowledge. Thus, we count the number of flows in each layer by the edges between two nodes and self-loops. The number of edges of node i from/to different nodes is denoted as

$$m_{in,i}^{[\alpha]} = k_{in,i}^{[\alpha]} - \ell_i^{[\alpha]} \quad \text{and} \quad m_{out,i}^{[\alpha]} = k_{out,i}^{[\alpha]} - \ell_i^{[\alpha]}, \tag{3}$$

and their total number is

$$M_\alpha = \frac{1}{2} \sum_{i,j=1, i \neq j}^{N_\alpha} a_{ij}^{[\alpha]} \quad \text{and} \quad L_\alpha = \sum_i^{N_\alpha} a_{ii}^{[\alpha]}. \tag{4}$$

where N_α is the number of nodes in α -th layer. Notably, we added the status layer index p to the definitions of the first layer characteristics such as in-degree of the i -th node $k_{in,i}^{[1,p]}$ and the total number of edges M_1^p .

Bow tie structure

As we showed in the edge-level analysis in Sect. 2, the knowledge flow in the drug pipeline seems to be circulated at the country level; however, whether the flows connect at the company level is unclear. Generally, the giant weakly connected components (GWCCs) of a directed network can be decomposed as giant strongly connected components (GSCCs), which is the largest size of the SCC in the GWC, its upstream and downstream portions (IN and OUT) known as the bow tie decomposition in the Web

Broder et al. (2000). This decomposition could help us understand the hierarchical and circular flows of the networks from a macroscopic perspective.

Community structure

Besides the macroscopic structure measured as a bow-tie structure, a community detection is a powerful tool for explaining densely connected networks' structural properties. We compared the structural properties between layers by using node attributes such as country and primary industry. Although we show the detailed analysis of community structure of each network in 6.2, we explain the method that is useful to extract community structures of a network here.

To find communities in the GWCC of the layers, we use the map equation method Rosvall and Bergstrom (2008), known as Infomap, which is one of the best performing community detection methods Lancichinetti and Fortunato (2009). The map equation method is a flow-based and information-theoretic approach to find an efficient code for minimizing the length of the description of the random walk for generating a module partition \mathcal{M} to divide n nodes into m communities. Then, the average single-step description length is defined as

$$L(\mathcal{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^m p_{i\circlearrowleft} H(\mathcal{P}_i). \quad (5)$$

The first term arises from the movements of the random walker across modules, where q_{\curvearrowright} is the probability that the random walker switches communities, and $H(\mathcal{Q})$ depicts the average description length of the community index codewords given by the Shannon entropy. The second term arises from the intra-community movement of the random walker, where the weight $p_{i\circlearrowleft}$ represents the fraction of the movements within the community, and $H(\mathcal{P}_i)$ represents the entropy of the intra-community movement. Furthermore, this method has been extended to a hierarchical map equation Rosvall and Bergstrom (2011) that decomposes a network into communities and sub-communities.

We detect the hierarchical communities by using the multi-coding Infomap method, and we use the "Level" index to represent the hierarchy of communities; communities at the 2nd level represent sub-communities at the 1st level. To characterize the hierarchical communities, we use node attributions with country, company type, and bow tie component for the drug pipeline layer, and country, primary industry, and bow tie component for the supply-chain and ownership layer, respectively.

Interlayer degree correlations

It is reasonable to assume that a hub node in the drug pipeline network could be a hub in the supply chain or ownership network. Thus, to verify this assumption, we investigate the degree correlations for each node between different layers. However, we must note that the edges in the drug pipeline layer E_1^p characterized by the status $p \in \{\text{Discovery, Phase I–III Clinical, Pre-registration, Registered, Launched}\}$, and the drug pipelines that have not reached the launched stage could disappear in the drug development cycle. Therefore, we only focus on the launched case here. Furthermore, each company's incoming edges in the drug pipeline layer are equal to the number of drugs with which

it is dealing. From this perspective, we divide the edges on which we focus in the drug pipeline layer into two types:

- Self-loops, $\ell_i^{[1,p]}$ at $p = \{\text{Launched}\}$, corresponding to the number of launched drugs that the company i discovered as the licensor and has launched. In this paper, we define that $\ell_i^{[1,p]}$ at $p = \{\text{Launched}\}$ represents the closed innovation.
- In-degrees removed self-loops, $m_{in,i}^{[1,p]}$ at $p = \{\text{Launched}\}$, corresponding to the number of launched drugs for which company i was not licensed originally but owns because of the licensee. In this paper, we define that $m_{in,i}^{[1,p]}$ at $p = \{\text{Launched}\}$ represents the open innovation.

Node and edge overlap

Although the dynamics of knowledge flows in the drug pipeline between companies remain unclear, we can raise a possible hypothesis for the edge-level similarity. When drug pipelines develop in a too closed situation, such that pharmaceutical companies do not have business with the same industry, the supply chain’s edge-level similarity ceases to appear. However, the alliances between the pharmaceutical companies, i.e., open innovation in the pharmaceutical industry, could share not only markets but also drug pipelines, which is assumed to appear as the edge-level similarity to the supply-chain layer. Furthermore, if the drug pipeline tends to be transferred to the owner of its licensor, we might observe the edge-level similarities to the ownership network. Although knowledge of the pharmaceutical industry might flow along with the flow of control, it is challenging to observe it in our MLN when the mergers and acquisition (M&A) causes it. To confirm the above hypothesis, we compute the overlapping of nodes and edges as

$$O(X_\alpha, X_\beta) = \frac{|X_\alpha \cap X_\beta|}{|X_\alpha \cup X_\beta|}, \tag{6}$$

where X_α is a set of nodes/edges at the α -th layer. We measure the overlap by the fraction of nodes/edges appearing in both layers over the aggregate number of nodes/edges of the two layers. We ignore the multiple edges and self-loops in the drug pipeline layer.

To evaluate the statistical significance of the finding stated above more precisely, we compute the probabilities (p values) when the expected number of overlapped edges is larger than the observed value by using a statistical test. Here, the null hypothesis is that we have no edge overlap between the two layers. First, we assume that the probability of generating the α -th layer having the x overlapping edges obeys the binomial distribution, $x \sim B(n, p(\beta|\alpha))$, where $n = |E_\alpha \cup E_\beta|$. We define the conditional probability that the edge connects between the two nodes of the overlapping knowledge flow layer ($\beta = 1$) and the α -th layer as

$$p(\beta = 1|\alpha) = p_1 \times p_\alpha = 1 \times \frac{\langle \bar{k}^{[\alpha]} \rangle}{|V_\alpha|}, \tag{7}$$

where $\langle \bar{k}^{[\alpha]} \rangle$ is the half value of the averaged total degree of the α -th layer. Therefore, the last term in the right-hand side of this equation corresponds to the probability of finding

the edges for randomly selected pairs of nodes. Here, the drug pipeline layer ($\beta = 1$) is independent variable of the α -th layer, $p_1 = 1$, by definition.

Results

In this section, first, the results of the single-layer analysis were explained, and then the multi-layer analysis was performed based on the results.

Single-layer analysis

In preparation for MLN analysis, we carried out an analysis of the bow-tie structure and the community structure of every single layer: the drug pipeline network, the supply chain network, and the ownership network. Bow tie analysis suggests that the knowledge flows in drug pipelines have similar macroscopic characteristics as those of the supply chain network. For a comparison of the structural properties between layers, we conducted the community detection using the Infomap method and characterized each community in terms of node attributes such as country and primary industry. For the detailed discussion of single-layer analysis, see Appendix 1. As discussed in Sect. 2, the knowledge flow between the licensor and licensee to circulate at the country level and the global level because of the US companies. We observed a large size of SCC in the bow-tie structure at the company level in the drug pipeline network, which is a large-scale circular flow between pharmaceutical companies.

Interlayer degree correlations

We constructed the MLN to investigate the knowledge flow between companies in drug pipelines using the drug pipeline, supply-chain, and ownership data. Table 3 summarizes the number of nodes, edges, and self-loops for each layer of the MLN, where we list the number of the cases successfully combined by the Coritellis and Capital IQ datasets in parentheses. Table 4 shows the knowledge flows in drug pipelines for each status of the drug development cycle.

We show the relations between the number of launched drugs and the number of supplier–customer (ownership) links in the upper (lower) part of Fig. 5. Note that the number of drugs corresponds to the edges attributed to $p = \{\text{Launched}\}$ at the drug pipeline layer, and Fig. 5 does not contain companies not dealing with launched drugs. This is the reason that the number of data points in Fig. 5 is less than the number of nodes in Table 3.

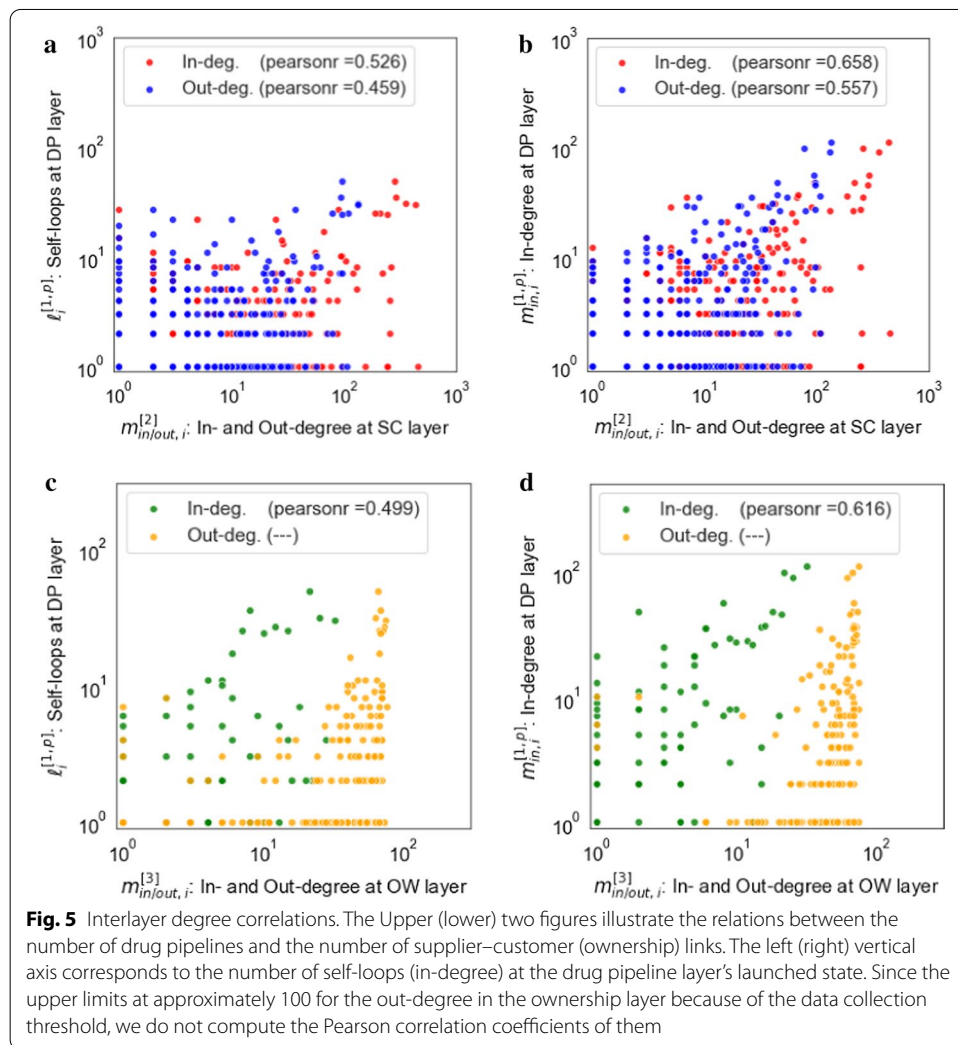
Table 3 The number of nodes, edges, and self-loops before combining Coritellis and Capital IQ datasets

	G_1 : Drug pipeline		G_2 : Supply chain		G_3 : ownership	
	Whole	(Combined)	Whole	(Combined)	Whole	(Combined)
N_α : # of nodes	7785	(2379)	503,840	(1954)	69,552	(830)
M_α : # of edges	11,478	(3769)	1,379,344	(3115)	1,147,937	(158)
L_α : # of self-loops	20,876	(10,927)	5	(0)	24	(0)

We also list the number after the combination

Table 4 The number of edges and self-loops for each pipeline status in the drug pipeline layer

p : pipeline status	M_p^E : # of edges	L_p^S : # of self-loops	Total
Discovery	4488	13,668	18,156
Phase I clinical	818	1716	2534
Phase II clinical	1255	1941	3196
Phase III clinical	538	630	1168
Pre-registration	165	189	354
Registered	120	155	275
Launched	4065	2527	6592



As defined in Sect. 3, $\ell_{in,i}^{[1,p]}$ and $m_{in,i}^{[1,p]}$ at $p = \{\text{Launched}\}$ represents the closed and open innovation, respectively. Thus, we can investigate the relation between layers from Fig. 5a and c for the closed innovation, and Fig. 5b and d for the open innovation, respectively. We also calculated the Pearson correlation coefficients as denoted on these figures. Note that

the upper limits at approximately 100 for the out-degree in the ownership layer because of the threshold in data collection. This is why we do not compute the Pearson correlation coefficients relating to out-degree in the ownership layer. Notably, the companies having larger $m_{in,i}^{[1,p]}$ at $p = \{\text{Launched}\}$ have a larger number of supplier–customer or ownership relations. As a result, we have weak and positive interlayer degree correlations between the drug pipeline and both supply chain and ownership. Especially, we observed more strong correlation in the open innovation, which is measured by $m_{in,i}^{[1,p]}$ at $p = \{\text{Launched}\}$ in Fig. 5b and d, than the closed innovation in Fig. 5a and c. Therefore, we could expect that open and closed innovation is useful for some companies with a larger number of suppliers–customer or ownership relations in the pharmaceutical industry– and they appear more remarkably in open innovation.

Node and edge overlap

Table 5 shows the node and edge overlap between the two layers. We demonstrate the overlapping of nodes and edges between the layers by measuring the fraction of nodes/edges appearing in both layers over the aggregate number of nodes/edges of the two layers. There is a lower proportion of node/edge overlap present at the ownership layer, which is more remarkable than the GWCC case.

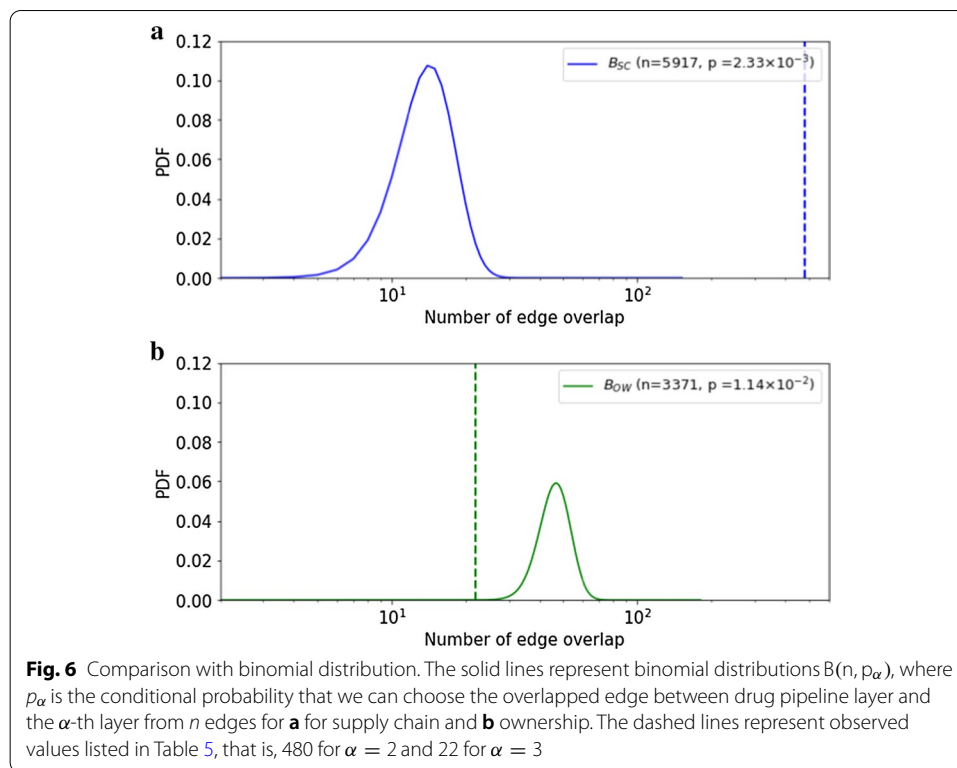
Table 5 shows that the overlap between the knowledge flow layer and the supply chain layer is more extensive than other combinations of layers in both the node overlap and edge overlap. Now, the conditional probabilities $p(1|\alpha)$ are given as

$$p(1|\alpha) = \begin{cases} 2.33 \times 10^{-3} & (\alpha = 2) \\ 1.14 \times 10^{-2} & (\alpha = 3) \end{cases} \text{ where } \langle \bar{k}^{[\alpha]} \rangle = \begin{cases} 2.68 & (\alpha = 2) \\ 1.01 & (\alpha = 3) \end{cases}, \tag{8}$$

Table 5 Overlap of nodes and edges between the layers, as measured by the fraction of nodes/edges appearing in both layers over the aggregate number of nodes/edges of the two layers

	Drug pipeline	Supply chain	Ownership
(a) Node overlap between two layers			
Drug pipeline	1.000		
Supply chain	0.730	1.000	
Ownership	0.042	0.057	1.000
(b) Edge overlap between two layers			
Drug pipeline	1.000		
Supply chain	0.081	1.000	
Ownership	0.007	0.007	1.000
(c) Number of node overlap between two layers $ V_\alpha \cap V_\beta $			
Drug pipeline	1574		
Supply chain	1149	1149	
Ownership	67	66	74
(d) Number of edge overlap between two layers $ E_\alpha \cap E_\beta $			
Drug pipeline	3318		
Supply chain	480	3079	
Ownership	22	23	75

We ignore the multiple edges in the drug pipeline layer



for the GWCC of supply chain ($\alpha = 2$) and ownership ($\alpha = 3$) network. Figure 6 compares the comparisons of the actual number of edge overlaps with the above binomial distributions, where Fig. 6a shows the supply chain and (b) the ownership network. Consequently, the p-value shows a value approximately equal to zero (one) for the supply-chain (ownership) layer. Therefore, the null hypothesis is rejected (adopted) for the supply chain (ownership) layer, proving that the observed edge overlaps between the knowledge flow and the supply chain (ownership) layers are (are not) statistically significant.

Therefore, the pharmaceutical industry's knowledge flows are related to the flow of products associated with the supply chain, rather than the dependency based on the ownership network. This suggests that pharmaceutical companies use open innovation to share the drug pipeline and spread the business market as the supply chain. Although we do not use the ownership data, including M&A, the relation between the knowledge flow of the drug pipeline and ownership network is rare. Future work, therefore, should compare the traditional M&A strategy of pharmaceutical companies with their sustainable growth.

The similarity between the knowledge flows of drug pipelines and supply-chain networks is also observed at the edge level, i.e., the licensor (licensee) of drug pipelines tends to be the supplier (customer), and a hub of knowledge flows of the drug pipelines tends to be their hub in the supply chain. Our results suggest a strong connection between an open innovation in the pharmaceutical industry and firms' activities regarding the supply chain. Generally, the pharmaceutical industries promote drug discovery and a clinical trial using artificial intelligence (AI) methods. For example, large pharmaceutical

companies, such as Pfizer and Novartis, deployed AI systems for drug discovery and clinical trials by IBM Pharmatutor.org, Fleming (2018). This collaboration suggests an increase in the need for the relation between drug pipelines and supply chains in other industries. Based on the analysis, our findings, therefore, agree with the reported situation of pharmaceutical innovation.

Implications

Open innovation is a lucrative field within industries relying on innovation. Specifically, the pharmaceutical industry represents knowledge-intensive industries, and open innovation should significantly boost pharmaceutical collaborations and reduce the cost of long-term R & D processes or the M & A in closed innovation.

Consequently, the development of a new drug and trial requires high investment, but with a low possibility of success. Private companies, such as small and medium enterprises and ventures, end up funding innovations because of high investments. Furthermore, public companies, such as big pharmaceutical companies, are always looking for new available targets. Thus, to boost pharmaceutical innovation, based on globalized and IT innovation, it is crucial to understand the dynamics of innovation regarding the firm-level economy's various interactions.

To reveal pharmaceutical innovation's actual situation, we proposed an MLN framework using the drug pipeline, global supply chain, and ownership data. Furthermore, we investigated the characteristics of the knowledge flows between the licensor and licensee, defined by the drug pipeline data, from the proposed framework. Several seeds of the drug pipeline were discovered in the US and provided to other countries. Larger firms, such as public companies having an individual market in the supply chain, can hold many drug pipelines as a licensee, causing pharmaceutical knowledge flows cross-borders among countries. The macroscopic structure of pharmaceutical knowledge flows provides us knowledge about the characteristics of each country. Because the US firms provide drug pipeline seeds, they compose the upper stream of the entire network of the knowledge flow of the drug pipeline. Although firms in JPN do not have many drug seeds, they deal with the second-largest launched drug pipeline. Several firms in JPN are interdependent with others on drug development. From a macroscopic structure, we can confirm that most of the firms in JPN are located in the largest size of the SCC. By contrast, the firms in CHN, having a high self-sufficiency rate of drug pipelines, do not involve the circular flow of knowledge of drug pipelines with others. Furthermore, the tendency of closed innovation in CHN is seen from a closed community of pharmaceutical firms in CHN in the global supply chain network.

Finally, an essential question for future studies is to understand how some operations, such as policy making, could boost pharmaceutical innovation. Thus, our study provides a framework for future studies to assess pharmaceutical innovation's characteristics regarding various firm-level activities by using MLN representation. Although our results are inspiring, more investigations are required for further improvements. The traditional M&A strategy of pharmaceutical companies could be evaluated based on their sustainable growth by combining the historical M&A data with our networks. Furthermore, our approach can investigate how the risk, patent cliff in the pharmaceutical

industry will spread to other sectors in the future. We hope that future studies could reduce unnecessary risk.

Conclusion

Drug development is time-consuming from the start of research to obtaining approval, and the probability of success with a candidate compound is extremely low. We evaluated the characteristics of the flow and localization of knowledge during drug development in the global pharmaceutical industry.

We analyzed the MLN constructed with the drug pipeline network layer, the global supply chain network layer, and the global ownership network layer. First, we focused on the flows in an individual layer of MLNs, licenses as knowledge, goods, and capital. We identified the bow tie and the community structure of each network layer. The obtained bow-tie structure suggested that the knowledge flows in drug pipelines have macroscopic characteristics similar to the supply chain network, characterized as the larger SCC size. The obtained communities in three layers were characterized by country, category of the company, and bow tie component.

After analyzing individual layers, we studied the knowledge flows in the MLN, primarily whether the knowledge flows are related to the flow of goods in the supply chain or the flow of capital in the ownership. Based on the results obtained from the MLN analysis, we conducted a statistical test. We verified the significance of the overlapping of edges between drug pipeline and supply-chain layers. Our results suggest a strong connection between open innovation in the pharmaceutical industry and its economic activities in the supply chain.

Acknowledgements

We are grateful to Y. Fujiwara, H. Aoyama, H. Iyetomi, and H. Yoshikawa for their insightful comments and encouragement. We would like to thank the anonymous referees for helpful comments on earlier drafts of the manuscript.

Authors' contributions

HG provided key contributions in performing the network analysis and the visualization. WS conceptualized multilayer network representation. MJ made a survey of the related work, conceptualized the drug pipeline. YI supervised the project and formulated of the statistical test of node and edge overlap between network layers. All authors read and approved the final manuscript.

Funding

The present study was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grants-in-Aid for Scientific Research (B), Grant No. 17KT0034 (2017-2019).

Data availability statement

The "Cortellis Competitive Intelligence" data are available from Clarivate Analytics (website: <https://clarivate.com/cortellis/solutions/competitive-intelligence-and-analytics/>), however, restrictions apply to the availability of these data. The "Capital IQ platform" data that support the findings of this study are available from Standard & Poor's (website: <https://www.spglobal.com/marketintelligence/en/solutions/sp-capital-iq-platform>) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Standard & Poor's offers various company-level data in accordance with customers' demand, such as detailed company descriptions, corporate structures, global supply chain data, global ownership data, etc.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ College of Science and Technology, Nihon University, Chiba 2748501, Japan. ² National Institute of Science and Technology Policy, Tokyo 1000013, Japan. ³ Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Kyoto 6068306, Japan.

Appendix 1: Single-layer analysis

We summarize the bow-tie structure's characteristics and the community structure of the drug pipeline network, the supply chain network, and the ownership network.

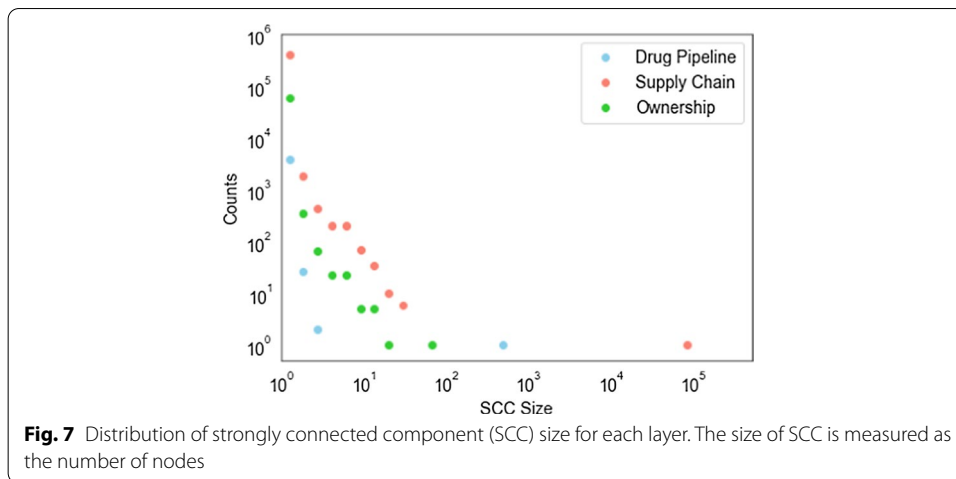


Table 6 The sizes of different components based on the bow tie decomposition for each layer

Components	Drug pipeline		Supply chain		Ownership	
	Counts	Ratio (%)	Counts	Ratio (%)	Counts	Ratio (%)
IN	1020	23.88	100,976	21.17	33,011	56.82
GSCC	498	11.66	80,225	16.82	63	0.11
OUT	1313	30.74	218,344	45.79	501	0.86
TE	1440	33.72	77,332	16.22	24,525	42.21
GWCC	4271	100.00	476,877	100.00	58,100	100.00

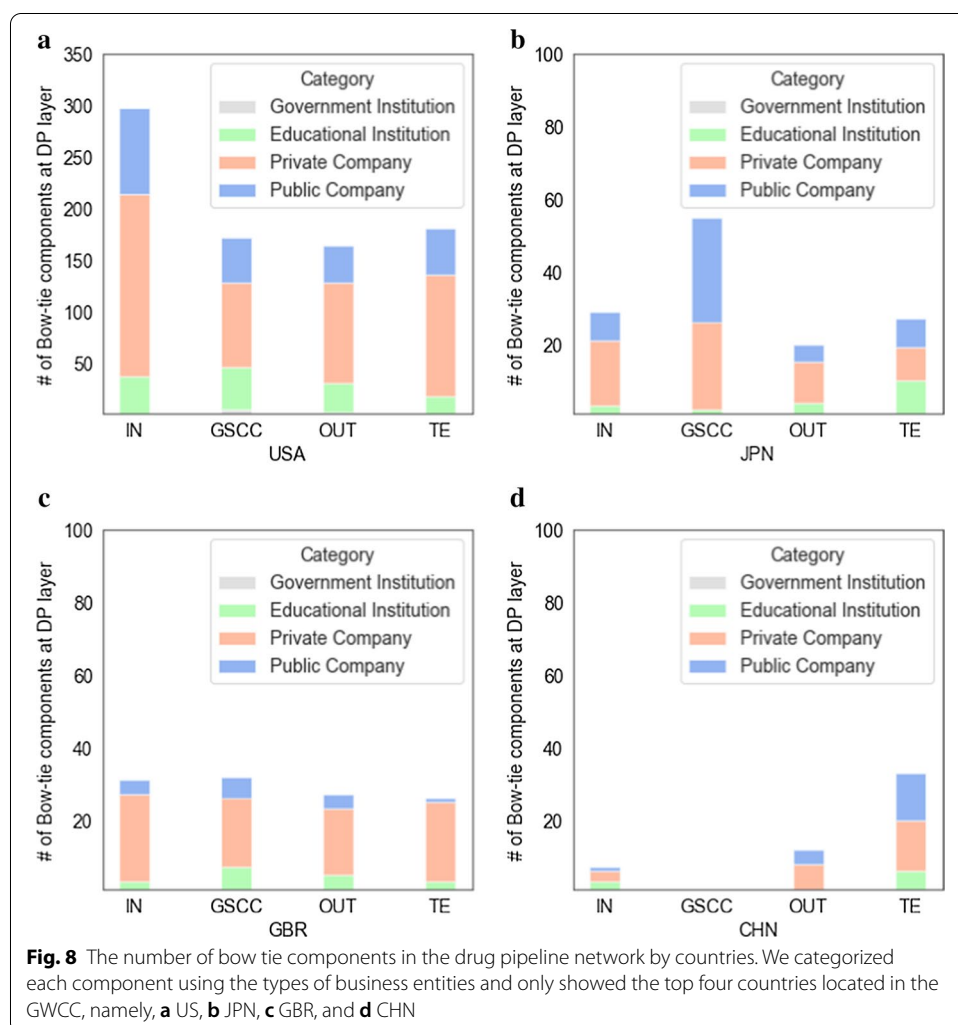
Bow tie structure

Although the knowledge flow in the drug pipeline seems to be circulated at the country level by the edge-level analysis, whether the flows connect at the network level or not remains unclear.

First, we computed the number and sizes of SCC for each layer. We show the distributions of SCC size for each layer in Fig. 7. Compared to the size of GSCC in the supply chain layer with the second-largest SCC, it is clear that the supply chain network contains an immensely large circular flow. On the other hand, there is no significantly larger SCC than the others in the ownership network. In the drug pipeline network, it is apparent that only small SCCs except for GSCC.

We show the bow-tie decomposition results for each layer network in Table 6, where we define the GWCC components not belonging to the GSCC, IN, and OUT components as Tendril (TE). The ownership network has an entirely different bow-tie structure than the others. The ownership network does not have a large SCC, such as a large circular flow of control. Conversely, the drug pipeline layer’s bow-tie structure has similar fractions of components than those of the supply-chain layer. Our results suggest that the knowledge flows in drug pipelines have similar macroscopic characteristics as those of the supply chain network, especially the larger SCCs in the flows.

Furthermore, we investigated the role of each country in the knowledge circulation of the drug pipeline layer. Figure 8 shows the number of each bow-tie component of the drug pipeline layer by country, where we listed the top four countries located in GWCC. We categorized each component using the type of business entities; Private Company, Public Company, Educational, and Government Institutions. As shown in Fig. 8a, most companies in the US are located on the IN component, which is consistent with most drug pipelines at discovery status in the US (Sect. 2). Figure 8b shows that several public companies in JPN belong to GSCC of the drug pipeline networks. Although the GWCC in the drug pipeline layer is composed mainly of firms in the US, these companies should provide the drug pipeline seeds to the whole system, and the leading company in JPN contributes to the circular flow of knowledge. Furthermore, as shown in Fig. 8c and d, the companies in CHN does not contribute to the largest SCC in the GWCC, and this structure is not found in other countries.



Community structure

For a comparison of the structural properties between layers, we conducted the community detection known as the Infomap method and characterized each community using node attributes such as country and primary industry. We detected the hierarchical communities by using the multi-coding Infomap method. Table 7 summarizes the statistics in terms of hierarchical communities, where “Level” index represents the hierarchy of communities; communities at the 2nd level represent sub-communities at the 1st level. “#com.” is the number of all communities at each level, “#irr.com.” is the number of irreducible communities, which are communities that do not have any sub-communities. We determined the number of nodes “#nodes” such as firms and institutions, in irreducible communities at each level. We find that most of the firms belong to 2nd level communities except for the ownership network layer. Therefore, we limit our discussion of the properties of sub-communities to those of the 2nd level.

In Table 8, we show the characteristics of the largest community at the first layer for the drug pipeline layer and the three largest communities at the second level, where we use the attributions of country, bow tie component, and category of the company for this layer, and measure the community size by using the number of nodes in the community. Then, comparing the fraction of attributions at the first and second levels reveals that each community has similar characteristics, including 10% of public companies. The knowledge of the drug pipelines flows into the set of several private companies and a few leading firms and is widely distributed in the largest community.

Second, we demonstrate the characteristics of the communities in the other layers, using the primary industry attributes instead of the category of companies. As shown in Table 9, the primary industry of the firms forms the communities at the first level in the supply-chain layer, and the third-largest community labeled fourth has a significant fraction of the pharmaceutical sector. To investigate the relationship between the knowledge flows in the pharmaceutical industry, we investigate the second-level sub-communities in the community labeled fourth in Table 10. Note that the community comprising the pharmaceutical firms in CHN is the second-largest sub-community at the supply chain layer. These trends align with the previously discussed structure of the knowledge flows in drug pipelines in CHN, which is too close.

Finally, Table 11 shows the characteristics of the six largest communities at the first level. Table 12 shows the characteristics of the three largest sub-communities of the

Table 7 Statistics for the communities detected by using the multi-coding Infomap method

Level	G ₁ : Drug pipeline			G ₂ : Supply chain			G ₃ : Ownership		
	#com.	#irr.com.	#nodes	#com.	#irr.com.	#nodes	#com.	#irr.com.	#nodes
1	784	778	778	425	70	479	447	429	709
2	410	320	2754	41,532	35,586	381,343	7819	6689	29,953
3	287	279	701	22,591	22,155	92,599	12,185	11,671	23,706
4	16	16	38	1028	1022	2425	1731	1687	3531
5				15	15	31	112	112	201

“#com.” is the number of all communities, “#irr.com.” is the number of irreducible communities, which are communities that do not have any sub-communities. “#nodes” represents the number of nodes such as firms and institutions in irreducible communities

Table 8 Characteristics of the largest community at the first level for the drug pipeline layer and the three largest sub-communities at the second level

G_1 : Drug pipeline

Index	Size	Country	Category	Bow tie
(a) the largest community				
1:	3380	US (48.5)	Private company (76.7)	TE (31.5)
		JPN (7.6)	Public company (12.6)	IN (27.3)
		GBR (7.2)	Educational institution (9.9)	OUT (27.1)
				GSCC (14.1)
(b) 3 largest sub-communities				
1:1:	68	US (66.7)	Private company (73.5)	OUT (36.8)
		ISR (7.4)	Educational institution (11.8)	IN (32.4)
			Public Company (10.3)	GSCC (17.6)
				TE (13.2)
1:3:	66	US (59.4)	Private company (83.3)	OUT (37.9)
		CAN (6.2)	Public company (10.6)	GSCC (25.8)
		JPN (6.2)		TE (18.2)
		GBR (6.2)		IN (18.2)
1:7:	63	US (50.0)	Private company (81.0)	OUT (42.9)
		CAN (7.7)	Public company (12.7)	IN (28.6)
		GBR (7.7)	Educational institution (6.3)	TE (15.9)
				GSCC (12.7)

The figures in parentheses indicate the fraction of attribution, and we have only listed the attributions having a share of more than 5%

largest community at the 1st level in the ownership layer. Consequently, based on ownership, there are no significant characteristics in the largest communities. However, at this level, it is challenging to relate the knowledge flows of the drug pipelines in this stage, future work should focus on the relation to the community of dependency flow, in contrast to the knowledge flows by the mergers and acquisitions (M&A).

Table 9 Characteristics of the six largest communities at the first level for the supply chain layer

G₂: supply chain

Index	Size	Country	Primary industry	Bow tie
3:	43,636	US (32.3)	Application software (7.6)	OUT (53.5)
		CHN (9.1)	Technology distributors (7.3)	IN (17.3)
		JPN (7.5)	Semiconductors (5.2)	GSCC (15.1)
		TWN (6.3)		TE (14.1)
1:	30,972	US (43.2)	Application software (8.8)	OUT (44.5)
		CHN (9.4)	Movies and entertainment (7.5)	IN (22.9)
		GBR (7.1)		GSCC (20.2)
4:	30,430	US (52.0)	Pharmaceuticals (15.7)	OUT (35.6)
		CHN (6.0)	Biotechnology (11.1)	IN (28.0)
		GBR (5.2)	Health care equipment (8.2)	GSCC (19.7)
			Health care facilities (7.0)	TE (16.7)
5:	28,298	IND (70.4)	Construction and engineering (5.3)	OUT (48.6)
		US (6.1)		GSCC (23.8)
2:	25,365	US (47.6)	Apparel, accessories and luxury goods (8.5)	OUT (41.5)
		GBR (5.8)	Packaged foods and meats (7.1)	IN (23.7)
		CHN (5.4)		GSCC (17.7)
6:	17,857	US (36.2)	Oil and gas exploration and production (15.1)	OUT (36.8)
		CHN (8.3)	Oil and gas storage and transportation (7.8)	IN (27.7)
		GBR (5.6)	Oil and gas equipment and services (6.7)	GSCC (23.2)
			TE (12.4)	

The figures in parentheses indicate the fraction of attribution, and we have only listed the attributions having a share of above 5%

Table 10 Characteristics of the three largest communities in the fourth community at the second level for the supply chain layer

G₂: Supply chain

Index	Size	Country	Primary industry	Bow tie
4:2:	792	US (97.0)	Construction and engineering (70.5)	IN (94.3)
4:7:	509	CHN (99.0)	Health care distributors (47.5)	OUT (41.8)
			Pharmaceuticals (35.6)	GSCC (31.2)
			Biotechnology (7.6)	IN (26.9)
4:22:	364	US (47.0)	Biotechnology (52.4)	TE (84.6)
		DEU (7.5)	Life sciences tools and services (17.7)	OUT (14.6)
		GBR (6.9)	Health care equipment (8.1)	
			Health care distributors (6.5)	
		Pharmaceuticals (5.6)		

The figures in parentheses indicate the fraction of attribution, and we have only listed the attributions with a share above 5%

Table 11 Characteristics of the six largest communities at the first level for the ownership layer

*G*₃: Ownership

Index	Size	Country	Primary industry	Bow tie
1:	54845	US (21.9) IND (9.0) CAN (6.4) CHN (5.7) AUS (5.2) GBR (5.2) JPN (5.1)	Asset management and custody banks (10.1)	IN (56.8) TE (42.2)
7:	644	US (24.0) IND (10.3) CAN (6.2) AUS (5.9) JPN (5.4)	Asset management and custody banks (11.6)	IN (57.3) TE (41.1)
6:	532	US (22.2) IND (8.3) CAN (7.5) AUS (6.2) JPN (5.8) CHN (5.1)	Asset management and custody banks (11.1)	IN (59.2) TE (39.7)
3:	525	US (22.3) IND (9.2) CHN (5.9) CAN (5.9)	Asset management and custody banks (11.8)	IN (57.3) TE (41.9)
12:	403	US (18.9) IND (10.4) CAN (7.7) AUS (5.7)	Asset management and custody banks (10.5)	IN (54.8) TE (44.2)
16:	364	US (24.4) CAN (11.3) GBR (8.7) IND (8.7)	Asset management and custody banks (10.2)	IN (57.5) TE (41.2)

The figures in parentheses indicate the fraction of attribution, and we have only listed the attributions with a share above 5%

Table 12 Characteristics of the largest community at the first level for the ownership layer and the three largest communities at the second level

*G*₃: Ownership

Index	Size	Country	Primary industry	Bow tie
1:1:	8,860	US (22.4) IND (8.8) CAN (6.2) CHN (5.8) GBR (5.4) JPN (5.1)	Asset management and custody banks (10.8)	IN (56.2) TE (42.8)
1:2:	2,761	US (21.2) IND (9.1) CAN (6.7) AUS (6.1) CHN (5.1) GBR (5.1)	Asset management and custody banks (10.0)	IN (58.1) TE (40.9)
1:5:	2,674	US (20.6) IND (9.1) CAN (7.6) GBR (5.5) CHN (5.4) AUS (5.3) JPN (5.1)	Asset management and custody banks (9.4)	IN (55.3) TE (43.5)

The figures in parentheses indicate the fraction of attribution, and we have only listed the attributions with a share above 5%

Received: 13 March 2020 Accepted: 19 October 2020

Published online: 02 November 2020

References

- Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Souma W (2010) *Econophysics and companies: statistical life and death in complex Business Networks*. Cambridge University Press, Cambridge
- Bianconi G (2013) Statistical mechanics of multiplex networks: entropy and overlap. *Phys Rev E* 87:062806. <https://doi.org/10.1103/PhysRevE.87.062806>
- Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A, Wiener J (2000) Graph structure in the web. *Comput Netw* 33(1–6):309–320
- Buldryev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464(7291):1025
- CapitalIQ: Capital IQ platform data that support the findings of this study are available from Standard & Poor's. <https://www.spglobal.com/marketintelligence/en/solutions/sp-capital-iq-platform/>
- Comanor WS, Scherer FM (2013) Mergers and innovation in the pharmaceutical industry. *J Health Econ* 32(1):106–113. <https://doi.org/10.1016/j.jhealeco.2012.09.006>
- Cortellis: Cortellis competitive intelligence data are available from Clarivate. <https://clarivate.com/cortellis/solutions/competitive-intelligence-and-analytics/>
- Donges JF, Schultz HCH, Marwan N, Zou Y, Kurths J (2011) Investigating the topology of interacting networks. *Eur Phys J B* 84(4):635–651. <https://doi.org/10.1140/epjb/e2011-10795-8>
- Fleming N (2018) How artificial intelligence is changing drug discovery. *Nature* 557(7706):55–55
- Kurant M, Thiran P (2006) Layered complex networks. *Phys Rev Lett* 96:138701. <https://doi.org/10.1103/PhysRevLett.96.138701>
- Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. *Phys Rev E* 80(5):056117
- Mazzola E, Bruccoleri M, Perrone G (2015) Supply chain of innovation and new product development. *J Purch Supply Manag* 21(4):273–284. <https://doi.org/10.1016/j.pursup.2015.04.006>
- Munos B (2009) Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov* 8(12):959
- Narayana SA, Pati RK, Vrat P (2014) Managerial research on the pharmaceutical supply chain - a critical review and some insights for future directions. *J Purch Supply Manag* 20(1):18–40. <https://doi.org/10.1016/j.pursup.2013.09.001>
- Osabe Y, Jibu M (2013) Development of new indicators for the launch of a japanese version of the nih (1): An overview and future prospects of pharmaceutical industry based on new indicators. *J Inf Process Manag* 56(7):448–458 in Japanese
- Osabe Y, Jibu M (2013) Development of new indicators for the launch of a japanese version of the nih (2): a technology-specific overview and future prospects of pharmaceutical industry. *J Inf Process Manag* 56(9):611–621 in Japanese
- Osabe Y, Jibu M (2014) Development of new indicators for the launch of amed (7): trend analysis for us nih funding. *J Inf Process Manag* 57(6):395–406 in Japanese
- Osabe Y, Jibu M (2014) Development of new indicators for the launch of amed (8): Knowledge flows of the pharmaceutical innovation system. *J Inf Process Manag* 57(8):562–572 in Japanese
- Osabe Y, Jibu M (2014) Development of new indicators for the launch of a japanese version of the nih (3): analysis of entities conducting pharmaceutical r&d activities. *J Inf Process Manag* 56(10):685–696 in Japanese
- Osabe Y, Jibu M (2014) Development of new indicators for the launch of a japanese version of the nih (4): refined indicators for patents linked with marketed drugs and pipelines. *J Inf Process Manag* 57(1):29–37 in Japanese
- Osabe Y, Jibu M (2014) Development of new indicators for the launch of a japanese version of the nih (5): analysis based on refined indicators for patents linked with marketed drugs and pipelines. *J Inf Process Manag* 57(3):178–186 in Japanese
- Osabe Y, Jibu M (2014) Development of new indicators for the launch of amed (6): a disease-specific overview and future prospects of pharmaceutical industry. *J Inf Process Manag* 57(5):323–333 in Japanese
- Pharmatutor.org: TOP 10 Pharmaceutical industries using Artificial Intelligence (AI). <https://www.pharmatutor.org/articles/top-ten-pharmaceutical-industries-using-artificial-intelligence>
- Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci* 105(4):1118–1123
- Rosvall M, Bergstrom CT (2011) Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6(4):18209
- Scherer FM (2010) Chapter 12 - pharmaceutical innovation. In: Hall BH, Rosenberg N (eds) *Handbook of the economics of innovation*, Vol. 1. Handbook of the economics of innovation, vol 1, pp 539–574. North-Holland (2010). [https://doi.org/10.1016/S0169-7218\(10\)01012-9](https://doi.org/10.1016/S0169-7218(10)01012-9). <http://www.sciencedirect.com/science/article/pii/S0169721810010129>
- Yang J, Wang W, Chen G (2009) A two-level complex network model and its application. *Phys A* 388(12):2435–2449. <https://doi.org/10.1016/j.physa.2009.02.046>
- Zou S-R, Zhou T, Liu A-F, Xu X-L, He D-R (2010) Topological relation of layered complex networks. *Phys Lett A* 374(43):4406–4410. <https://doi.org/10.1016/j.physleta.2010.08.073>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.