

RESEARCH

Open Access



# Community structure based on circular flow in a large-scale transaction network

Yuichi Kichikawa<sup>1\*</sup>, Hiroshi Iyetomi<sup>1,3</sup>, Takashi Iino<sup>1</sup> and Hiroyasu Inoue<sup>2</sup>

\*Correspondence:

[ykichikawa@m.sc.niigata-u.ac.jp](mailto:ykichikawa@m.sc.niigata-u.ac.jp)

<sup>1</sup>Faculty of Science, Niigata University, Niigata, 950-2181, Japan  
Full list of author information is available at the end of the article

## Abstract

The objective of this study is to shed new light on the industrial flow structure embedded in microscopic supplier-buyer relations. We first construct directed networks from actual data from interfirm transaction relations in Japan; as one example, the dataset compiled by the Tokyo Shoko Research, Ltd. in 2016 contains five million links between one million firms. Then, we analyze the industrial flow structure of such a large-scale network with a special emphasis on its hierarchy and circularity. The Helmholtz-Hodge decomposition enables us to break down the flow on a directed network into two flow components: gradient flow and circular flow. The gradient flow between a pair of nodes is given by the difference of their potentials obtained by the Helmholtz-Hodge decomposition. The gradient flow runs from a node with higher potential to a node with lower potential; hence, the potential of a node shows its hierarchical position in a network. On the other hand, the circular flow component illuminates feedback loops built in a network. The potential values averaged over firms classified by the major industrial category describe hierarchical characteristics of sectors. The ordering of sectors according to the potential agrees well with the general idea of the supply chain. We also identify industrially integrated clusters of firms by applying a flow-based community detection method to the extracted circular flow network. We then find that each of the major communities is characterized by its main industry, forming a hierarchical supply chain with feedback loops by complementary industries such as transport and services.

**Keywords:** Production network, Helmholtz-Hodge decomposition, Visualization

## Introduction

In general, interactions between individuals are considered to play an important role in the economy. For instance, firms are connected to each other directly or indirectly through their business transactions. A firm buys materials from suppliers and sells its products to customers. These transactions are so essential to firms that one cannot isolate the dynamics of individual firms from the entire economic system. Firms' production activities thus give rise to a complex network; also, examining economic phenomena from the perspective of networks can provide a variety of new insights into economic phenomena.

Conventionally, the industrial structure and economic ripple effects have been studied on the basis of the input-output tables (Leontief 1986). Furthermore, a network-theoretic point of view was incorporated into the input-output analysis to elucidate complex interindustrial flow structures within or across the sectors (Slater 1977; 1978; Carvalho

2008; McNerney et al. 2013; Contreras and Fagiolo 2014). However, such classification of firms by industry may be too formal for a reliable macroscopic picture of the economy.

Recently, firm-level network analyses based on a comprehensive database of interfirm transaction relations have begun to appear (Atalay et al. 2011; Acemoglu et al. 2012; Cainelli et al. 2012; Luo et al. 2012; Watanabe et al. 2015; Letizia and Lillo 2018; Goto et al. 2017). Economists as well as physicists have recognized the importance of taking an explicit account of interfirm links in order to understand economic issues, such as the origin of business cycles and the possibility of a chain reaction in firm bankruptcies.

Very recently, we have studied (Chakraborty et al. 2018) the structure of a Japanese production network with one million firms and five million supplier-customer links. We first constructed a directed production network from the actual data of interfirm transaction relations and found that they form a tightly knit structure with a giant strongly connected component surrounded by two half-shells constituting incoming-flow and outgoing-flow components for the core. The hierarchical structure of communities was then elucidated by a flow-based multilevel community detection method (Rosvall and Bergstrom 2011), and most of the irreducible communities were found to be on the second level. The composition of some of the major communities, including overexpressions of industrial and regional components, as well as hierarchical connections between the communities, was studied in detail.

The hierarchy of the production network is expected to emerge from self-organization of the supply chain in the industrial system. This is the general view on evolutionary processes in complex systems (Holland 2000; Anderson 1972). Here, we emphasize that we should also pay attention to the inner loops of production, giving rise to a nonlinear feedback mechanism in the system, because they can be engines for economic growth. The priority production system adopted by the Japanese government just after World War II is an illustrative application of this idea (Vestal 1995). It was intended to stimulate recovery of the nation's economy so damaged by concentrating public investment into coal mining and steel production. Production of steel needs electricity generated from coal, and mining of coal needs machinery made of steel. Such an industrial loop formed by the three industries, mining, steel production, and machinery manufacturing, led to autonomous growth of the economy.

The objective of this study is to advance the previous empirical analysis (Chakraborty et al. 2018) on the industrial flow structure embedded in microscopic supplier-buyer relations with a special emphasis on its circularity. To delve further into the flow structure of the transaction network with firms as nodes, we take advantage of a mathematical tool called the Helmholtz-Hodge decomposition (Jiang et al. 2011; Bhatia et al. 2013). It allows us to decompose the flow on a directed network into a gradient flow component and a circular flow component.

This paper is organized as follows. First, we give salient features of the dataset to be analyzed for transaction relations between firms in Japan. “[Helmholtz-Hodge decomposition](#)” section is devoted to a mathematical formulation of the Helmholtz-Hodge decomposition for the present analyses. In “[Bow-tie decomposition](#)” section, we revisit the walnut structure of the Japanese production network that was found in the previous study (Chakraborty et al. 2018), using the bow-tie decomposition and a visualization technique of networks. In “[Results and discussion](#)” section, we build a network consisting of only the circular flow components and detect communities in the network to

elucidate circular flow structure in the production network. The final section summarizes the results obtained here.

### Interfirm transaction data

The present analysis is based on the big data of 4,974,802 transaction relations between 1,066,037 firms in Japan that was collected by the Tokyo Shoko Research, Ltd. (TSR) in 2016.<sup>1</sup> These data virtually cover the entire amount of industrial activities in Japan. We regard firms as nodes and transaction relations between them as directed links spanning from suppliers to customers to construct the latest production network in Japan. Since information on the volume of each transaction is not available, we assume that all the links have the same weight.

In addition to the information on transactions between firms, various attributes of individual firms are available. For simplicity of analyses, we use two attributes of each firm, namely, industrial sector and geographical location of the head office. Firms are categorized into 20 sectors and 47 prefectures. Readers are referred to the previous paper (Chakraborty et al. 2018) for more detailed information on the dataset.<sup>2</sup>

### Helmholtz-Hodge decomposition

In general, one can write flow  $F_{ij}$  running from node  $i$  to node  $j$  in a directed network as follows:

$$F_{ij} = F_{ij}^{(p)} + F_{ij}^{(c)}, \quad (1)$$

where we assume that the magnitude of  $F_{ij}$  on the network is given by the following:

$$|F_{ij}| = \begin{cases} 1 & \text{(singly connected in one way)} \\ 0 & \text{(doubly connected in both ways)} \\ 0 & \text{(not connected)} \end{cases} \quad (2)$$

Since information on the volume of transactions is not available in the TSR dataset, we adopt such a simplified flow structure. The first term  $F_{ij}^{(p)}$  on the right-hand side of Eq. (1) denotes the gradient flow from node  $i$  to node  $j$  which is given by the following:

$$F_{ij}^{(p)} = w_{ij} (\phi_i - \phi_j), \quad (3)$$

where  $\phi_i$  is the Helmholtz-Hodge potential associated with node  $i$  and  $w_{ij}$  is a positive weight for linkage between nodes  $i$  and  $j$ . We assume that the weight  $w_{ij}$  takes the following values depending on how the two nodes are connected:

$$w_{ij} = \begin{cases} 1 & \text{(singly connected in one way)} \\ 2 & \text{(doubly connected in both ways)} \\ 0 & \text{(not connected)} \end{cases} \quad (4)$$

The Helmholtz-Hodge potential of nodes in a directed network identifies their hierarchical positions in its flow structure. In the network built with only gradient flow, nodes are perfectly ranked; the gradient flow always runs from a node with higher potential to a node with lower potential. On the other hand, the second term  $F_{ij}^{(c)}$  denotes the circular

<sup>1</sup>This is the largest connected component in the network obtained from the original data, containing 99.3% of all active firms listed in the data.

<sup>2</sup>In this study, sector I (Wholesale & retail trade) in (Chakraborty et al. 2018) was separated into two sectors, Wholesale and Retail trade.

flow component in which incoming flow and outgoing flow are exactly balanced at each node:

$$\sum_j F_{ij}^{(c)} = 0, \tag{5}$$

so that there is no hierarchy among nodes in the circular flow network. The circular flow component illuminates feedback loops embedded in the system.

Additionally, one can determine the potential  $\phi_i$  for every node by minimizing the squared difference between the actual flow and the gradient flow:

$$I = \frac{1}{2} \sum'_{i < j} w_{ij}^{-1} \left( F_{ij} - F_{ij}^{(p)} \right)^2, \tag{6}$$

where the double summation excludes pairs of nodes that are not connected. This is a variational formulation of the Helmholtz-Hodge decomposition. Subtracting the gradient flow thus determined from the original flow leaves the loop flow. In addition, to remove arbitrariness in the potential determination, we impose the following condition on  $\phi_i$ :

$$\sum_i \phi_i = 0. \tag{7}$$

To quantify to what extent the flow of a directed network has hierarchical and circular characteristics, we introduce two measures for the gradient flow and circular flow components associated with each node  $i$  as follows:

$$\xi_i^{(p)} = \frac{\frac{1}{2} \sum'_j w_{ij}^{-1} \left( F_{ij}^{(p)} \right)^2}{\sum'_{j < k} w_{jk}^{-1} \left( F_{jk} \right)^2}, \tag{8}$$

$$\xi_i^{(c)} = \frac{\frac{1}{2} \sum'_j w_{ij}^{-1} \left( F_{ij}^{(c)} \right)^2}{\sum'_{j < k} w_{jk}^{-1} \left( F_{jk} \right)^2}. \tag{9}$$

Summation of  $\xi_i^{(p)}$  and  $\xi_i^{(c)}$  over all nodes yields generalized formulas for the gradient and loop ratios (Fujiki and Haruna 2014; Haruna and Fujiki 2016), respectively:

$$\gamma = \sum_i \xi_i^{(p)}, \tag{10}$$

$$\lambda = \sum_i \xi_i^{(c)}. \tag{11}$$

Because of the orthogonality between the gradient flow and the circular flow vectors, the sum of the two ratios amounts to unity:

$$\gamma + \lambda = 1 \tag{12}$$

If a network is completely hierarchical (circular),  $\gamma = 1$  (0) and  $\lambda = 0$  (1). One can thus use either of the two ratios to characterize the overall flow structure of a directed network. This can be considered as ranking the nodes according to the hierarchical structure of the network.

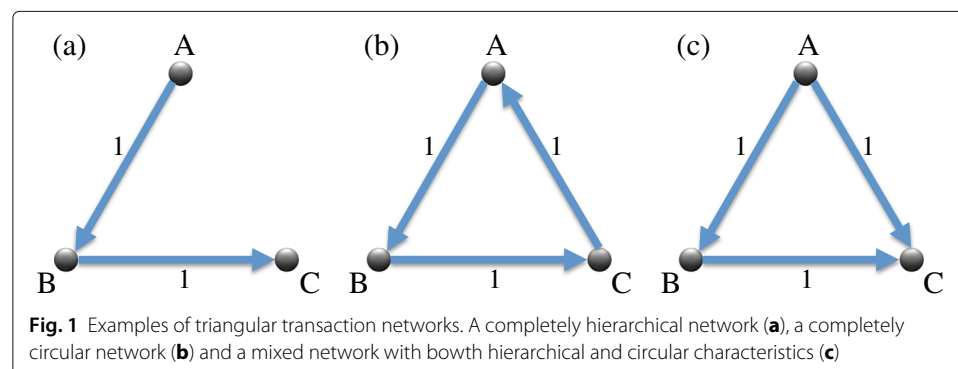
There are a number of works on hierarchy and ranking in complex directed networks (De Bacco et al. 2018; Tatti 2015; Letizia et al. 2018; Johnson et al. 2014). In such problems, the presence of cyclic parts of the network causes ranking conflicts. Studies have been conducted in the context of how to minimize the inconsistency encountered when considering ranking in actual networks. Previous work (Tatti 2015; Letizia et al. 2018)

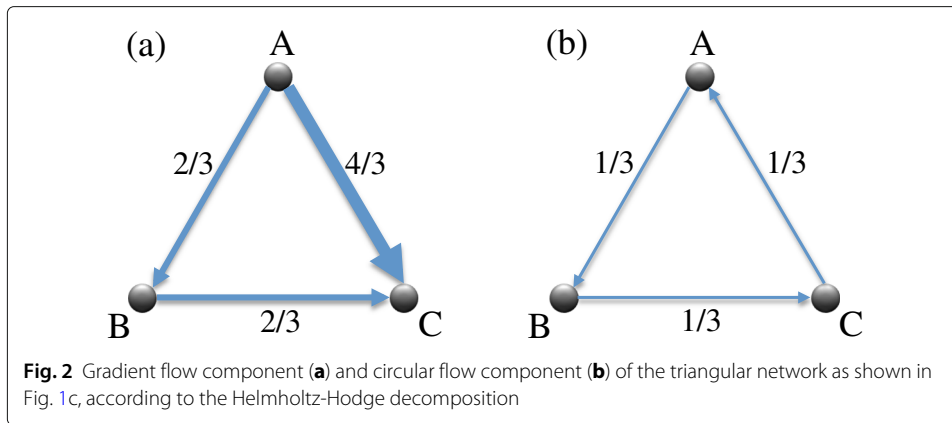
accomplishes this by minimizing the penalty function called agony, and De Bacco et al. (2018) by minimizing the energy of the physical model. Determining  $\phi$  to minimize Eq. (6) corresponds to the Helmholtz-Hodge decomposition. The expression of energy minimization of the physical model is one of the variants of the Helmholtz-Hodge decomposition. Generally, the method of optimizing the penalty function is computationally expensive, but the computational cost of the Helmholtz-Hodge decomposition and physical model is not so cumbersome because it only needs to solve a set of linear equations. The major difference between Helmholtz-Hodge decomposition and other methods is the Helmholtz-Hodge decomposition allows us to treat hierarchies and cycles on equal footing. The Helmholtz-Hodge decomposition has a strong advantage of providing a unified representation of the flow structure of a directed network not only in terms of hierarchy but also in terms of circularity. In this paper, we focus on circularity as well as hierarchy, taking advantage of the Helmholtz-Hodge decomposition.

Finally, we illustrate the Helmholtz-Hodge decomposition with examples of triangular transaction networks in Fig. 1. The first example shown in Fig. 1a is a completely hierarchical network with  $\phi_A = 1, \phi_B = 0$ , and  $\phi_C = -1$ , while the second one in Fig. 1b is a completely circular network. The third example in Fig. 1c is a mixed network with both hierarchical and circular characteristics. Its gradient flow component, Fig. 2a, is determined by  $\phi_A = 2/3, \phi_B = 0$ , and  $\phi_C = -2/3$ . The circular flow component is a loop of flow with magnitude  $1/3$ , as shown in Fig. 2b.

### Bow-tie decomposition

To elucidate flow structure in the TSR transaction network, we begin with the bow-tie decomposition of the network, which has been widely used to understand the flow structure of various complex networks including the worldwide web and metabolic networks. The decomposition classifies nodes in a directed network according to the way in which they are mutually connected: IN component, GSCC (giant strongly connected component), OUT component, and others. The GSCC is the largest group of nodes in which any pairs of nodes are connected bidirectionally by two directed paths. The IN component is a collection of nodes that have a path to the GSCC, but no reverse path to come back from the GSCC. The OUT component is defined in the other way around, that is, a collection of nodes that are reachable only from the GSCC. From their definition, these classifications of nodes provide an overall view on the hierarchical structure of the network. In the previous paper (Chakraborty et al. 2018), however, we named such a structure of the TSR network the “walnut” structure instead of the bow-tie structure after its shape. Because





the IN and OUT components are not as separated as the two wings of a bow-tie, they are more similar to two halves of a walnut shell, surrounding the central GSCC core.

Table 1 lists the numbers of firms belonging to the IN, GSCC, OUT and other components of the TSR network. The results are compared with the corresponding numbers of firms averaged over 1000 random networks with the same degree distribution as that of the original network. We observe no significant difference in the bow-tie parameters between the original and randomized networks. However, Table 1 also shows that complete randomization of the network destroys the bow-tie structure; virtually all nodes constitute the GSCC.

In contrast, the distributions of the Helmholtz-Hodge potential shown by the histogram in Fig. 3 exhibit a significant difference between the two deep networks; the flow structure of the network is influenced by the randomization process.

The potential distributions of IN, GSCC, and OUT in the original network are well-overlapped compared with the randomized network with the same degree sequence. In particular, the distributions of IN and OUT of the randomized network are quite separated, but the corresponding distributions of the original network are substantially overlapped. For a quantitative argument, we define the following overlap integral of two distribution functions:

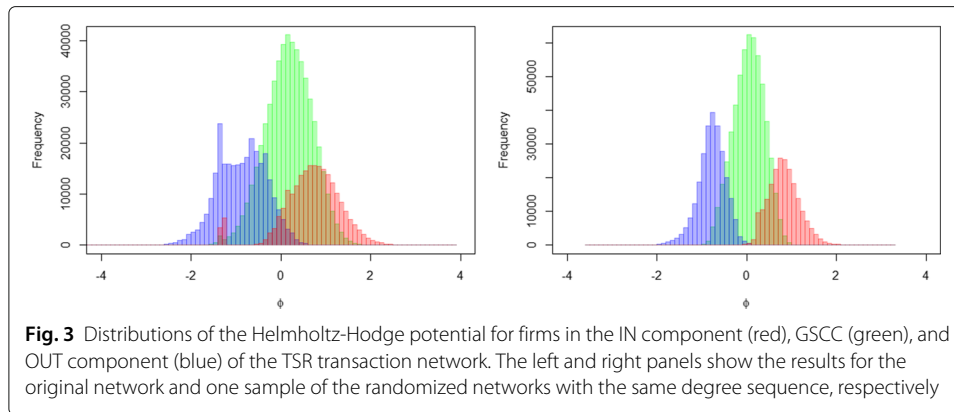
$$J = \frac{\int f(x) \cdot g(x) dx}{\sqrt{\int f^2(x) dx} \sqrt{\int g^2(x) dx}}. \tag{13}$$

The overlap integral  $J$  takes a value within the range of  $0 \leq J \leq 1$ ;  $J$  takes the unity for  $f(x) \propto g(x)$ . In the TSR network, the overlap integral of the potential distributions for the

**Table 1** The numbers of firms belonging to the IN, GSCC, OUT and other components of the TSR transaction network

	TSR	Random TSR	Completely Random
IN	219,927	212,415 (181.9)	10,395 (105.1)
GSCC	530,174	560,876 (212.5)	1,045,150 (154.6)
OUT	278,880	263,412 (201.7)	10,388 (108.1)
Others	37,056	25,701 (162.5)	9 (3.0)
Total	1,066,037	1,062,404 (82.4)	1,065,942 (10.1)

The results are compared with the corresponding numbers of firms averaged over 1000 random networks with the same degree distribution and 1000 completely random networks with the same number of links as that of the original network; the figures in parentheses show the standard deviation associated with each of the average values

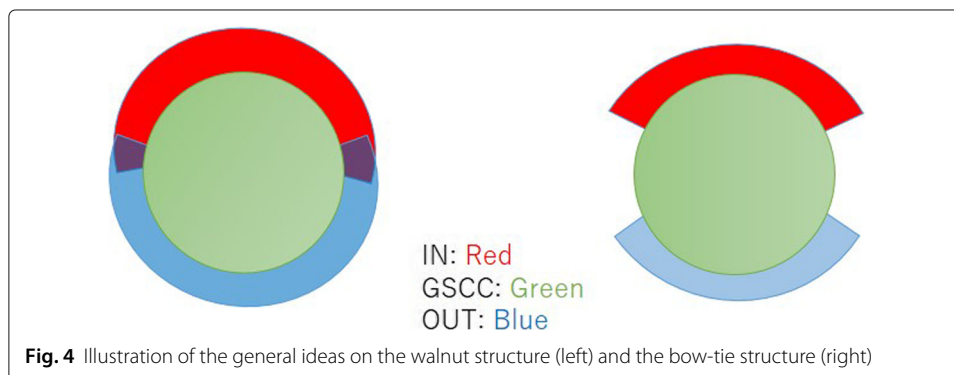


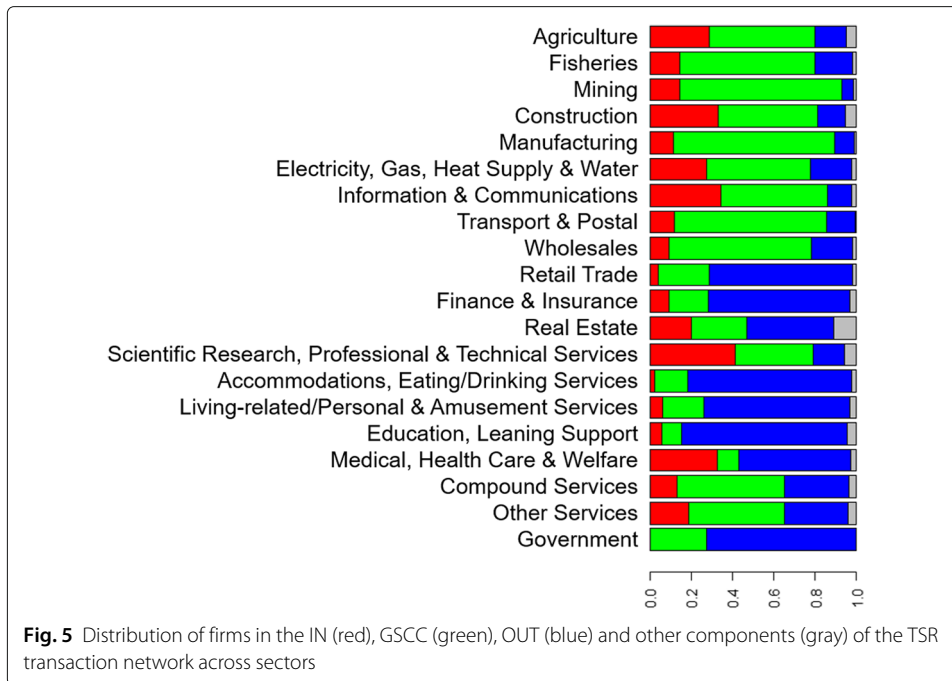
IN and OUT components is  $J^{(\text{TSR})} = 0.125$ . On the other hand, randomized networks with the same degree sequence take much smaller values of  $J$ , for instance,  $J_{0.01}^{(\text{rand})} = 0.00021$  and  $J_{0.05}^{(\text{rand})} = 0.00020$ , where  $J_{0.01}^{(\text{rand})}$  and  $J_{0.05}^{(\text{rand})}$  are the 1% and 5% significance level of  $J$  for 1,000 samples, respectively. These numerical results establish the conceptual difference between the bow-tie structure and the walnut structure, as shown in Fig. 4, in a statistically meaningful way. We emphasize that the structure is not essentially determined by the degree distribution, but by more detailed properties on the linkage of the network.

Figure 5 shows the distribution of firms in such bow-tie components of the TSR network across sectors. The sectors such as Construction, Information & Communications, and Scientific Research, Professional & Technical Services are important constituents in the IN component. Mining, Manufacturing, Transport & Postal, and Wholesale sectors are key players in the GSCC. The important sectors in the OUT component include Retail Trade, Finance & Insurance, Accommodations, Eating/Drinking Services, Living-related/Personal & Amusement Services, and Education, Learning Support. We thus see that each component of the bow-tie structure in the production network has its own industrial characteristics. The main industries in the GSCC form an integrated core of economic activities in Japan.

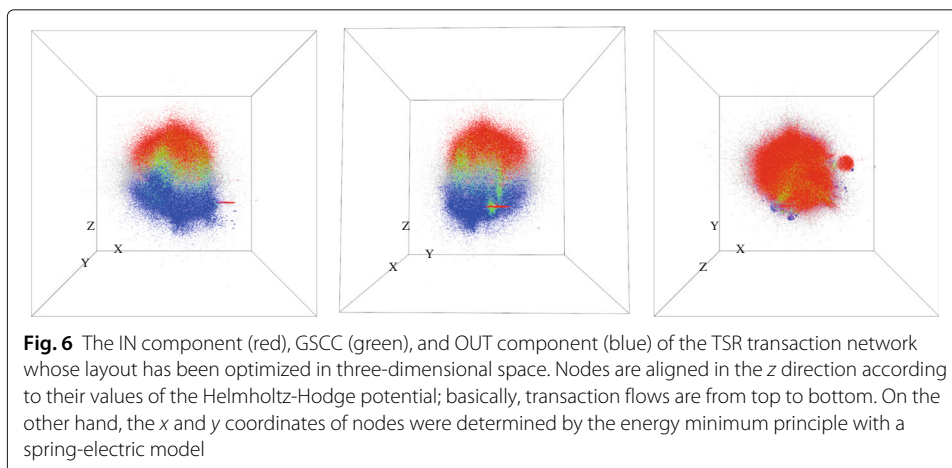
## Results and discussion

We first obtained an optimized layout of the network in three-dimensional space by incorporating information of the Helmholtz-Hodge potential for individual nodes. The result is displayed in Fig. 6. Nodes are aligned in the  $z$  direction according to their values of the

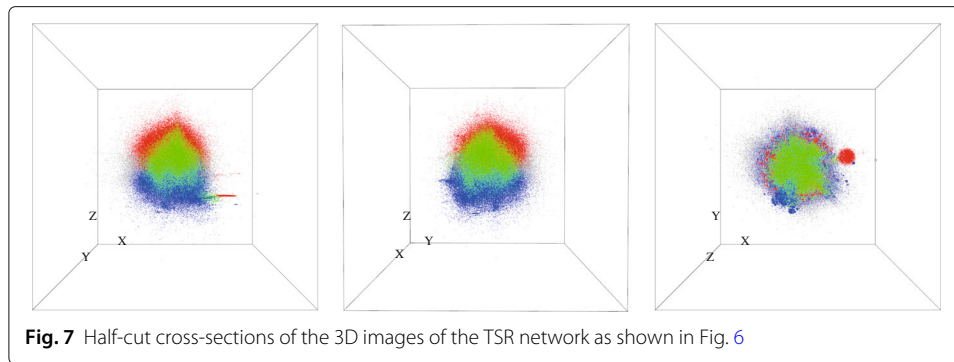




Helmholtz-Hodge potential; basically, transaction flows are from top to bottom. On the other hand, the  $x$  and  $y$  coordinates of nodes are determined by minimizing the potential energy in a spring-electric model in which nodes with direct transaction relations are connected to each other by a spring and all nodes have an identical electric charge to maintain distance from disconnected nodes. In Fig. 6, nodes belonging to the different walnut components are distinguished with different colors. Figure 7 shows half-cut cross-sections of the 3D images of the network, as shown in Fig. 6. The walnut structure is also clearly visible in this visualization. The GSCC is certainly sandwiched between the IN component on the upstream side and the OUT component on the downstream side. However, the potential values in the three components are distributed so widely that even the potential distributions of the peripheral components are not well separated. These







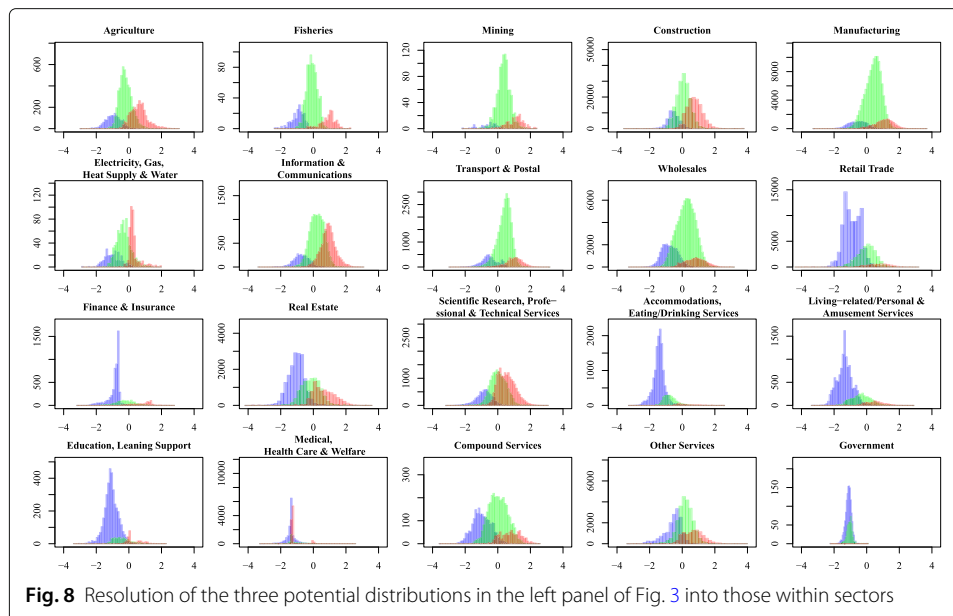
results agree with our naming convention of the flow structure of the transaction network as a walnut structure.

Figure 8 resolves the three potential distributions in Fig. 3 into those within individual sectors. The averaged value of the potential for firms in each sector is given in Table 2. The results, listed in their descending order, describe hierarchical characteristics of the sectors in the transaction network. For instance, the manufacturing sector is located at the upstream side compared with the wholesale and retail trade sectors. The hierarchical ordering of sectors is in harmony with the general idea of the supply chain. However, the potential values are widely distributed from upstream to downstream even within the same sectors, except for Finance & Insurance, Medical, Health Care & Welfare, and Government. This fact indicates that major sectors such as Manufacturing, Construction, Wholesale and Retail trades have appreciable hierarchical structure in and of themselves.

We turn our attention to the gradient ratio  $\gamma$  and loop ratio  $\lambda$  for the whole network and the GSCC of the transaction network.<sup>3</sup> The results are shown in Table 3 together with the corresponding results for the two kinds of random networks in parallel with Table 1. The hierarchy is significantly developed in the original network compared with the randomized networks. This is understandable because hierarchical structure is in general a manifestation of self-organization in complex systems (Holland 2000; Anderson 1972); it is a formation of supply chains in the economic system. Although randomizing the network with a preserved degree distribution does not change the walnut structure considerably, the randomization procedure has an appreciable influence on the balance between the hierarchy and circularity of the network. We have similar results for the GSCC of the original network. The hierarchy is slightly stronger than the circularity even in the GSCC consisting only of nodes that are mutually connected in both ways. In contrast, the circularity dominates the flow structure of the corresponding network that has been completely randomized.

The hierarchical flow is dominant in the IN component, which has mainly one-way flow to the GSCC because of its definition. This is also true for the OUT component. On the other hand, the GSCC has a more complicated flow structure; both hierarchical and circular flow components coexist in it. This is because any pairs of nodes in the GSCC are connected bidirectionally by at least two directed paths. We thus expect that firms in the GSCC constitute the core of the production activities, while firms in the IN and OUT parts, forming a thin layer for the GSCC, are just peripherals.

<sup>3</sup>We reiterated the Helmholtz-Hodge decomposition of the GSCC for its gradient and loop ratios.



For the purpose of this study, therefore, we hereafter concentrate on the flow structure of the GSCC, especially its circularity. To identify important loops in the circular flow network on the GSCC, we adopt the map equation method (Rosvall and Bergstrom 2008; 2011) for community detection. It is an information-theoretic method based on an idea that random walkers should stay in looping communities for a long

**Table 2** The averaged values of the Helmholtz-Hodge potential for firms in individual sectors, which are listed in their descending order corresponding to the direction of upstream to downstream in the TSR transaction network

Industry	Mean of the potential
Mining	0.412
Manufacturing	0.388
Transport & Postal	0.353
Information & Communications	0.348
Construction	0.216
Scientific Research, Professional & Technical Services	0.103
Wholesale	0.084
Other Services	0.041
Agriculture	-0.081
Fisheries	-0.100
Compound Services	-0.188
Electricity, Gas, Heat Supply & Water	-0.274
Real Estate	-0.380
Finance & Insurance	-0.562
Retail Trade	-0.602
Education, Learning Support	-0.910
Living-related/Personal & Amusement Services	-0.998
Government	-1.100
Medical, Health Care & Welfare	-1.247
Accommodations, Eating/Drinking Services	-1.253

The average of the potential for all firms is set to zero

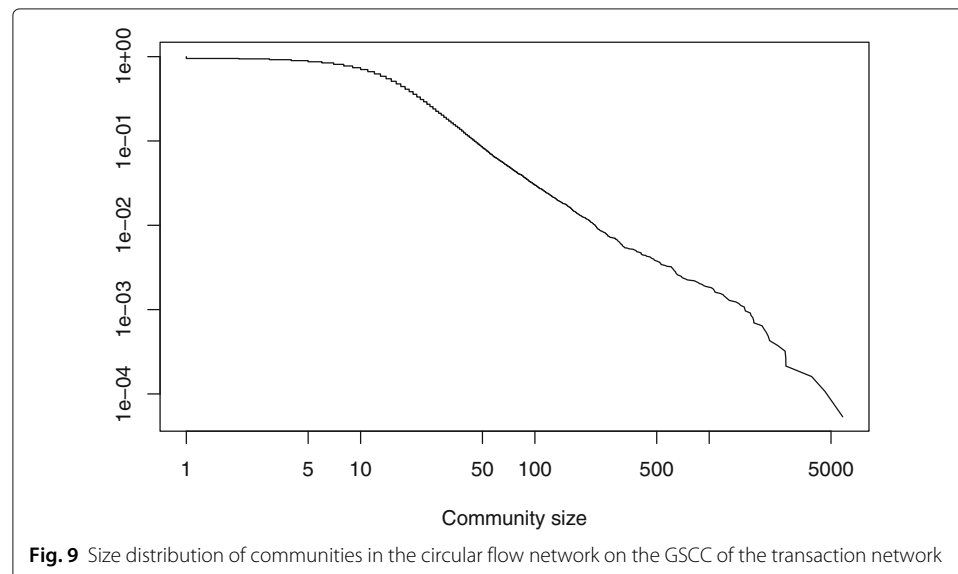
**Table 3** Gradient ratio  $\gamma$  and loop ratio  $\lambda$  for the whole network and the GSCC of the TSR transaction network

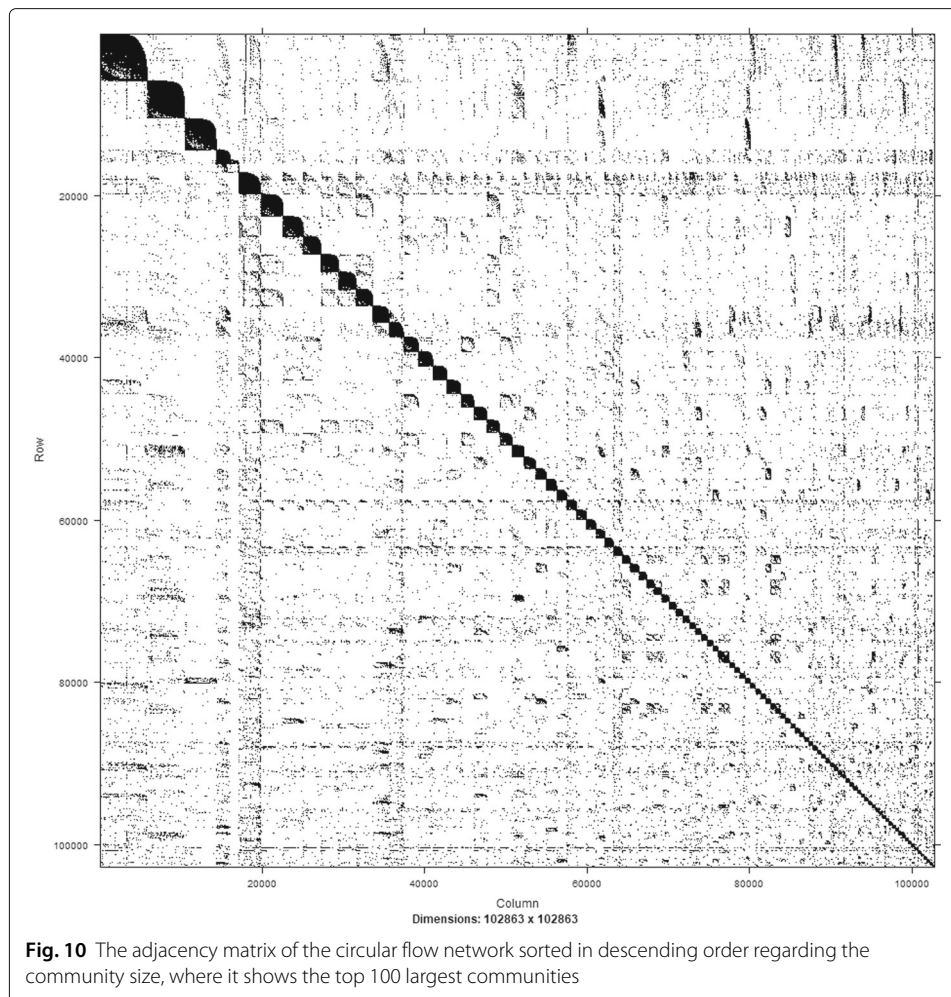
	$\gamma$	$\lambda$
Whole network		
TSR	0.648	0.352
Random TSR	0.537 (0.0001)	0.463 (0.0001)
Completely Random	0.214 (0.0003)	0.786 (0.0003)
GSCC		
TSR	0.537	0.463
Random TSR	0.386 (0.0001)	0.614 (0.0001)
Completely Random	0.200 (0.0002)	0.800 (0.0002)

The results are compared with the corresponding ratios averaged over 1000 random networks with the same degree distribution and 1000 completely random networks with the same number of links as that of the original network; the figures in parentheses show the standard deviation associated with each of the average values

time. Figure 9 demonstrates that the communities so detected have a size distribution of the long-tail form. The total number of communities is 18,660, and the largest community has approximately 5000 firms. Figure 10 depicts the adjacency matrix of the circular flow network in which nodes are ordered according to the community assignment. It shows the community detection works well because links are sparse between the communities and are considerably dense within the communities. The 10 largest communities are illuminated in Fig. 11 with the same node configuration as in Figs. 6 and 7.

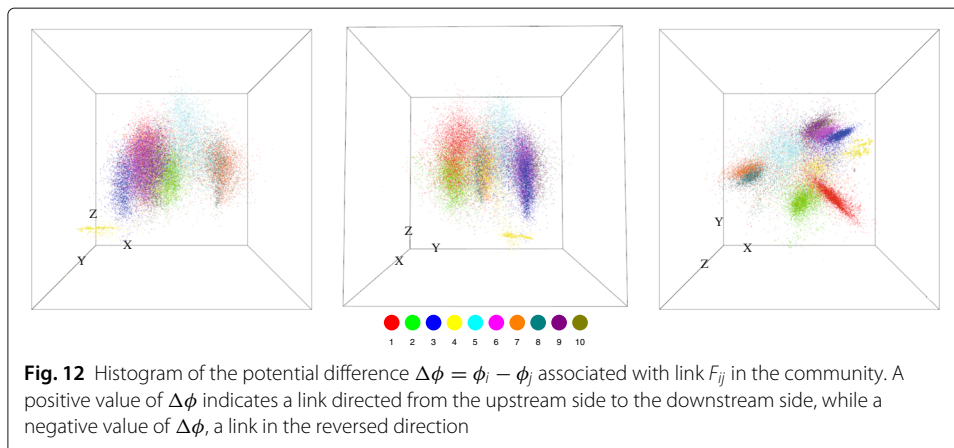
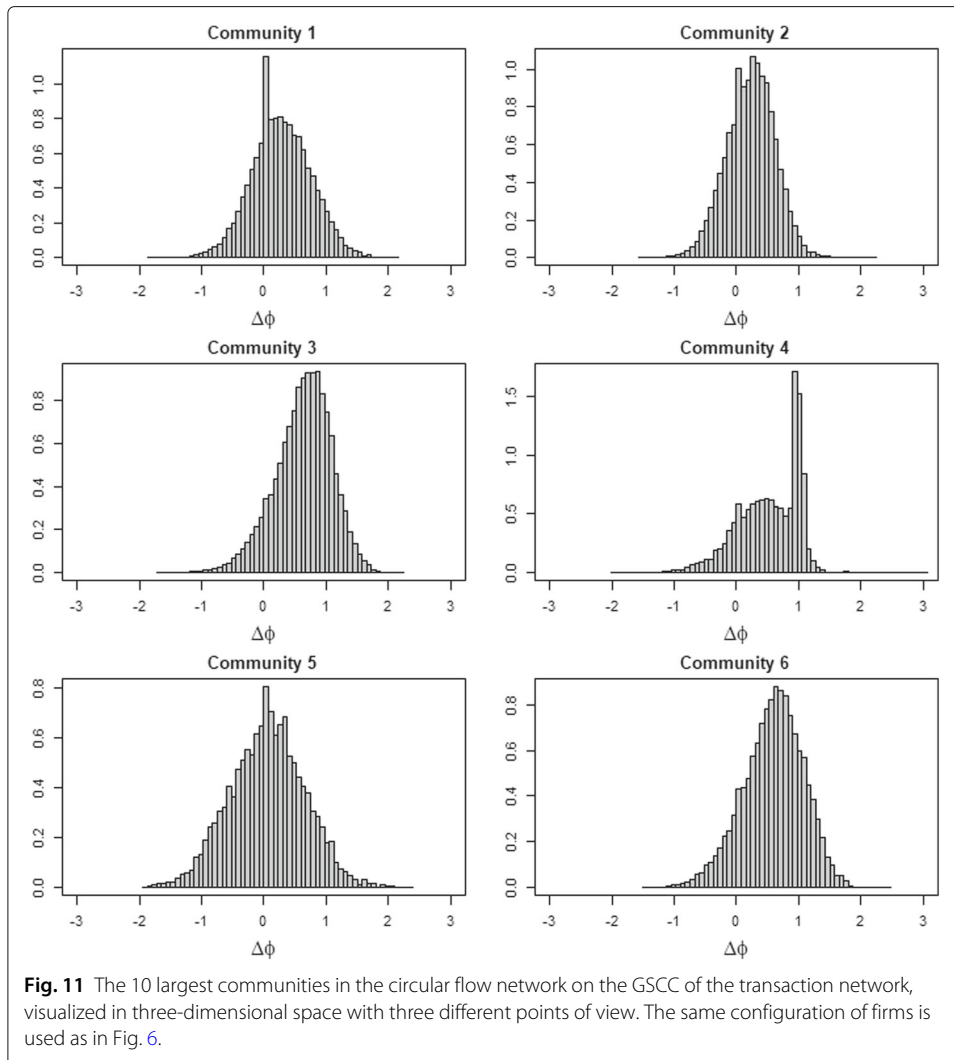
Figure 12 shows the histogram of the Helmholtz-Hodge potential difference  $\Delta\phi$  of links for the 1st-6th communities, where  $\Delta\phi = \phi_i - \phi_j$  is the potential difference between nodes  $i$  and  $j$  at both ends of links  $F_{ij}(> 0)$ . A positive value of  $\Delta\phi$  indicates a link directed from the upstream side to the downstream side, while a negative value of  $\Delta\phi$ , a link in the reversed direction. The distribution of  $\Delta\phi$  is significantly shifted to the positive side except for the 5th community. It means that the main flow from the upstream to





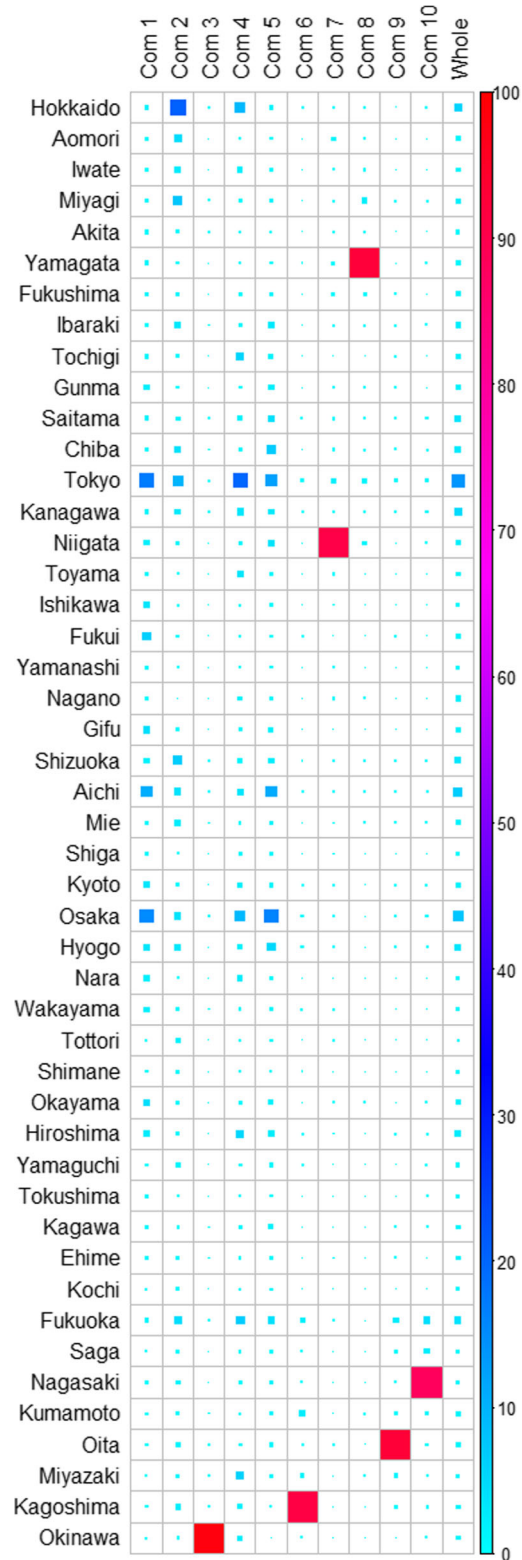
the downstream dominates over the feedback flow in those communities. On the other hand, the 5th community shows the distribution of  $\Delta\phi$  that is quite symmetrical around  $\Delta\phi = 0$ . This indicates that the exceptional community has well-developed circular flow structure, which will be addressed later.

One can characterize the major communities by industrial and regional affiliations of their constituent firms. Figure 13 shows the industrial characterization of the 10 largest communities. They are divided into two contrastive groups. The first, second, fourth and fifth largest communities are mainly featured by manufacturing and wholesale sectors; the medical, health care & welfare sector is additionally important for the fourth community. On the other hand, the remaining 6 communities are featured by the construction sector. Figure 14 shows the regional characterization of the 10 largest communities. Additionally, all of the major communities have prominent regional characteristics. The manufacturing- and wholesale-dominant communities are basically metropolitan communities except for the second largest community, in which Hokkaido and some provincial prefectures play a key role. In contrast, the distribution of the regional affiliations in the construction-dominant communities are well-localized at prefecture level. All prefectures except Nara have at least one of their own construction communities of more than 100 firms.





We will look into the 6 largest communities in more detail. Tables 4 and 5 list the number of firms in each industry type that belong to the 6 largest communities. They are grouped by the middle classification of the TSR industry classification scheme (99 industries). Communities 1, 2, 4, and 5 were groups of firms in which manufacturing and wholesales dominate. Community 1 includes many firms in the manufacture and wholesale of textile and apparel industries. Community 2 is mainly composed of firms in the fisheries cooperative, wholesale and retail trade of seafood, and the manufacture of food from seafood. Community 4 is characterized by medical and health services, and the manufacture and wholesale of pharmaceutical products with representative firms. Most of



**Fig. 14** Regional characterization of the 10 largest communities obtained for the circular flow network on the GSCC by share of prefectures in which their constituent firms are located. The size of each square is proportional to share of the corresponding prefecture in the specified community. Additionally, its share value (in percentage) is represented by the color coding

**Table 4** Number of firms for each industry type that belongs to the 6 largest communities

Industry type	#firms
Community 1	
Manufacture of textile mill products	2547
Wholesale trade (textile and apparel)	1222
Miscellaneous living-related and personal services	226
Retail trade (dry goods, apparel and apparel accessories)	184
Miscellaneous wholesale trade	171
Wholesale trade (building materials, minerals and metals, etc.)	93
Professional services	81
Miscellaneous of manufacturing industries	78
Construction work, general work including public and private construction work	74
Miscellaneous business services	73
Community 2	
Wholesale trade (food and beverages)	1850
Manufacture of food	1383
Retail trade (food and beverage)	151
Fisheries	150
Cooperative association	124
Miscellaneous wholesale trade	120
Road freight transport	99
Aquaculture	60
Miscellaneous retail trade	57
Warehousing	41
Community 3	
Construction work, general work including public and private construction work	998
Construction work by specialist contractor, except equipment installation work	802
Equipment installation work	672
Wholesale trade (building materials, minerals and metals, etc.)	195
Wholesale trade (machinery and equipment)	165
Miscellaneous retail trade	83
Technical services	82
Miscellaneous wholesale trade	74
Machinery and equipment	72
Road freight transport	69

Industry types are according to the middle classification of the TSR industry type classification

the medical and health services are general hospitals and clinics. Community 5 contains many firms in the manufacture and wholesale of metal products and construction. On the other hand, most of the firms in communities 3 and 6 are those in the construction industry. Although the industrial distributions of the construction communities resemble each other closely, they are clearly distinguished by their regional characteristics. In fact, communities 3 and 6 are dominated by firms in Okinawa and Kagoshima, respectively. We note that Okinawa is considerably isolated from the mainland in the whole industry.

In this way, one can characterize the manufacture and wholesale communities by their own products. Although the communities thus extracted represent dense parts of the loop flow network, they mainly include a series of firms in the production line from manufacturing to retail trade through wholesales, the so-called supply chain, themselves forming a hierarchical flow structure.



**Table 5** Continuation of Table 4

Industry type	#firms
Community 4	
Medical and other health services	956
Manufacture of chemical and allied products	349
Miscellaneous wholesale trade	287
Miscellaneous retail trade	141
Wholesale trade (machinery and equipment)	133
Wholesale trade (food and beverages)	69
Wholesale trade (building materials, minerals and metals, etc.)	67
Miscellaneous business services	55
Manufacture of food	49
Professional services	42
Community 5	
Wholesale trade (building materials, minerals and metals, etc.)	529
Construction work by specialist contractor, except equipment installation work	404
Manufacture of fabricated metal products	305
Construction work, general work including public and private construction work	218
Manufacture of iron and steel	212
Road freight transport	165
Manufacture of production machinery	120
Equipment installation work	89
Manufacture of general-purpose machinery	79
Wholesale trade (machinery and equipment)	62
Community 6	
Construction work, general work including public and private construction work	993
Construction work by specialist contractor, except equipment installation work	529
Equipment installation work	411
Wholesale trade (building materials, minerals and metals, etc.)	126
Miscellaneous retail trade	90
Manufacture of ceramic, stone and clay products	62
Wholesale trade (machinery and equipment)	53
Road freight transport	48
Agriculture	31
Technical services	31

We are now in a position to identify firms that strongly contribute to the feedback structure in the major communities. For this purpose, we use the ratio  $\xi_i^{(c)}/\xi_i^{(p)}$  of the circular and gradient flow components for each node  $i$ . A firm characterized by a large value of this ratio proves to be important for the circular flow structure of the communities. Details of this method are given with an illustrative example in the Appendix A. Tables 6 and 7 list the number  $n_c$  of firms for each industry in the community, the number  $n'_c$  of firms with the top 10% of the  $\xi_i^{(c)}/\xi_i^{(p)}$  value within community and the ratio  $n'_c/n_c$  of them. In community 1, the  $n_c$  of the manufacture and wholesale of textile and apparel are large, but their  $n'_c$  values are low, so that the corresponding  $n'_c/n_c$  values are lower than those of other industries. This indicates that firms of the manufacture and wholesale of textile and apparel have strong hierarchy. On the other hand, the  $n'_c/n_c$  value of the industry that is not related to textiles or apparel is high. Therefore, this fact shows

**Table 6** Table of the number of firms  $n_c$  for each industry type in the community, and the number of firms  $n'_c$  that has the top 10% of the  $\xi^{(C)}/\xi^{(P)}$  value within the community

Industry type	$n'_c$	$n_c$	$n'_c/n_c$
Community 1			
Manufacture of textile mill products	144	2547	0.057
Wholesale trade (textile and apparel)	71	1222	0.058
Miscellaneous wholesale trade	37	171	0.216
Miscellaneous retail trade	21	58	0.362
Equipment installation work	19	59	0.322
Wholesale trade (building materials, minerals and metals, etc.)	19	93	0.204
Construction work by specialist contractor, except equipment installation work	18	71	0.254
Miscellaneous of manufacturing industries	18	78	0.231
Road freight transport	17	59	0.288
Wholesale trade (machinery and equipment)	17	71	0.239
Community 2			
Wholesale trade (food and beverages)	154	1850	0.083
Manufacture of food	89	1383	0.064
Miscellaneous wholesale trade	20	120	0.167
Retail trade (food and beverage)	20	151	0.132
Road freight transport	14	99	0.141
Miscellaneous retail trade	13	57	0.228
Fisheries	12	150	0.08
Cooperative association	12	124	0.097
Aquaculture	8	60	0.133
Construction work, general work including public and private construction work	7	40	0.175
Community 3			
Construction work by specialist contractor, except equipment installation work	46	802	0.057
Construction work, general work including public and private construction work	40	998	0.04
Equipment installation work	35	672	0.052
Wholesale trade (machinery and equipment)	33	165	0.2
Road freight transport	31	69	0.449
Machinery and equipment	18	72	0.25
Wholesale trade (building materials, minerals and metals, etc.)	15	195	0.077
Miscellaneous wholesale trade	12	74	0.162
Manufacture of ceramic, stone and clay products	10	65	0.154
Information services	10	29	0.345

The top-10 industry types of  $n'_c$  are listed in decreasing order

that a relatively small number of firms that are not related to the industry that characterizes this community are in an important position for the circular structure. In community 2, the manufacture of food and wholesale trade (food and beverages) exhibit high hierarchy, and the miscellaneous retail trade and construction work, which is general work including public and private construction work, have high circularity. In communities 3 and 6, firms in the construction industry form the main stream flow. On the other hand, firms in the manufacture of ceramic, stone and clay products (community 3) or fabricated metal products (community 6), miscellaneous wholesale trade and road freight transport have large contributions to its feedback structure. Community 4 is a community in which medical and health services dominate, but the number of firms of medical and health services is only 7 out of 956 firms with the top 10% of the  $\xi_i^{(C)}/\xi_i^{(P)}$  value. Additionally,

**Table 7** Continuation of Table 6

Industry type	$n'_c$	$n_c$	$n'_c/n_c$
Community 4			
Miscellaneous wholesale trade	19	287	0.066
Manufacture of chemical and allied products	18	349	0.052
Miscellaneous retail trade	16	141	0.113
Wholesale trade (machinery and equipment)	14	133	0.105
Wholesale trade (food and beverages)	13	69	0.188
Miscellaneous business services	13	55	0.236
Professional services	10	42	0.238
Advertising	10	28	0.357
Technical services	10	23	0.435
Agriculture	9	25	0.36
Community 5			
Construction work by specialist contractor, except equipment installation work	43	404	0.106
Manufacture of fabricated metal products	33	305	0.108
Road freight transport	25	165	0.152
Wholesale trade (building materials, minerals and metals, etc.)	23	529	0.043
Construction work, general work including public and private construction work	21	218	0.096
Manufacture of production machinery	18	120	0.15
Equipment installation work	17	89	0.191
Manufacture of general-purpose machinery	9	79	0.114
Miscellaneous wholesale trade	8	44	0.182
Miscellaneous business services	8	25	0.32
Community 6			
Construction work, general work including public and private construction work	44	993	0.044
Construction work by specialist contractor, except equipment installation work	35	529	0.066
Equipment installation work	26	411	0.063
Wholesale trade (building materials, minerals and metals, etc.)	16	126	0.127
Road freight transport	13	48	0.271
Miscellaneous retail trade	13	90	0.144
Wholesale trade (machinery and equipment)	12	53	0.226
Manufacture of fabricated metal products	11	28	0.393
Miscellaneous wholesale trade	9	24	0.375
Manufacture of lumber and wood products	6	26	0.231

the manufacture and wholesale of pharmaceutical products exhibits high hierarchy, but technical services and advertising indicate high circularity. In community 5, construction work and wholesale trade (building materials, minerals and metals, etc.) show high hierarchy, while miscellaneous wholesale trade, equipment installation work and road freight transport largely contribute to the circularity. In fact, the miscellaneous wholesale trade particularly includes iron and steel primary product wholesale, steel crude product wholesale and iron scrap wholesale trade, indicating that recycling steps of iron and steel are incorporated into the steel industry. This is why the flow structure of community 5 is so highly circular, as has been demonstrated.

As we have seen, the production network is certainly quite hierarchical, but the GSCC contains the circular flow component to an appreciable extent. In general, it is hard to reveal the circular flow structure embedded in the production network because it is

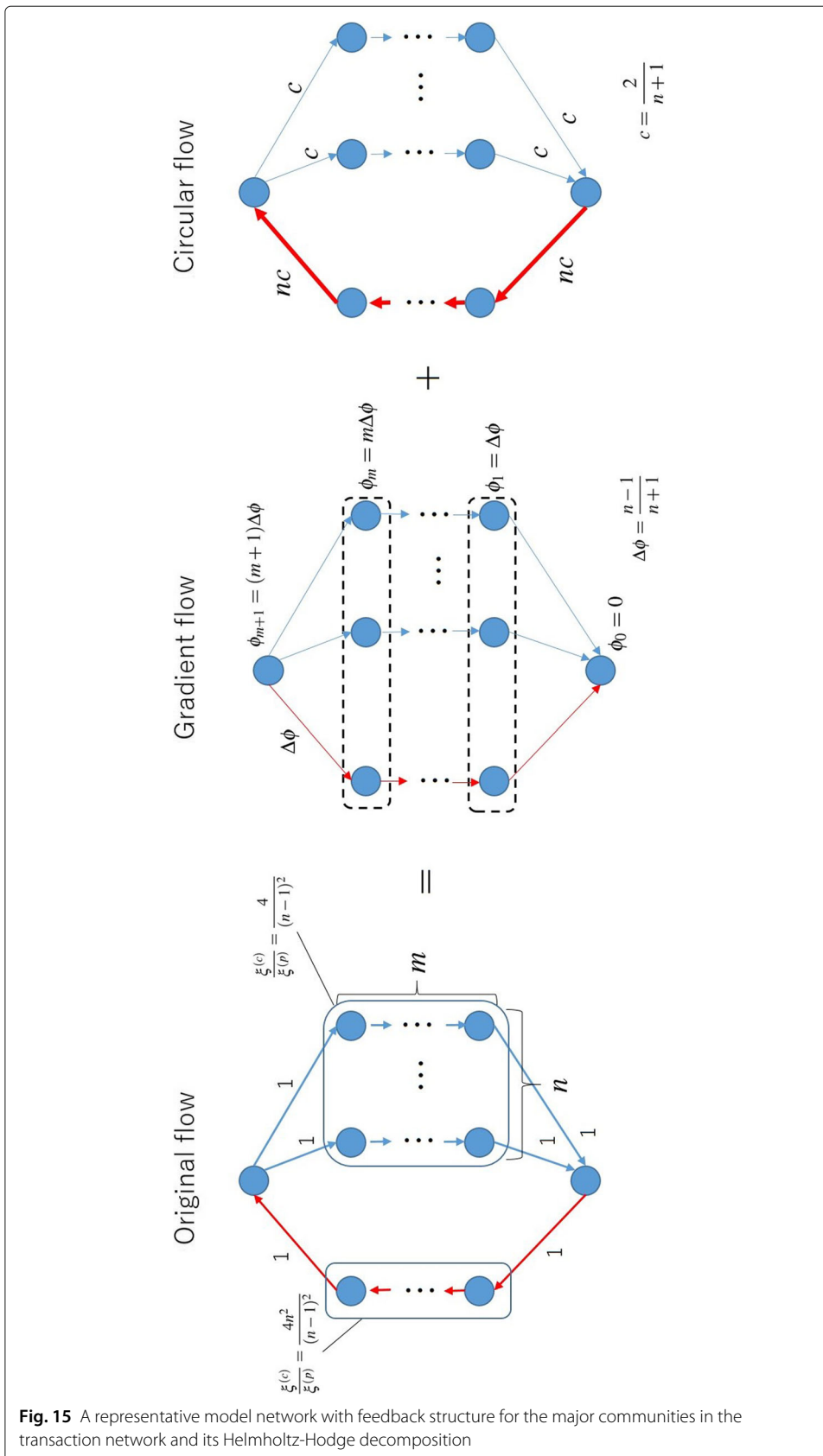
hidden by its strong hierarchy. To overcome the difficulty, we exclusively examined the circular flow component and then applied the community detection to the network thus constructed. Each of the communities that we have detected consist of a hierarchical supply chain of the main industry and feedback loops formed by firms in industries that complement the main industry. Specifically, we found that the transport industry plays an important role in forming the feedback structure for many of the major communities in the production network. In the previous study (Chakraborty et al. 2018), on the other hand, the original flow network was decomposed into communities. Consequently, we were unsuccessful in detecting such industrially integrated clusters of firms as have been reported here.

## Conclusions

The comprehensive dataset of interfirm transaction relations in Japan enabled us to study the industrial flow structure of the nation's production network with a sound microscopic foundation. Particularly, we emphasized its hierarchy and circularity. The network was first decomposed into the walnut components according to their flow properties: IN, GSCC, OUT, and others. The flow structure of the walnut components except the GSCC is mainly hierarchical. By adopting the Helmholtz-Hodge decomposition, we separated the flow structure of the GSCC of the network into two components: gradient flow and circular flow. The gradient flow between a pair of firms is given by the difference of their potentials, and hence, the potential of a firm identifies its hierarchical position in the transaction network. On the other hand, the circular flow component illuminates feedback loops built in the network. The potential values averaged over firms classified by the major industrial category describe hierarchical characteristics of sectors. The order of sectors determined by the potential calculation agrees well with the general idea of the supply chain. We also identified dominant clusters of firms forming feedback loops by applying the map equation method to the extracted circular flow network. We found that both hierarchical and loop structures coexist within the major sectors, such as construction, manufacturing, and wholesales. We measured the magnitude of the contribution to the circular structure from each firm in the major communities. The measurement indicates that the main industry that characterizes the community exhibits high hierarchy and low circularity. On the other hand, most of the firms that contribute to the circular flow structure belong to industries complementary to the main industry, such as the transportation industry. These results suggest limitations of the conventional industrial classification scheme in analyzing economic activities, which may be replaced by a new classification scheme for firms based on the actual interfirm transactions.

## Appendix A: Identification of nodes significantly contributing to the feedback structure.

The major communities in the production network have many links from the upstream side to the downstream side and fewer feedback links from the downstream side to the upstream side. Here, we seek a method to identify nodes that contribute significant feedback from the downstream side to the upstream side in such partial networks. As an illustrative example, we adopt the model network given in Fig. 15, composed of  $m$  layers between the most upstream node and the most downstream node with  $n$  parallel



**Fig. 15** A representative model network with feedback structure for the major communities in the transaction network and its Helmholtz-Hodge decomposition

flows passing through them and one reversed flow forming the feedback structure. This network can be decomposed into the gradient flow and the circular flow components by the Helmholtz-Hodge decomposition, as illustrated in Fig. 15. For this gradient flow component, the potential difference  $\Delta\phi$  between the adjacent layers takes the constant value of  $(n-1)/(n+1)$  throughout the intermediate layers, and the potential difference between the most upstream and the most downstream is thereby  $(m+1)\Delta\phi$ . All of the gradient flows are thus directed from top to bottom with flux of  $\Delta\phi$ . In the circular flow component, each of the  $n$  parallel flows is also directed downward with flux of  $c = 2/(n+1)$ . These parallel flows join at the bottom node and go back to the top node to form a feedback loop with flux of  $nc = 2n/(n+1)$ . On the feedback path, the direction of the loop flow is the same as that of the original flow, while that of the gradient flow is reversed from the original direction ( $F_{ij}$  and  $F_{ij}^{(c)}$  have the same sign, while  $F_{ij}^{(p)}$  has the opposite sign). Quantitatively, the circular flow component plays a relatively important role for a node on the feedback path compared with that for a node on the main stream lines. This is as guaranteed by the following calculations.

Thus, the relative magnitude of  $\xi_i^{(c)}$  in reference to  $\xi_i^{(p)}$  for every node  $i$  is a good measure to identify the nodes that are important in the feedback structure.

For a node  $i$  on the feedback path,  $\xi_i^{(p)}$  and  $\xi_i^{(c)}$  are calculated as follows:

$$\xi_i^{(p)} = \frac{|\Delta\phi|^2}{\sum_{i<j} |F_{ij}|^2} \propto \left(\frac{n-1}{n+1}\right)^2, \tag{14}$$

$$\xi_i^{(c)} = \frac{(nc)^2}{\sum_{i<j} |F_{ij}|^2} \propto \frac{4n^2}{(n+1)^2}, \tag{15}$$

hence:

$$\frac{\xi_i^{(c)}}{\xi_i^{(p)}} = \frac{4n^2}{(n-1)^2}. \tag{16}$$

For a node  $j$  on the main stream lines,  $\xi_j^{(c)}/\xi_j^{(p)}$  is likewise given by the following:

$$\frac{\xi_j^{(c)}}{\xi_j^{(p)}} = \frac{4}{(n-1)^2}. \tag{17}$$

If  $n \gg 1$ , then Eq. (16) takes a much larger value than Eq. (17).

**Abbreviations**

GSCC: Giant Strongly Connected Component; TSR: Tokyo Shoko Research

**Acknowledgments**

This study has been conducted as a part of the project “Large-scale Simulation and Analysis of Economic Network for Macro Prudential Policy” undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This research used computational resources of the K computer provided by the RIKEN Center for Computational Science through the HPCI System Research project (Project ID: hp170242, hp180177).

**Authors’ contributions**

The paper was written collaboratively. The coauthors read and approved the final manuscript.

**Funding**

This research was also supported by MEXT as Exploratory Challenges on Post-K computer (Studies of Multilevel Spatiotemporal Simulation of Socioeconomic Phenomena) and JSPS KAKENHI Grant Numbers JP15KT0052, JP17KT0034, and JP18K03451.

**Availability of data and materials**

The data that support the findings of this study are available from Tokyo Shoko Research, Ltd., but restrictions apply to the availability of these data, which were used under license for the current study, and therefore are not publicly available. However, data are available from the authors upon reasonable request and with permission of Tokyo Shoko Research, Ltd.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Faculty of Science, Niigata University, Niigata, 950-2181, Japan. <sup>2</sup>Graduate School of Simulation Studies, University of Hyogo, Kobe, 650-0047, Japan. <sup>3</sup>The Canon Institute for Global Studies, Tokyo, 100-6511, Japan.

Received: 1 March 2019 Accepted: 30 August 2019

Published online: 26 October 2019

### References

- Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A (2012) The network origins of aggregate fluctuations. *Econometrica* 80(5):1977–2016
- Anderson PW (1972) More is different. *Science* 177(4047):393–396
- Atalay E, Hortacsu A, Roberts J, Syverson C (2011) Network structure of production. *Proc Natl Acad Sci* 108(13):5199–5202
- Bhatia H, Norgard G, Pascucci V, Bremer P-T (2013) The Helmholtz-Hodge decomposition—a survey. *IEEE Trans Vis Comput Graph* 19(8):1386–1404
- Cainelli G, Montresor S, Vittucci Marzetti G (2012) Production and financial linkages in inter-firm networks: structural variety, risk-sharing and resilience. *J Evol Econ* 22(4):711–734
- Carvalho VM (2008) *Aggregate Fluctuations and the Network Structure of Intersectoral Trade*. The University of Chicago, Chicago
- Chakraborty A, Kichikawa Y, Iino T, Iyetomi H, Inoue H, Fujiwara Y, Aoyama H (2018) Hierarchical communities in the walnut structure of the Japanese production network. *PLoS ONE* 13(8):0202739
- Contreras MGA, Fagiolo G (2014) Propagation of economic shocks in input-output networks: A cross-country analysis. *Phys Rev E* 90(6):062812
- De Bacco C, Larremore DB, Moore C (2018) A physical model for efficient ranking in networks. *Sci Adv* 4(7):8260
- Fujiki Y, Haruna T (2014) Hodge decomposition of information flow on complex networks. In: *Proceedings of the 8th International Conference on Bioinspired Information and Communications Technologies*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Gent. pp 103–112
- Goto H, Takayasu H, Takayasu M (2017) Estimating risk propagation between interacting firms on inter-firm complex network. *PLoS ONE* 12(10):0185712
- Haruna T, Fujiki Y (2016) Hodge decomposition of information flow on small-world networks. *Frontiers Neural Circ* 10:77
- Holland JH (2000) *Emergence: From Chaos to Order*. OUP Oxford, Oxford
- Jiang X, Lim L-H, Yao Y, Ye Y (2011) Statistical ranking and combinatorial Hodge theory. *Math Program* 127(1):203–244
- Johnson S, Domínguez-García V, Donetti L, Muñoz MA (2014) Trophic coherence determines food-web stability. *Proc Natl Acad Sci* 111(50):17923–17928
- Letizia E, Lillo F (2018) *Corporate payments networks and credit risk rating*. Available at SSRN 3075019
- Leontief W (1986) *Input-output Economics*. Oxford University Press, Oxford
- Letizia E, Barucca P, Lillo F (2018) Resolution of ranking hierarchies in directed networks. *PLoS ONE* 13(2):0191604
- Luo J, Baldwin CY, Whitney DE, Magee CL (2012) The architecture of transaction networks: a comparative analysis of hierarchy in two sectors. *Ind Corp Chang* 21(6):1307–1335
- McNerney J, Fath BD, Silverberg G (2013) Network structure of inter-industry flows. *Phys A Stat Mech Appl* 392(24):6427–6441
- 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
- Slater P (1977) The determination of groups of functionally integrated industries in the United States using a 1967 interindustry flow table. *Empir Econ* 2(1):1–9
- Slater P (1978) The network structure of the United States input-output table. *Empir Econ* 3(1):49–70
- Tatti N (2015) Hierarchies in directed networks. In: *2015 IEEE International Conference on Data Mining*. IEEE, New York. pp 991–996
- Vestal JE (1995) *Planning for Change: Industrial Policy and Japanese Economic Development 1945-1990*. Clarendon Press, Oxford
- Watanabe T, Uesugi I, Ono A (2015) *The Economics of Interfirm Networks*, vol. 4. Springer, New York

### Publisher's Note

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