RIETI Discussion Paper Series 18-E-026

Hierarchical Communities in the *Walnut* Structure of Japanese Production Networks

Abhijit CHAKRABORTY
University of Hyogo

IYETOMI Hiroshi
Niigata University

INOUE Hiroyasu
University of Hyogo

AOYAMA Hideaki
RIETI

KICHIKAWA Yuichi
Niigata University

IINO Takashi
Niigata University

FUJIWARA Yoshi
University of Hyogo

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 18-E-026

April 2018

Hierarchical Communities in the *Walnut* Structure of Japanese Production Networks*

Abhijit CHAKRABORTY[1], KICHIKAWA Yuichi [2], IYETOMI Hiroshi [2], IINO Takashi [2]

INOUE Hiroyasu [1], FUJIWARA Yoshi [1], and AOYAMA Hideaki [3]

[1]Graduate School of Simulation Studies, The University of Hyogo

[2]Faculty of Science, Niigata University

[3]Graduate School of Science, Kyoto University

Abstract

The structure of Japanese production networks with one million firms and five million supplier-customer links is studied. It is found that they form a tightly-knit structure with a core giant strongly connected component (GSCC) surrounded by IN and OUT components constituting two half-shells for the GSCC, which we name the *Walnut* structure after its shape. The hierarchical structure of communities is studied by the Infomap method and most of the irreducible communities are found to be on the second level. Composition of some of the major communities, including overexpression of industrial and regional nature, as well as connections between the communities, is studied in detail. The findings obtained here cast doubt on the validity and accuracy of the conventional input-output analysis, which is expected to be useful if firms in the same sectors would be well connected with each other.

Keywords: Production network, Hierarchical structure, Community, Infomap, Overexpression, Input-output analysis

JEL classification: D57, D85, L14

*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 was also supported by MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena), Grant-in-Aid for Scientific Research (KAKENHI) by JSPS Grant Numbers 25400393, 17H02041, and the Kyoto University Supporting Program for Interaction-based Initiative Team Studies: SPIRITS, as part of the Program for Promoting the Enhancement of Research Universities, MEXT. We are grateful to Y. Ikeda, W. Souma and H. Yoshikawa for their insightful comments and encouragements, and to the seminar participants at RIETI for their helpful comments and suggestions.

Introduction

Macro economy is the aggregation of the dynamic behaviour of agents who interact with each other under diverse external (non-economic) conditions. Economic agents are numerous; they include consumers, workers, firms, financial institutions, government agencies, countries. Their interactions define economic networks, where nodes are the economic agents and links (edges) connect agents interact with each other. Therefore there are various kinds of economic networks depending on the nature of interactions, which form overlapping multi-level network of networks. Thus any evidence-based scientific investigation of macro economy must be based on understanding of the real nature of the interactions and the economics network of networks they form. Micro level understanding of economic agents also require the same understanding: without knowing whom a firm trades with, how can anyone hope to see into the future of the firm? Therefore, in studying the economic dynamics, either agent-based modelling/simulations or other means of systematic studies such as Debt-Rank [1–5], it is highly important to use the actual network information. Without it, it is difficult to justify the validity of the result to the real world of economy.

In this paper, we study the structure of one of the most important network, the production network, which is formed by firms as nodes and trade relationship as links [6–9]. The data was collected in Japan by TSR (Tokyo Shoko Research Inc.) by means of inquiry to firms as to who are the top five suppliers and the top five customers. Although large firms with many suppliers and customers submit replies that are quite incomplete, they are supplemented from the other side of trade: smaller firms submit replies that include large firms, who are the important trade partners. By combining all the submissions from either side of trade into one database, large firms get connected with numerous smaller firms, which provides a good approximation to the real complete picture. One might worry that since some of the trades last for a short time, even only once as a firm seek a good deal for just one particular occasion, and thus cast doubt on the definition of the trade network. This form of data collection solves this problem: it is most implausible that replies contain one-time trade but rather only the firms above certain trade frequency are likely to be listed. Thus, this production network of the real economic world is of high importance for all aspects of the scientific study of both macro and micro economy.

Before one goes into the agent-model building and simulations, in order to understand dynamics on this network and eventually reaching into the realm of economic fluctuations, business cycles and systemic crisis, and also each firms' growth and decline, one needs to understand its structure. For this purpose, we first describe its overall statistics and visualization in the next Section, and we propose to call its unique overall structure *"Walnut"* structure. This is quite different from what is expected from the existence of the IN-GSCC-OUT components: In the trade network, the flow of materials and goods start from imported/mined/harvested raw materials such as oil, iron and other metals, foods etc. The firms who engaged in this business form IN components. Then they are processed to various parts such as semiconductors or powdered food by firms in GSCC components, before they are made into consumer goods by firms in the OUT components. One might think that this existence of IN-GSCC-OUT components similar to web network implies the bow-tie structure [10]. But the production network is different. Ties among firms form much tighter network form an overall structure that cannot be called bow-tie. Then we study the community structure and reveal its hierarchical nature using the Infomap method [11]. Level-by-level study of communities and "irreducible" communities (communities that are not decomposed into sub-communities at the lower level) are identified. We also study overexpression of some of the major communities are also studied to identify both the industrial sector and regional decomposition. Complex nature of links between the communities are studied. Discussion

and conclusion with future prospects are offered at the end. Some of the supporting materials are included as Appendices.

Production network data and its basic structure

Our data for production network is based on a survey done by Tokyo Shoko Research (TSR), one of the leading credit research agencies in Tokyo, supplied to us through the Research Institute of Economy, Trade and Industry (RIETI). We utilize the two datasets of 'TSR Kigyo Jouhou' (firm information), which contains basic financial information for more than a million firms, and 'TSR Kigyo Soukan Jouhou' (firm correlation information) with several million links of supplier-customer and ownership links, and a list of bankruptcies. Both of them were compiled on July 2016. (For some of the earlier studies of the production network, see [6–9].)

Let us denote a supplier-customer link as $i \to j$, where firm i is a supplier to another firm j, or equivalently, j is a customer of i. We extracted only the supplier-customer links for pairs of "active" firms to exclude inactive and failed firms by using an indicator flag for them in the basic information. Eliminating self-loops and parallel edges (duplicate links recorded in the data), we have a network of firms as nodes and supplier-customer links as edges. The network has the largest connected component, when viewed as an undirected graph, namely the giant weakly connected component (GWCC) comprising of 1,066,037 nodes (99.3% of all the active firms) and 4,974,802 edges.

In addition to the network, several attributes of each node are available; financial information of firm-size, which is measured as sales, profit, number of employees and their growth, major and minor classification into industrial sectors, details of products, the firm's main banks, principal shareholders, and miscellaneous information including geographical location. For the purpose of our study, let us focus on two attributes of each firm, namely industrial sector and geographical location of head office.

The industrial sectors are categorized hierarchically into 20 divisions, 99 major groups, 529 minor groups and 1,455 (Japan Standard Industrial Classification, November 2007, Revision 12). See Table 7 in the Supporting information for the number of firms in each division of industrial sector. Each firm has industry classification according to the sector it belongs to as primary (also secondary and tertiary, if any) industry. The geographical location is converted into a level of one of 47 prefectures or into one of 9 regions (Hokkaido, Tohoku, Kanto, Tokyo, Chubu, Kansai, Chugoku, Shikoku, and Kyushu). See Table 8 in the Supporting information for the number of firms in each regional area of Japan.

In terms of the flow of goods and services (and money in the reverse direction) 1he firms are classified to three categories; "IN" component, "GSCC" (Giant Strongly Connected Component), and "OUT" component. This structure is called "bow-tie" in the well-known study of the Web [10]. The GWCC can be decomposed into the parts defined as follows:

GWCC Giant weakly connected component: the largest connected component when viewed as an undirected graph. An undirected path exists for an arbitrary pair of firms in the component.

GSCC Giant strongly connected component: the largest connected component when viewed as a directed graph. A directed path exists for an arbitrary pair of firms in the component.

IN The firms from which the GSCC is reached via a directed path.

OUT The firms that are reachable from the GSCC via a directed path.

TE "Tendrils"; the rest of GWCC

It follows from the definitions that

$$\mathrm{GWCC} = \mathrm{GSCC} + \mathrm{IN} + \mathrm{OUT} + \mathrm{TE} \tag{1}$$

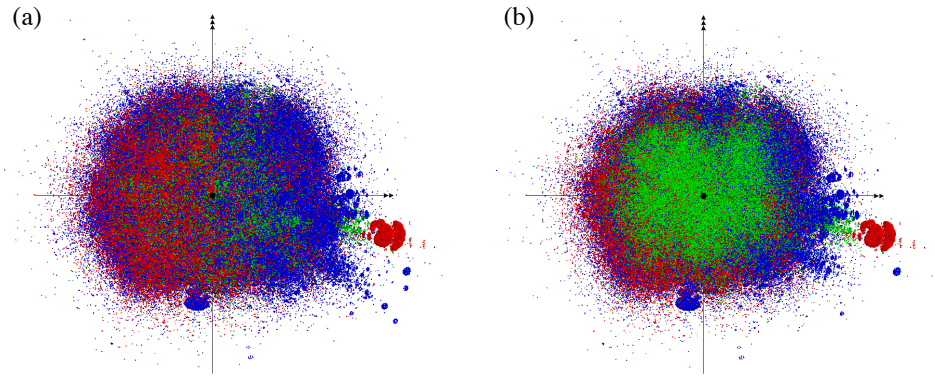


Figure 1: **Visualization of the network in three-dimensional space** A surface view of the network is shown in the panel (a) and a cross-sectional view cut through its center, in the panel (b). The red, green, and dots represent firms in the IN, GSCC, and OUT components, respectively.

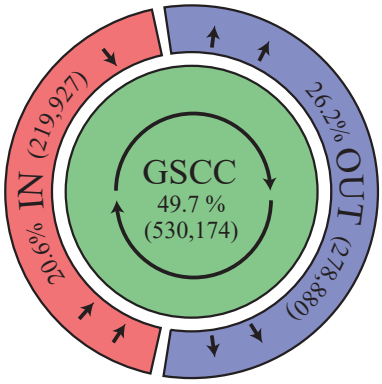


Figure 2: **The Walnut structure.** The production network form a walnut structure. Areas of each components are approximately proportional to their sizes.

We, however, find it far more appropriate to call it ''Walnut" structure, as "IN" and "OUT" components are not as separated as in the two wings of "bow-tie", but is more like the two halves of a walnut shell, surrounding the central GSCC core. Let us see this in the following way. The number of firms in each component of GSCC, IN, OUT and TE is shown in Table 1. Half of the firms is inside the GSCC. 20% of the firms are in the upstream side or IN; 26% of them are in the downstream side or OUT.

To compare with the well-known "bow-tie structure" in the study [10] (in which GSCC is less than one-third of the GWCC), the GSCC in the production network occupies half of the system, meaning that most firms are interconnected in small geodesic

distances or shortest-path lengths in the economy. In fact, by using a standard graph layout algorithm based on a spring-electrostatic model in three-dimensional space [12], we can show in Fig. 1 by visual inspection how closely most firms are interconnected with each other.

Moreover, by examining shortest-path lengths from GSCC to IN and OUT as shown in Table 2, one can observe that those firms in the upstream or downstream sides are located mostly at a single step away from the GSCC. Such a feature in the economic network is different from the bow-tie structure of many other complex networks. For example, hyperlinks between web pages of a similar size (GWCC: 855,802, GSCC: 434,818 (51%), IN: 180,902 (21%), OUT: 165,675 (19%), TE: 74,407 (9%)) studied in [13] have a bow-tie structure such that the maximum distance from GSCC to either IN or OUT is 17, while more than 10% of web pages in IN or OUT are located at more than a single step from GSCC. This observation as well as Fig. 1 leads us to say that the production network has a "walnut" structure, rather than a bow-tie. We depict the schematic diagram in Fig. 2

Later we shall show how each densely connected module or community is located in the walnut structure. For a preview, see Figure 12.

Table 1: **Walnut structure: Sizes of different components**

Component	#firms	Ratio (%)
GSCC	530,174	49.7
IN	219,927	20.6
OUT	278,880	26.2
TE	37,056	3.5
Total	1,066,037	100

"Ratio" refers to the ratio of the number of firms to the total number of the firms in GWCC.

Table 2: **Walnut structure: shortest distances from GSCC to IN/OUT**

IN to GSCC			OUT to GSCC		
Distance	#firms	Ratio (%)	Distance	#firms	Ratio (%)
1	212,958	96.831	1	266,925	95.713
2	6,793	3.089	2	11,650	4.177
3	170	0.077	3	296	0.106
4	6	0.003	4	9	0.003
Total	219,927	100	Total	278,880	100

Left half shows the number of firms in the IN component that connects to GSCC firms with shortest distance 1–4. Left is for the OUT component.

Methods

Community Detection

Community detection is widely used to elucidate structural properties of large-scale networks. In general, real networks are highly non-uniform. Community detection singles out groups of nodes densely connected to each other in a network to divide it into modules. This enables us to have a coarse-grained view on structure of such complicated networks. One of the most popular methods of community division is to maximize the modularity index [14]. Modularity measures the strength of partition of a network into communities by comparing the fraction of links in given communities with the expected fraction of links if links were randomized with the same degree distribution as the original network has. However, it is well known that the modularity method suffers from a problem called resolution limit [15] when applied to large networks. That is, optimizing the modularity would fail to detect small communities even if they are well defined like cliques.

The map equation method [11] is another way to detect communities in a network. This method is found to be one of the best performing community detection technique when compared with others [16]. It is a flow-based and information-theoretic method depending on the map equation defined as

$$L(C) = q_{\curvearrowright} H(C) + \sum_{i=1}^{m} p^i_{\circlearrowright} H(\mathcal{P}^i) \ . \tag{2}$$

Here $L(C)$ measures the per step average description length of dynamics of a random walker migrating through links between nodes of a network with a given node partition $C = \{C_1, \cdots, C_\ell\}$ and consists of two parts. The first term arises from movements of the random walker across communities, where $q_{\curvearrowright}$ is the probability that the random walker switches communities and $H(C)$ is the average description length of the community index codewords given by the Shannon entropy. The second term arises from movements of the random walker within communities, where $p^i_{\circlearrowright}$ is the fraction of the movements within community C_i and $H(\mathcal{P}^i)$ is the entropy of codewords in module codebook i.

If the network has densely connected parts in which a random walker stays long time, one can compress the description length of the random walk dynamics on a network by using a two-level codebook for nodes adapted to such a community structure, an analogy to geographical maps in which different cities recycle the same street names such as main street. Therefore, obtaining the best community decomposition in the map equation framework amounts to searching for the node partition that minimizes the average description length $L(C)$.

As regards the resolution limit problem, any two-level community detection algorithms including the map equation are not able to get rid of the limitation. However, the map equation significantly mitigate the problem as has been shown by a recent theoretical analysis [17]. In practice, this is true for our network, as will be demonstrated later.

Recently, the original map equation method has been extended for networks of multiscale inhomogeneity. A network is decomposed into modules, their submodules, their subsubmodules and so forth. The hierarchical map equation [18] recursively searches for such a multilevel solution by minimizing the description length with possible hierarchical partitions. The map equation framework for community detection of networks is now more powerful. We thereby analyze the production network using this method. The code of the hierarchical map equation algorithm is available at http://www.mapequation.org.

In passing we recall that the community identification for nodes in our network is exclusive in this study. That is, each node belongs to a unique community at every

hierarchical level. However, such community assignment may be too restrictive for a small number of giant conglomerate firms such as Hitachi and Toshiba because of diversity of their business. The map equation is so flexible as to be able to detect overlapping community structure of a network in which any node can be a member of multiple communities [19]. However, we stick to the original algorithm as an initial step toward full account of the firm-to-firm transaction data.

The over-expression within communities and sub communities

Most of the real-world networks exhibit community structure [20]. Such communities are formed in a network based on the principle of homophily [21]. It indicates a node has a tendency to connect with other similar nodes. For example, ethnic and racial segregation is observed in our society [22], biological functions play key role in formation of communities in protein-protein interaction network [23], community structure in a stocks market shows similarity in their economic sector [24]. We find attributes that play crucial role on the formation of community structure in the production network using the following method.

We follow the procedure used in [25] to expose the statistically significant over-expression of different locations and sectors within a community. This method is developed from the statistical validation of over-expression of genes in specific terms of the Gene Ontology database [26]. In this procedure, a hypergeometric distribution $H(X|N, N_C, N_Q)$ is used to measure the probability that X randomly selected nodes from a community C of size N_C will have attribute Q. The hypergeometric distribution $H(X|N, N_C, N_Q)$ can be written as

$$H(X|N, N_C, N_Q) = \frac{\binom{N_C}{X}\binom{N-N_C}{N_Q-X}}{\binom{N}{N_Q}} , \tag{3}$$

where N_Q is the total number of elements with attribute Q in the system. Further, one can associate a *p value* $p(N_{C,Q})$ for $N_{C,Q}$ nodes having an attribute Q in a community C with the $H(X|N, N_C, N_Q)$ by the following relation:

$$p(N_{C,Q}) = 1 - \sum_{X=0}^{N_{C,Q}-1} H(X|N, N_C, N_Q) . \tag{4}$$

The attribute Q is over-expressed within the community C if $p(N_{C,Q})$ is found to be smaller than some threshold value p_c. As we are considering a multiple-hypothesis test, we need to choose p_c appropriately to exclude false positives. We take $p_c = 0.01/N_A$, as used in [25], which suffices Bonferroni correction [27]. Here, N_A represents total number of different attributes (In our study we have $N_A = 9$ for region attributes) over all the nodes of the system.

Results

Hierarchy of communitites

By using the Infomap method [11,18], we have detected hierarchical structure of communities as summarized in Table 3 with the number of firms at each level. This hierarchical structure is illustrated in Fig. 3, where the 2nd level communities are lined up from left to right on the descending order of the community size (number of firms), and the width of triangles reflects the number of subcommunities in that community. From these, we find that most of the subcommunites are on the 2nd level, and that most of the firms (94%) belong to level 2 communities. Compared with level 1 and 2, the 3rd to the 5th levels are of no significant importance. Therefore we limit our discussion of properties of the (sub)communities to the 2nd level, not further down. Past studies on the application of hierarchical map equation in real world networks [17, 18] show that dense networks have large communities at the finest level with shallow hierarchies, and sparse networks tend to have deep hierarchies. It is also observed that the depth of the hierarchies increases with network size. In the case of the California road network, a deep level of hierarchy is found because of the road network has geographical constraints which suppress shortcuts between different parts of the network [18]. In our production network we observe a relatively shallow hierarchy because it does not have such strict constraints.

Table 3: **Modular Level statistics**

Level	#com	#irr.com	#firms	Ratio (%)
1	209	106	830	0.078
2	65,303	60,603	998,267	93.643
3	18,271	17,834	61,748	5.792
4	1,544	1,539	5,168	0.485
5	10	10	24	0.002
Total		80,092	1,066,037	100.00

Results of the community detection based on the multi-coding Infomap method. "#com" is the number of all the communities, "#irr.com" is the number of irreducible communities, which are communities that do not have any subcommunities. "#firms" refers to the number of firms in the irreducible communities

We visualize the hierarchical decomposition of the whole network into communities and their subcommunities in Fig. 4. The configuration of nodes in three-dimensional space is the same as that in Fig. 1. We can see that the network is extremely complex with multi-scale inhomogeneity. The major communities at the 1st and the 2nd levels are characterized through the over-expression analysis as regards industrial sectors and regions in the subsequent subsections.

For the purpose of making the following discussion of communities transparent, let us adopt the following indexing convention: At the top modular level of the hierarchical tree structure, the communities are indexed by the rank of the size (number of firms in the community. Thus, the largest community at the top level is denoted "C_1". At lower levels, the rank of the size in that levels are added after ':'. For example, the community "$C_{1:5}$" is the fifth largest 2nd level community among all the 2nd-level communities that belong to the largest top-level community C_1.

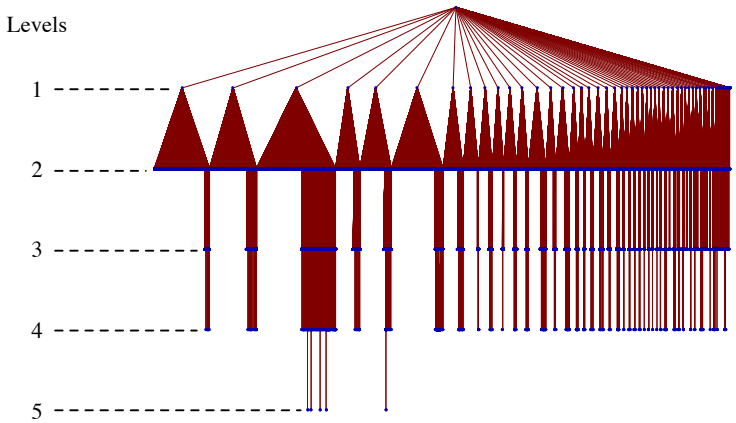


Figure 3: **Hierarchical structure of the communities** Five Levels of hierarchical community decomposition are illustrated. The width of the triangle originating from each community at the n-th level is proportional to the number of its subcomunities at the $(n+1)$-th level.

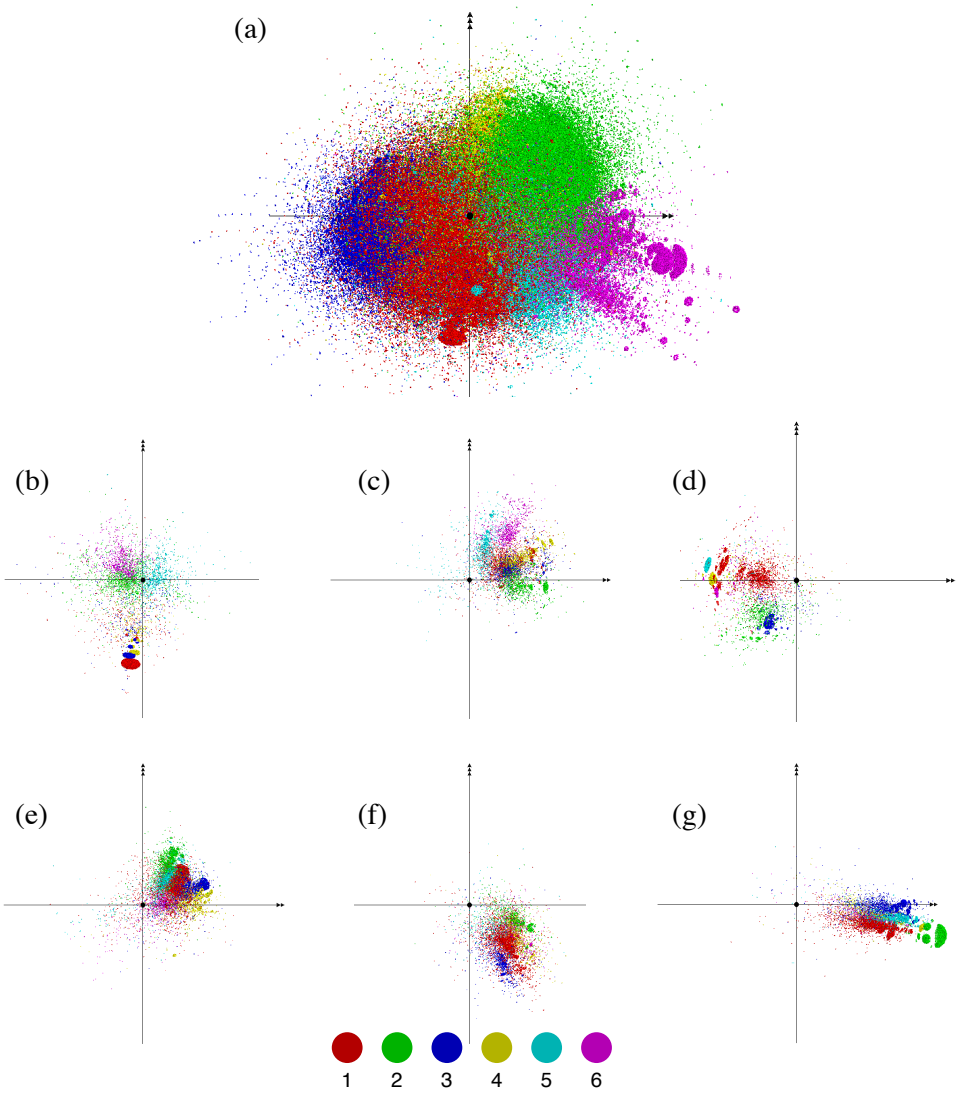


Figure 4: **Hierarchical decomposition of the whole network into communities and subcommunities.** The panel (a) highlights the 8 largest communities at the top modular level with different colors. Each of those communities is further decomposed into subcommunities as demonstrated in the panels, (b) through (g), where the 6 largest subcommunities in the first through 6th largest communities are highlighted.

Level-1 communities

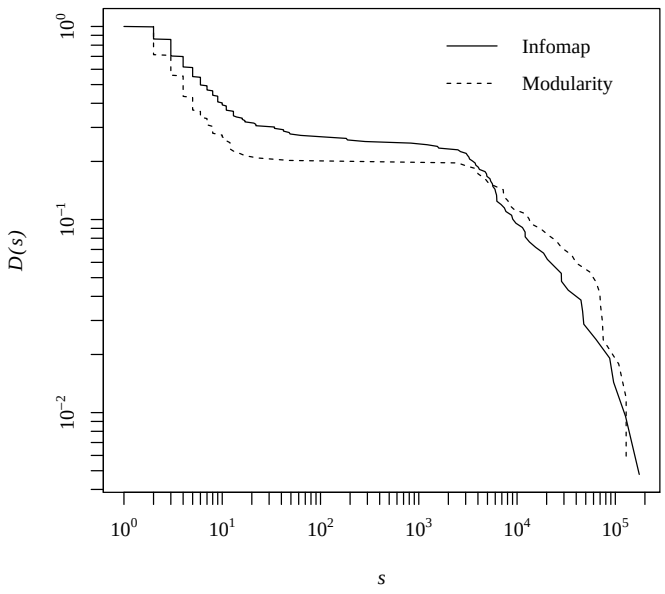


Figure 5: **The complementary cumulative distribution function $D(s)$ of the community sizes s at the top modular level.**

Figure 5 shows the complementary cumulative distribution of the community size at the top level, which is compared with the corresponding result obtained by the modularity maximization method[1] The two distributions are quite similar, indicating that two community structures resemble to each other. More detailed comparison between them is made in the SI 1. The bimodal nature of the distributions manifests the resolution limit problem. A small number of communities predominates the whole system. Among some 200 communities detected, for instance, the largest communities contain 100,000 - 200,000 firms. However, such extremely large communities are decomposed into subcommunities by the hierarchical map equation in a unified way. This is highly contrasted with the community detection based on modularity. One may go around the obstacle by applying the modularity maximization method recursively; communities are regarded as separated subnetworks for further decomposition. However, the procedure lacks a sound basis because it uses different null models to decompose the subnetworks [20].

The map equation is a method to divide a directed network into communities in which nodes are tightly connected in both directions. From its nature, the flows across communities thus detected should be biased in an either direction. Figure 6 confirms the expectation is really true. To quantify the polarizability of links between a pair of

[1]This results is obtained for the undirected network, that is, by ignoring the direction of the links. We have also done the analysis using the directed modularity analysis, but the result is not much different from this.

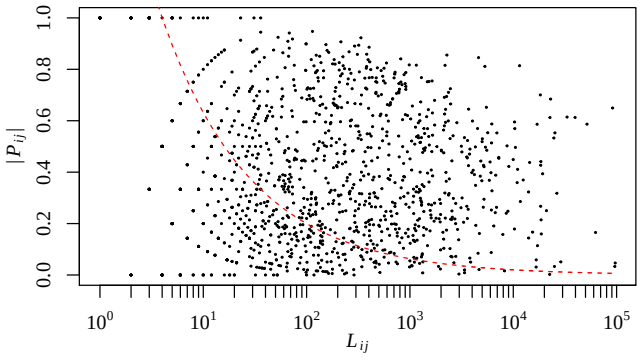


Figure 6: **Polarizability of direction of links interconnecting communities at top level.** Here 51 major communities with more than 1,000 firms are selected. The dashed curve shows a significance level corresponding to 2σ for the polarization ratio of intercommunity links, where random orientation of the links is adopted as a null model; see Eq. (6) for the standard deviation σ.

communities, we introduce the polarization ratio defined by

$$P_{ij} = \frac{A_{ij} - A_{ji}}{A_{ij} + A_{ji}}, \tag{5}$$

where A_{ij} is the total number of links spanning from communities i to j and A_{ji}, that of opposite links. If the linkage between communities i and j is completely polarized, P_{ij} becomes ± 1 depending on its direction; if the linkage is evenly balanced, $P_{ij} = 0$. If we assume that links has no preference with respect to their direction as a null hypothesis, the null model predicts that the polarization ratio for connection between communities i and j fluctuates around 0 with the standard deviation σ given by

$$\sigma = \frac{1}{\sqrt{L_{ij}}} \ , \tag{6}$$

where $L_{ij} = A_{ij} + A_{ji}$ is the total number of links between the two communities. Most of connections between communities with more than 100 links are significantly polarized in reference to the random orientation model for intercommunity links.

We apply the method of finding over-expression of the attributes in level-1 communities to uncover the factors that play crucial role on the formation of such communities. Our study considers the location and sector attributes. The location attributes are divided in 9 regions and the sector attributes are categorized in 20 divisions. The details about the six largest level-1 communities and also over-express attributes within it are tabulated in Table 4. We have further used a finer classification, i.e., 47 prefectures and 99 major sectors for which the results are listed in SI. We observe a strong connection between over-express sectors and over-express regions. In the largest community mainly Manufacturing sectors and heavily urbanized regions (Kanto, Tokyo, Chubu, Kansai) are over-expressed. The second largest community shows the over-expression of mainly agriculture, food industry (see SI) and rural regions (Hokkaido, Tohoku, Shikoku, Kyusyu-Okinawa). Construction sector dominates in the over-expression of

third largest community and corresponding over-express region indicates these firms are mainly based in Kanto and Tokyo. Transport and Wholesale retail trade are the dominate sectoral attributes in fourth largest community with Tohoku, Kanto, Chubu as over-express regions. The fifth largest community are mainly based on Tokyo and primary over-expressed sectors are Information and communications, Scientific research, professional and technical services. The six largest community primarily showing the grouping in medical and health care. In summary we conclude following facts that characterize the largest six communities:

- The largest community: Manufacturing sectors.
- The second largest community: Food sectors.
- The third largest community: Construction sectors.
- The fourth largest community: Wholesale and Retail Trade.
- The fifth largest community: IT sector and scientific research primarily based in Tokyo.
- The six largest community: Medical and health care.

Table 4: **Overexpressions of the level 1 communities**

Index	Size	#subcom	Region	Sector	IN	GSCC	OUT
1	175,150	7135	Kanto (0.21); Tokyo (0.14); Chubu (0.22); Kansai (0.21)	Manufacturing (0.33);	0.20	0.65	0.14
2	126,997	5455	Hokkaido (0.07); Tohoku (0.11); Shikoku (0.05); Kyusyu-Okinawa (0.13)	Agriculture (0.04); Manufacturing (0.18); Wholesale and retail (0.43); Accommodationss (0.11); Living-related (0.03); Compound services (0.02)	0.11	0.46	0.40
3	96,062	7339	Kanto (0.48); Tokyo (0.25)	Construction (0.64); Real estate (0.09); Scientific research (0.06);	0.39	0.38	0.16
4	87,647	2660	Tohoku (0.11); Kanto (0.22); Chubu (0.20)	Transport (0.15); Retail (0.38); Finance (0.05); Services, N.E.C. (0.17)	0.11	0.43	0.44
5	63,611	3631	Tokyo (0.40)	Information (0.25); Finance (0.01); Real estate (0.05); Scientific research (0.13); Living-related (0.05); Education (0.01); Services, N.E.C. (0.07)	0.26	0.45	0.26
6	47, 759	6214	Hokkaido (0.06); Tokyo (0.22); Chugoku (0.08); Shikoku (0.05); Kyusyu-Okinawa (0.13)	Wholesale and retail (0.28); Living-related (0.05); Medical (0.48)	0.24	0.21	0.52

"#subcom" is the total number of subcommunities included in each of the level 1 communities. The overexpression in terms of regions and sector-divisions in the six largest communities at level 1. The fraction of nodes having the particular attribute is also indicated within the parentheses. Those with fraction less than 0.01 are not listed. Also the fractions of the IN, GSCC, and OUT components are listed for each community.

Figure 7 is a coarse-grained diagram of the network shown in Fig. 1, where the 50 largest communities at the top level are represented by nodes and directed links connecting them in an either way are bundled into arrows. We prepared the diagram taking the following steps. We first calculated the centers of mass for the IN, GSCC, and OUT components in three-dimensional space. The three centers thus obtained determines a two-dimensional plane for the drawing. Secondly, we fixed its horizontal

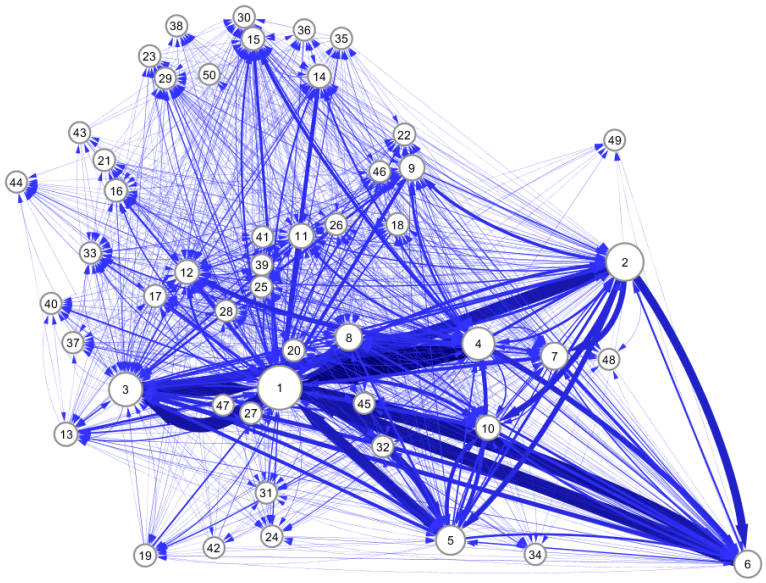


Figure 7: **Network of the 50 largest communities at the top level.** The major communities are depicted as nodes and the size of them is scaled to the size of their corresponding communities. A bundle of directed links connecting a pair of nodes in an either way is represented by an arrow, the width of which is proportional to the total number of their links.

axis to optimally represent the direction of flow from IN (left-hand side) to OUT (right-hand side) through GSCC; in fact, the three centers are almost aligned horizontally. Then we calculated the centers of mass of the major communities and projected them onto the two-dimensional plane to layout the major communities on it. Finally we connected those communities by arrows using information on the links between them.

The positions of communities in the horizontal direction reflect quite well their characteristics as regards the walnut structure such as given in Table 4. Among the 6 largest communities, the third community containing twice as many IN components as compared with the averaged concentration is located in the leftmost side. On the other hand, the sixth community with the largest OUT concentration is in the rightmost side. The second and fourth communities, which are dominated by the OUT components, are also in the right-hand side. The first community with the excess GSCC components is between the third community and those OUT-excess communities. The fifth community, whose composition is very close to the average one, is rather in middle of the walnut structure. Most of the remaining relatively small communities are localized on the left-hand side. This is understandable, because the IN and GSCC components tend to form integrated communities as will be shown later.

Level-2 communities

At the second level, some of the top level communities are decomposed to several sub-communities as shown in Table 9.

The cumulative distribution of the community size at this level is plotted in Fig 8. We use maximum likelihood estimation (MLE) [28] to quantitatively fit a statistically significant power-law decay for the tail of the CCDF having the functional form $D(s) \sim s^{-\gamma+1}$ with $\gamma = 2.50 \pm 0.02$. It indicates that the sizes of the communities are highly heterogeneous and spanning over several orders of magnitude.

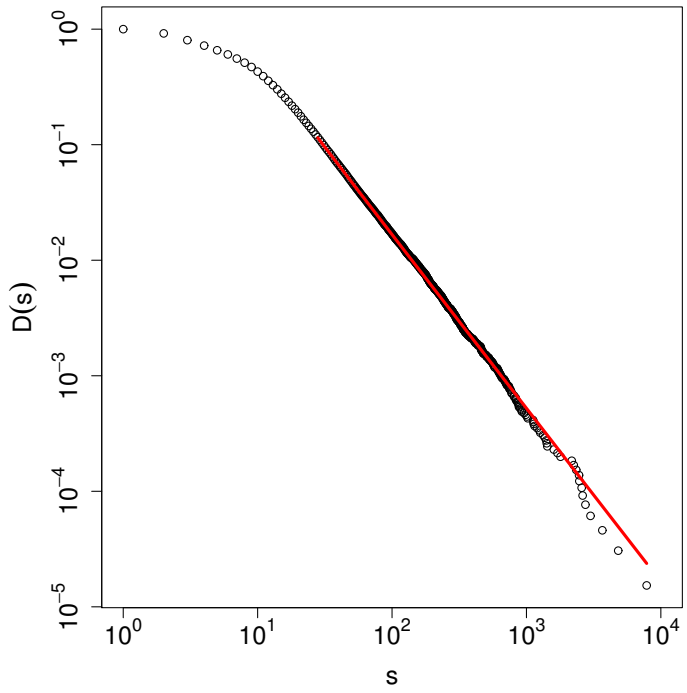


Figure 8: (color online) **The complementary cumulative distribution function $D(s)$ of the community sizes s at the second modular level.** A power-law fit to the data (red line) using the maximum likelihood estimation technique yields $D(s) \sim s^{-\gamma+1}$ with $\gamma = 2.50 \pm 0.02$, $s_{min} = 28.2 \pm 7.6$, and $p\ value = 0.976$.

We have also done analysis of overexpressions of selected sub communities. In case of sub communities, we observe Wholesale and retail trade is the dominate overexpress attribute in the largest five sub communities of the largest community. Kansai region is the only over-expressed region in the second largest sub community of the largest community. In the $C_{2:1}$ Transport and postal activities, Accommodations, eating and drinking services, Living related and personal services and amusement services are the dominate over-express sectors which are mainly based in urban region (Tokyo and Chubu). Manufacturing, Wholesale and retail trade in Tokyo and Kansai region is over-expressed in $C_{2:2}$. Wholesale and retail trade is the dominate over-express attribute in $C_{2:3}$, $C_{2:4}$ and $C_{2:5}$. The detail results are given in the Supplement.

The network diagram in Fig 9 shows the overlapping nature of the industrial sectors in communities. We construct the weighted undirected network of 97 major sectors from sector over expression data in second modular level. Here a weighted link of value 1 is formed between a pair of sectors if they are over expressed within the same community.

The link-weight of the network is found to be highly heterogeneous having a board nature of the distribution as shown in Fig. 10. The top five heavy weight links between sectors are listed in Table 5.

Table 5: **Top five heavy weight links between sectors:**

Rank	Node 1	Node 2	Weight
1	Retail trade (machinery and equipment)	Automobile maintenance services	48
2	Miscellaneous wholesale trade	Miscellaneous retail trade	28
3	Road passenger transport	Automobile maintenance services	21
4	Miscellaneous manufacturing industries	Miscellaneous wholesale trade	19
5	Road passenger transport	Retail trade (machinery and equipment)	19

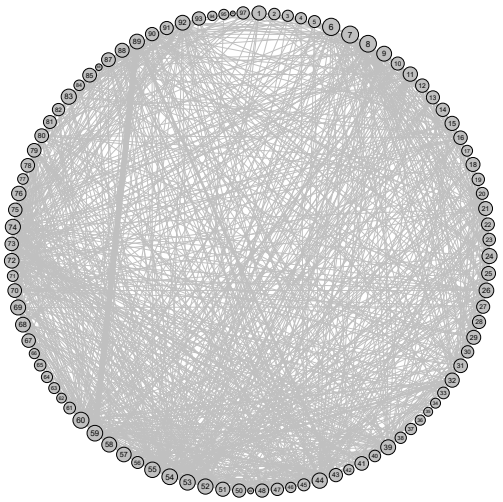


Figure 9: **Overexpression Network of Sectors.** The size of the node represents the fraction of firms belong to that particular sector.

Figure 11 is the same plot as Fig. 6 but for communities at the second modular level. We can confirm that links between the subcommunities are well polarized. Once again this result is consistent with the nature of the map equation which extracts communities of nodes tightly connected in a bidirectional way in a directed network.

Figure 12 shows how mixed are the IN, OUT, and GSCC components of the walnut structure in each of large communities with more than 50 firms at the second level, adopting a triangular diagram representation.[2] Here 3,011 communities containing more than 50 firms are selected with 421,779 firms in total. Suppose a community contains firms belonging to the IN, OUT, and GSCC parts whose fractions are given by x_1, x_2, and x_3, respectively. The walnut composition of the community is described by a point (x_1, x_2, x_3) on the plane of $x_1+x_2+x_3 = 1$ in three-dimensional space. One can thereby establish one-to-one correspondence between a point inside an equilateral triangle and

[2]We exclude firms belonging to TE out of this analysis; those are minor components of the walnut structure.

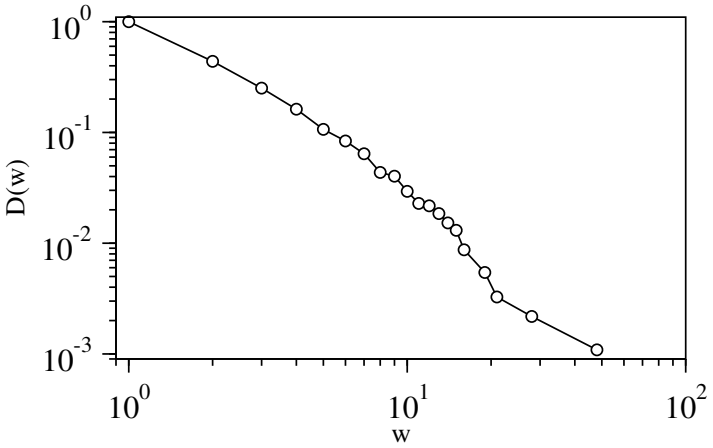


Figure 10: **The complementary cumulative distribution of link-weight in over-expression network.**

a composition of the three walnut components. Their averaged compositions over all firms in the selected communities are given by $\bar{x}_1 = 0.174$, $\bar{x}_2 = 0.333$, and $\bar{x}_3 = 0.493$. The triangular region in Figure 12 is then decomposed into six domains in reference to $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$: communities in the domain G ($x_1 < \bar{x}_1$, $x_2 < \bar{x}_2$, $x_3 > \bar{x}_3$) are GSCC-dominant; those in IG ($x_1 > \bar{x}_1$, $x_2 < \bar{x}_2$, $x_3 > \bar{x}_3$), GSCC-IN hybrid; those in I ($x_1 > \bar{x}_1$, $x_2 < \bar{x}_2$, $x_3 < \bar{x}_3$), IN-dominant; those in IO ($x_1 > \bar{x}_1$, $x_2 > \bar{x}_2$, $x_3 < \bar{x}_3$), IN-OUT hybrid; those in O ($x_1 < \bar{x}_1$, $x_2 > \bar{x}_2$, $x_3 < \bar{x}_3$), OUT-dominant; those in GO ($x_1 < \bar{x}_1$, $x_2 > \bar{x}_2$, $x_3 > \bar{x}_3$), GSCC-OUT hybrid. The total numbers of communities and firms in each domain are listed in Table 6. We thus observe that the IN components tend to merge with the GSCC components to form a single community. On the other hand, there are an appreciable number of communities dominated by the OUT components, leading to relatively few communities of IN-OUT and GSCC-OUT hybrids.

Though the IN components tend to to merge with the GSCC, we can see the large circle at the vertex of Figure 12. On the other hand, Table 2 shows that most nodes in the IN component have distance 1 to the GSCC. Therefore, one may think how there is a large community almost purely composed of nodes in the IN components of the Walnut shape (Figure 2). Actually, this indicates the interesting structure where the nodes are mutually connected and simultaneously they are connected to nodes in the GSCC. It can be precisely said the shell community of the walnut.

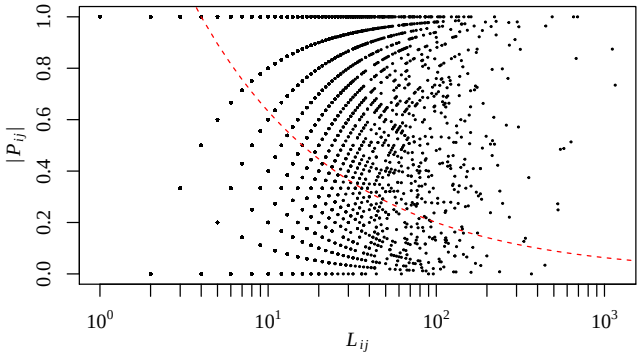


Figure 11: **Polarizability of direction of links interconnecting communities at the second level.** Here 1086 communities containing over 100 firms are selected. The dashed curve represents the same significance level as in Fig. 6.

Table 6: **Classification of communities at the second level by the walnut structure**

Domain	#com	#firms
G	1,010	114,399
IG	841	92,163
I	294	44,563
IO	80	14,362
O	640	139,986
GO	146	16,306
Total	3,011	421,779

"#com" and "#firms" refer to the total numbers of communities and firms, respectively, in each of the six domains defined in Fig. 12(b).

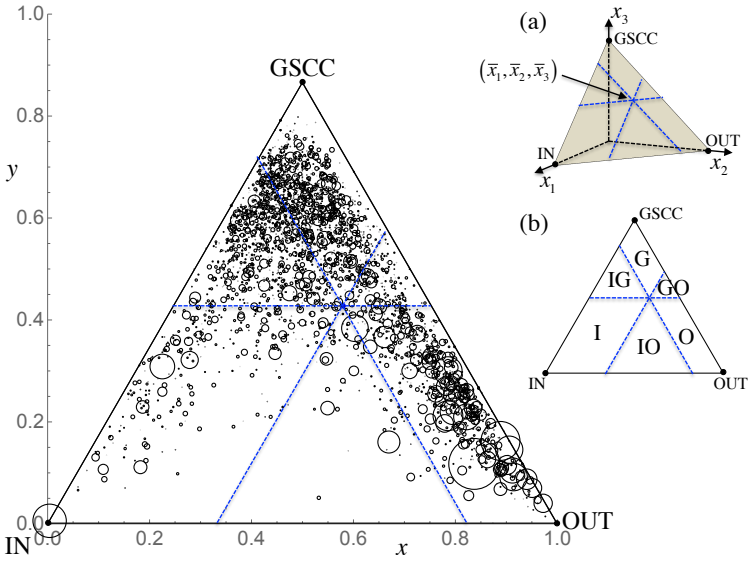


Figure 12: **Triangular diagram classifying communities at the second level by their relationship with the walnut structure.** Each community is depicted by a circle located at a point $(x,\ y)$ inside the equilateral triangle which corresponds to its composition $(x_1,\ x_2,\ x_3)$ of firms belonging to the IN, OUT, and GSCC parts, respectively, represented in three-dimensional space; the one-to-one correspondence between $(x,\ y)$ and $(x_1,\ x_2,\ x_3)$ is illustrated in the associated figure (a). The size of communities is reflected on the area of their associated circles. The triangular region is decomposed into six domains with the averaged composition $(\bar{x}_1,\ \bar{x}_2,\ \bar{x}_3)$ of the IN, OUT, and GSCC components over all firms as designated in the associated figure (b); see the text for more detailed information on the domain decomposition.

Comparison with Industrial Sectors

As is mentioned in Introduction Section, detecting communities from the supply-chain network is crucial for understanding agglomerative behaviours of firms. This is because detected communities are densely connected and it is plausible that those firms affect each other through the links.

On the other hand, industrial sectors are used to label firms and they are widely used in economics literature. If there is no difference between the detected communities and the industrial sectors, we do not have to bother to detect communities. Therefore, in this section, we show how the detected communities are different from industrial sectors in the sense of inter connections between groups.

Though there are different classifications of industrial sectors, we discuss one used in the input-output table [29]. This is because the input-output table is a major research domain in economics, and, more importantly, the purpose of the input-output table is to discuss money flow, which corresponds to the purpose of this paper.

As is mentioned earlier, there are 209 communities for the first level and 66,133 communities for the second level. On the other hand, input-output tables have sectoral classifications of 13, 37, 108, 190, and 397 which are nested. We choose 209 communities and 190 industrial sectors to compare because they are comparable in the numbers.

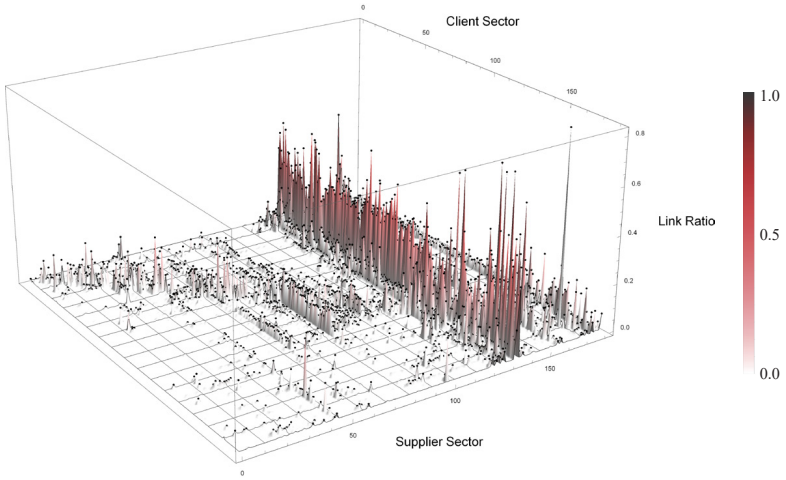
First, we have counted the number of links between communities and industrial sectors. Figure 13 shows the difference of them. These figures correspond to matrices that have numbers of links from row groups to column groups. Each element is divided by the sum of its row.

If intra-links within groups are dominant, diagonal elements of these matrices should show high density. As is shown in Figure 13, we can find the diagonal elements in communities are denser than other elements. However, the diagonal elements in sectors do not have dense links. We see a vertical line in the matrix instead. The suppliers in the line are 5111: Wholesale and 5112: Retailing and this result is natural because firms sell their products to the industrial sectors. The overall ratio of intra-links, i.e., (number of intra-group links)/(number of all links) are 20.9% for industrial sectors and 63.3% for communities.

We can conclude that the detected communities in this paper explicitly grasp the agglomeration of firms based on the supply-chain network compared to the commonly used industrial sectors. This result also tells us that community with densely connected firms consist of various industrial sectors and they have their own economies, i.e., small universes.

Through this paper, the network does not have weights on the links. However, Obviously, the transaction has a value and it has diversity and we can assume the weights on the basis of an assumption by using sales of firms. If we have totally different results with the results we have obtained, we might need extra discussion. However, the additional analyses based on the weight networks do not show significant difference. The detail is shown in Appendix Intra-link density for weighted links.

(a) Ratio of supplier links by sector



(b) Ratio of supplier links by community

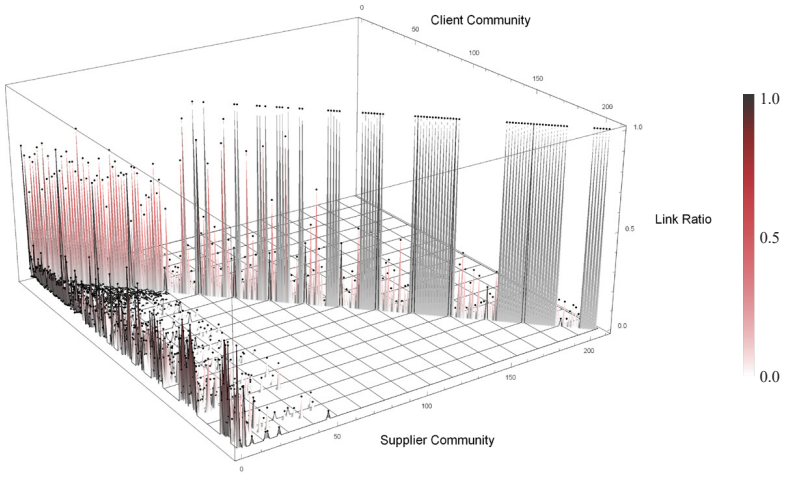


Figure 13: **Density of links over inter groups** These figures show how many links inter groups have. The top figure (a) shows the 3D plots for industrial sectors. The bottom figure (b) shows the 3D plots for communities.

Conclusion and Discussions

We have analyzed the overall structure and the hierarchical communities embedded in the production network of one million firms and five million links for trade relation ship in Japan at the year 2016, with the aim of setting the basis of simulations of macro/micro level dynamics of economy.

For the former, we have found the IN and OUT components (20 % and 26% of the firms) form tight shells (semi-sphere) of the GSCC component, which we name "Walnut" structure, not the "bow-tie" structure well-known for web network and others who has loose wings made of IN and OUT components.

As to the latter, we have used Infomap method to detect 5 layers of hierarchy of communities, most of the irreducible (those that do not have any lower level sub-communities) belonging to the second level. Furthermore, the size distribution of the second-level communities have clear power-law behavior at the large end. In addition to the large number of irreducible communities that is made of GSCC components mostly and those exist in IN shells or Out shells, there are fair number of communities made of IN and GSCC components, GSCC and OUT components, and even IN and OUT components. This is expected from the Walnut characterization of the overall structure: IN and OUT components are not far from each other like in bow-tie structure, but they form tight shells, whose ends are closely woven with each other. Furthermore, we have examined overexpression of the major communities in terms of industrial sectors and prefectures and have found they are not formed within a sector, but spans several sectors and prefectures. How they are shaped is different from community to community: in some cases they are formed surrounding goods and services related to particular item, such as foods. Sometimes they are made of small firms connected with a major hub such as the big construction company in the particular prefecture, or a medical insurance agency.

These findings have major implications in the study of macro economy: Consider economic crisis. once it starts due to a natural disaster in a particular region of a country, or a major failure of a big company, it is expected that it initially affects in the community where they are located. Then it will spread from other neighboring communities. This is so different from input-output analysis, who is expected to be useful on the basis of assumption that firms in the same sectors are well-connected with each other. Instead, the spread is community-wise.

The immediate applications of the hierarchical community structure studied in this paper includes the application to the analysis of large-scale modelling and simulation: Macro economy of a country or countries is an aggregation of products of economic dynamics on the trade network and other multitude of network of networks. Constructing models that span all the networks is an elaborating work. Instead, we may study one community at a time and then connect the results to obtain overall picture. Research on this direction are already starting and will appear in near future ([30–33]).

Acknowledgments

We are grateful to Y. Ikeda, W. Souma and H. Yoshikawa for insightful comments and encouragements. We are also grateful to Tokyo Shoko Research Ltd. and RIETI for making this research possible by providing us the production network data. This study was supported in part by the Project "Large-scale Simulation and Analysis of Economic Network for Macro Prudential Policy" undertaken at Research Institute of Economy, Trade and Industry (RIETI), MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena), Grant-in-Aid for Scientific Research (KAKENHI) by JSPS Grant Numbers 25400393, 17H02041 and

the Kyoto University Supporting Program for Interaction-based Initiative Team Studies: SPIRITS, as part of the Program for Promoting the Enhancement of Research Universities, MEXT, JAPAN

References

[1] Tesfatsion L, Judd K, editors. Agent-Based Computational Economics, Handbook of Computational Economics, vol.2. North Holland; 2006.

[2] Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G. DebtRank: too central to fail? Financial networks, the FED and systemic risk. Scientific reports. 2012;2.

[3] Abergel F, Aoyama H, Chakrabarti BK, Chakraborti A, Ghosh A. Econophysics of agent-based models. Springer; 2013.

[4] Caiani A, Russo A, Palestrini A, Gallegati M. Economics with Heterogeneous Interacting Agents. Springer; 2016.

[5] Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Souma W, Yoshikawa H. Macro-Econophysics – New Studies on Economic Networks and Synchronization. Cambridge University Press; 2017.

[6] Fujiwara Y, Aoyama H. Large-scale structure of a nation-wide production network. The European Physical Journal B-Condensed Matter and Complex Systems. 2010;77(4):565–580.

[7] Fujiwara Y. Omori law after large-scale destruction of production network. Progress of Theoretical Physics Supplement. 2012;194:158–164.

[8] Iino T, Iyetomi H. Community Structure of a Large-Scale Production Network in Japan. In: Watanabe T, Uesugi I, Ono A, editors. The Economics of Interfirm Networks. Tokyo: Springer Japan; 2015. p. 39–65. Available from: `https://doi.org/10.1007/978-4-431-55390-8_3`.

[9] Chakraborty A, Krichene H, Inoue H, Fujiwara Y. Characterization of the community structure in a large-scale production network in Japan. arXiv preprint arXiv:170600203. 2017;.

[10] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, et al. Graph structure in the Web. Computer Networks. 2000;33(1-6):309–320.

[11] Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences. 2008;105(4):1118–1123.

[12] Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Software: Practice and Experience. 1991;21(11):1129–1164. doi:10.1002/spe.4380211102.

[13] Leskovec J, Lang K, Dasgupta A, Mahoney M. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. Internet Mathematics. 2009;6(1):29–123.

[14] Newman ME. Fast algorithm for detecting community structure in networks. Physical review E. 2004;69(6):066133.

[15] Fortunato S, Barthélemy M. Resolution limit in community detection. Proceedings of the National Academy of Sciences. 2007;104(1):36–41.

[16] Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis. Physical review E. 2009;80(5):056117.

[17] Kawamoto T, Rosvall M. Estimating the resolution limit of the map equation in community detection. Phys Rev E. 2015;91:012809. doi:10.1103/PhysRevE.91.012809.

[18] Rosvall M, Bergstrom CT. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PloS one. 2011;6(4):e18209.

[19] Viamontes Esquivel A, Rosvall M. Compression of Flow Can Reveal Overlapping-Module Organization in Networks. Phys Rev X. 2011;1:021025. doi:10.1103/PhysRevX.1.021025.

[20] Fortunato S. Community detection in graphs. Physics reports. 2010;486(3):75–174.

[21] Currarini S, Jackson MO, Pin P. An economic model of friendship: Homophily, minorities, and segregation. Econometrica. 2009;77(4):1003–1045.

[22] Echenique F, Fryer Jr RG. A measure of segregation based on social interactions. The Quarterly Journal of Economics. 2007;122(2):441–485.

[23] Chen J, Yuan B. Detecting functional modules in the yeast protein–protein interaction network. Bioinformatics. 2006;22(18):2283–2290.

[24] Onnela JP, Chakraborti A, Kaski K, Kertesz J, Kanto A. Dynamics of market correlations: Taxonomy and portfolio analysis. Physical Review E. 2003;68(5):056110.

[25] Tumminello M, Miccichè S, Lillo F, Varho J, Piilo J, Mantegna RN. Community characterization of heterogeneous complex systems. Journal of Statistical Mechanics: Theory and Experiment. 2011;2011(01):P01019.

[26] Drăghici S. Data analysis tools for DNA microarrays. CRC Press; 2003.

[27] Miller Jr RG. Normal univariate techniques. In: Simultaneous statistical inference. Springer; 1981. p. 37–108.

[28] Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661–703.

[29] Leontief W. Quantitative Input and Output Relations in the Economic Systems of the United States. The Review of Economics and Statistics. 1936;18(3):105–125.

[30] Krichene H, Chakraborty A, Inoue H, Fujiwara Y. Business cycles correlation and systemic risk of the Japanese supplier-customer network. PloS one. 2017;12(10):e0186467.

[31] Krishene H, Arata Y, Chakraborty A, Fujiwara Y, Inoue I. How firms choose their partner in the Japanese supplier-customer network? *to appear*. 2018;.

[32] Arata Y. The Risk of the Propagation of Corporate Failure: an empirical analysis. *to appear*. 2018;.

[33] Kichikawa Y, Iino T, Iyetomi H, Inoue H. Hierarchical and Circulating Flow Structure in the Japanese Production Network. *to appear*. 2018;.

[34] Inoue H, Todo Y. Propagation of negative shocks through firm networks: evidence from simulation on comprehensive supply-chain data. RIETI Discussion Papers. 2017;17-E-044:1–21.

[35] Rand WM. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association. 1971;66(336):846–850. doi:10.1080/01621459.1971.10482356.

[36] Jaccard P. tude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Socit Vaudoise des Sciences Naturelles. 1901;37:547–579.

Supporting information

Data classifications

Table 7 is the list of number of firms in the 20 divisions of the industrial sectors.

Table 7: **Industrial sectors and the firm distribution**

ID	Code	Sector	# Firms	%
1	A	Agriculture	9,841	0.92
2	B	Fisheries	1,211	0.11
3	C	Mining	1,268	0.12
4	D	Construction	357,199	33.51
5	E	Manufacturing	156,188	14.65
6	F	Electricity, Gas, Heat Supply & Water (*EGW)*	1,470	0.14
7	G	*Information* & Communications	26,539	2.49
8	H	*Transport* & Postal	36,736	3.45
9	I	Wholesale & *Retail* Trade	254,251	23.85
10	J	*Finance* & Insurance	7,506	0.70
11	K	Real Estate	41,837	3.92
12	L	*Scientific Research*, Professional & Technical Services	42,030	3.94
13	M	*Accommodations*, Eating/Drinking Services	17,322	1.62
14	N	*Living-related*/Personal & Amusement Services	17,365	1.63
15	O	*Education*, Learning Support	4,655	0.44
16	P	*Medical*, Health Care & Welfare	30,154	2.83
17	Q	Compound Services	6,472	0.61
18	R	Other Services	52,190	4.90
19	S	Government	1,803	0.17
20	T	Unable to classify	0	0.0

Number of firms in classification by industrial sectors, which is based on the division in the Japan Industrial Sector Classification. The words in italic are used as abbreviation in the main text.

Table 8 is list of the firms numbers in the 8 regions and city of Tokyo of Japan, which is illustrated in Figure 14.

Table 8: **Regional areas and the firm distribution**

id	region	#firms	%
1	Hokkaido	54,423	5.11
2	Tohoku	87,374	8.20
3	Kanto	187,186	17.56
4	Tokyo	146,614	13.75
5	Chubu	196,477	18.43
6	Kansai	168,701	15.83
7	Chugoku	69,312	6.50
8	Shikoku	40,397	3.79
9	Kyusyu-Okinawa	115,553	10.84

Number of firms in each regional areas, determined by the geographical location of their main office regional area. “Kanto” means “Kanto less Tokyo, as the greater ”Tokyo” belongs to ”Kanto” region.

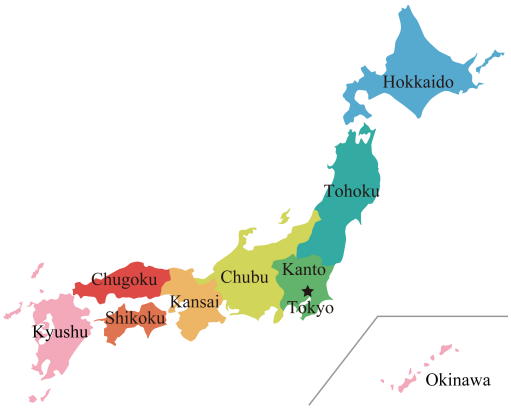


Figure 14: **Eight regions and Tokyo in Japan.**

Overexpressions: level 1

In the main text, we have presented the overexpression of 20 sector divisions and 9 regions in the selected large communities. Here, we further expose the overexpression of 99 sectors and 47 prefectures in the following communities.

Rank: 1

Over-expression of prefectures: Ibaraki; Gunma; Saitama; Chiba; Tokyo; Kanagawa; Nagano; Shizuoka; Aichi; Mie; Osaka; Hyogo; Hiroshima

Over-expression of sectors: Equipment installation work; Manufacture of plastic products, except otherwise classified; Manufacture of rubber products; Manufacture of iron and steel; Manufacture of non-ferrous metals and products; Manufacture of fabricated metal products; Manufacture of generalpurpose machinery; Manufacture of production machinery; Manufacture of business oriented machinery; Electronic parts, devices and electronic circuits; Manufacture of electrical machinery, equipment and supplies; Manufacture of information and communication electronics equipment; Manufacture of transportation equipment; Production, transmission and distribution of electricity; Heat supply; Wholesale trade, general merchandise; Wholesale trade (building materials, minerals and metals, etc); Wholesale trade (machinery and equipment); Retail trade (machinery and equipment); Machine, etc. repair services, except otherwise classified

Rank: 2

Over-expression of prefectures: Hokkaido; Aomori; Iwate; Miyagi; Akita; Yamagata; Fukushima; Niigata; Yamanashi; Nagano; Shizuoka; Tottori; Shimane; Tokushima; Kagawa; Ehime; Kochi; Saga; Nagasaki; Kumamoto; Oita; Miyazaki; Kagoshima

Over-expression of sectors: Agriculture; Fisheries, except Aquaculture; Aquaculture; Manufacture of food; Manufacture of beverages, tobacco and feed; Railway transport; Warehousing; Wholesale trade, general merchandise; Wholesale trade (food and beverages); Miscellaneous Wholesale trade; Retail trade, general merchandise; Retail trade (food and beverage); Non-store retailers; Financial institutions for cooperative organizations; Non-deposit money corporations, including lending and credit card business; Real estate lessors and managers; Accommodations; Eating and drinking places; Food take out and delivery services; Miscellaneous living-related and personal services; Services for amusement and recreation; Social insurance, social welfare and care services; Cooperative associations, N.E.C; Miscellaneous services

Rank:3

Over-expression of prefectures: Ibaraki; Saitama; Chiba; Tokyo; Kanagawa; Osaka

Over-expression of sectors: Mining and quarrying of stone; Construction work, general including public and private construction work; Construction work by specialist contractor, except equipment installation work; Equipment installation work; Manufacture of ceramic, stone and clay products; Collection, purification and distribution of water and sewage collection, processing and disposal; Railway transport; Financial auxiliaries; Real estate agencies; Real estate lessors and managers; Goods rental and leasing; Technical services, N.E.C.; Social insurance, social welfare and care services; Automobile maintenance services; Political, business and cultural organizations; Local government services

Rank: 4

Over-expression of prefectures: Aomori; Iwate; Miyagi; Akita; Fukushima; Ibaraki; Tochigi; Gunma; Saitama; Chiba; Kanagawa; Ishikawa; Shizuoka; Aichi; Mie; Saga

Over-expression of sectors: Manufacture of petroleum and coal products; Manufacture of transportation equipment; Road Passenger transport; Road freight transport; Water transport; Warehousing; Services incidental to transport; Wholesale trade (machinery and equipment); Retail trade (machinery and equipment); Miscellaneous retail trade; Insurance institutions, including insurance agents brokers and services; Goods

rental and leasing; School education; Waste disposal business; Automobile maintenance services

Rank: 5

Over-expression of prefectures: Tokyo; Kanagawa; Osaka

Over-expression of sectors: Printing and allied industries; Manufacture of business oriented machinery; Electronic parts, devices and electronic circuits; Manufacture of information and communication electronics equipment; Miscellaneous manufacturing industries; Communications; Brodcasting; Information services; Services incidental to internet; Video picture information, sound information, character information production and distribution; Air transport; Wholesale trade (machinery and equipment); Retail trade (machinery and equipment); Miscellaneous retail trade; Non-store retailers; Banking; Financial institutions for cooperative organizations; Non-deposit money corporations, including lending and credit card business; Financial products transaction dealers and futures commodity dealers; Financial auxiliaries; Real estate lessors and managers; Goods rental and leasing; Professional services, N.E.C.; Advertising; Technical services, N.E.C.; Miscellaneous living-related and personal services; Services for amusement and recreation; School education; Miscellaneous education, learning support; Employment and worker dispatching services; Miscellaneous business services; Political, business and cultural organizations

Overexpressions: Second modular level

Short summary of the overexpression of 20 sector divisions and 9 regions in level 2 communities is given in Table 9 and 10.

Further, we present the overexpression of 99 sectors and 47 prefectures in the following subcommunities.

Five largest sub communities

Rank: 1

Over-expression of prefectures: Hokkaido; Tokyo; Fukui; Yamanashi; Tokushima; Kagawa; Ehime

Over-expression of sectors: Information services; Financial products transaction dealers and futures commodity dealers; Insurance institutions, including insurance agents brokers and services; Real estate agencies; Real estate lessors and managers; Professional services, N.E.C.; Miscellaneous living-related and personal services

Rank: 2

Over-expression of prefectures: Ibaraki; Tochigi; Gunma; Saitama; Chiba; Kanagawa; Nagano; Gifu; Shizuoka; Aichi; Mie; Osaka; Hyogo; Nara; Wakayama

Over-expression of sectors: Miscellaneous retail trade; Medical and other health services

Rank: 3

Over-expression of prefectures: Iwate; Tochigi; Kanagawa; Kyoto; Hiroshima; Fukuoka; Miyazaki; Kagoshima

Over-expression of sectors: Medical and other health services

Rank: 4

Over-expression of prefectures: Hokkaido; Iwate; Tochigi; Gunma; Gifu; Aichi; Kyoto; Fukuoka; Kagoshima

Over-expression of sectors: Manufacture of plastic products, except otherwise classified; Manufacture of business oriented machinery; Wholesale trade (machinery and equipment); Miscellaneous Wholesale trade; Real estate lessors and managers; Services for amusement and recreation

Rank: 5

Over-expression of prefectures: Hokkaido; Chiba; Kanagawa; Gifu; Aichi

Over-expression of sectors: Wholesale trade, general merchandise; Wholesale trade (machinery and equipment); Retail trade (machinery and equipment)

Five largest sub communities of the largest community:

Rank: 6

Over-expression of prefectures: Hokkaido; Aomori; Iwate; Akita; Niigata; Toyama; Ishikawa; Fukui; Okayama; Yamaguchi; Tokushima; Kagawa; Ehime; Kochi; Kumamoto; Oita; Kagoshima

Over-expression of sectors: Retail trade (machinery and equipment)

Rank: 16

Over-expression of prefectures: Chiba; Osaka; Wakayama

Over-expression of sectors: Construction work by specialist contractor, except equipment installation work; Manufacture of textile products; Manufacture of iron and steel; Road freight transport; Warehousing; Services incidental to transport; Wholesale trade (textile and apparel); Wholesale trade (building materials, minerals and metals, etc); Waste disposal business

Rank: 25

Over-expression of prefectures: Hokkaido; Aomori; Akita; Toyama; Yamaguchi; Tokushima; Kochi; Saga; Okinawa

Over-expression of sectors: Retail trade (machinery and equipment)

Rank: 38

Over-expression of prefectures: Hokkaido; Iwate; Akita; Niigata; Toyama; Fukui

Over-expression of sectors: Retail trade (machinery and equipment)

Rank: 39

Over-expression of prefectures: Hokkaido; Shimane; Miyazaki; Kagoshima; Okinawa

Over-expression of sectors: Manufacture of furniture and fixtures; Wholesale trade (machinery and equipment); Miscellaneous Wholesale trade; Eating and drinking places

Five largest sub communities of the second largest community:

Rank: 8

Over-expression of prefectures: Tokyo; Yamanashi; Nagano; Okinawa

Over-expression of sectors: Road Passenger transport; Water transport; Accommodations; Miscellaneous living-related and personal services; Miscellaneous education, learning support; Employment and worker dispatching services; Political, business and cultural organizations

Rank: 23

Over-expression of prefectures: Tokyo; Kanagawa; Kyoto; Osaka

Over-expression of sectors: Manufacture of textile products; Manufacture of furniture and fixtures; Manufacture of leather tanning, leather products and fur skins; Miscellaneous manufacturing industries; Wholesale trade, general merchandise; Wholesale trade (textile and apparel); Miscellaneous Wholesale trade; Retail trade, general merchandise; Retail trade (woven fabrics, apparel, apparel accessories and notions)

Rank: 27

Over-expression of prefectures: Saitama; Aichi; Osaka

Over-expression of sectors: Manufacture of food; Manufacture of production machinery; Wholesale trade (food and beverages)

Rank: 28

Over-expression of prefectures: Saitama; Tokyo; Kanagawa; Kyoto; Osaka; Hyogo

Over-expression of sectors: Manufacture of beverages, tobacco and feed; Wholesale trade (food and beverages); Retail trade, general merchandise; Retail trade (food and beverage)

Rank: 29

Over-expression of prefectures: Ibaraki; Tochigi; Gunma; Chiba

Over-expression of sectors: Agriculture; Forestry; Construction work, general including public and private construction work; Manufacture of chemical and allied product; Miscellaneous Wholesale trade; Miscellaneous retail trade

Five largest sub communities of the third largest community:

Rank: 12

Over-expression of prefectures: Hokkaido; Miyagi; Tokyo; Aichi; Hiroshima; Kagawa; Fukuoka

Over-expression of sectors: Construction work by specialist contractor, except equipment installation work; Manufacture of fabricated metal products; Goods rental and leasing; Technical services, N.E.C.; Employment and worker dispatching services; Miscellaneous business services

Rank: 41

Over-expression of prefectures: Tokyo; Osaka

Over-expression of sectors: Real estate agencies; Real estate lessors and managers; Professional services, N.E.C.; Advertising

Rank: 46

Over-expression of prefectures: Tokyo

Over-expression of sectors: Professional services, N.E.C.; Technical services, N.E.C.; Social insurance, social welfare and care services; Political, business and cultural organizations; Local government services

Rank: 53

Over-expression of prefectures: Iwate; Akita; Tochigi; Niigata; Ishikawa; Nagano; Shizuoka; Aichi; Mie; Nara; Wakayama; Okayama; Hiroshima; Yamaguchi; Kagawa; Fukuoka; Nagasaki; Oita

Over-expression of sectors: Construction work, general including public and private construction work; Construction work by specialist contractor, except equipment installation work

Rank: 87

Over-expression of prefectures: Miyagi; Tochigi; Niigata; Toyama; Gifu; Shizuoka; Aichi; Mie; Tottori; Shimane; Hiroshima; Tokushima; Kagawa; Fukuoka; Saga; Oita; Okinawa

Over-expression of sectors: Construction work, general including public and private construction work; Construction work by specialist contractor, except equipment installation work

Five largest sub communities of the fourth largest community:

Rank: 1

Over-expression of prefectures: Hokkaido; Tokyo; Fukui; Yamanashi; Tokushima; Kagawa; Ehime

Over-expression of sectors: Information services; Financial products transaction dealers and futures commodity dealers; Insurance institutions, including insurance agents brokers and services; Real estate agencies; Real estate lessors and managers; Professional services, N.E.C.; Miscellaneous living-related and personal services

Rank: 5

Over-expression of prefectures: Hokkaido; Chiba; Kanagawa; Gifu; Aichi

Over-expression of sectors: Wholesale trade, general merchandise; Wholesale trade (machinery and equipment); Retail trade (machinery and equipment)

Rank: 11

Over-expression of prefectures: Saitama; Kanagawa; Ehime

Over-expression of sectors: Retail trade (machinery and equipment)

Rank: 17

Over-expression of prefectures: Hokkaido; Osaka

Over-expression of sectors: Manufacture of transportation equipment; Wholesale trade (machinery and equipment); Retail trade (machinery and equipment); Machine, etc. repair services, except otherwise classified

Rank: 24

Over-expression of prefectures: Osaka

Over-expression of sectors: Retail trade (woven fabrics, apparel, apparel accessories and notions); Automobile maintenance services

Five largest sub communities of the fifth largest community:

Rank: 4

Over-expression of prefectures: Hokkaido; Iwate; Tochigi; Gunma; Gifu; Aichi; Kyoto; Fukuoka; Kagoshima

Over-expression of sectors: Manufacture of plastic products, except otherwise classified; Manufacture of business oriented machinery; Wholesale trade (machinery and equipment); Miscellaneous Wholesale trade; Real estate lessors and managers; Services for amusement and recreation

Rank: 22

Over-expression of prefectures: Hokkaido; Tochigi; Gunma; Chiba; Yamanashi; Tokushima

Over-expression of sectors: Road freight transport; Miscellaneous retail trade; Advertising

Rank: 26

Over-expression of prefectures: Nagano; Shizuoka; Aichi; Mie

Over-expression of sectors: Manufacture of lumber and wood products, except fourniture; Miscellaneous manufacturing industries; Miscellaneous Wholesale trade; Miscellaneous retail trade; Miscellaneous education, learning support; Machine, etc. repair services, except otherwise classified

Rank: 42

Over-expression of prefectures: Gifu; Osaka

Over-expression of sectors: Manufacture of business oriented machinery; Miscellaneous manufacturing industries; Miscellaneous Wholesale trade; Miscellaneous retail trade; Goods rental and leasing; Services for amusement and recreation

Rank: 50

Over-expression of prefectures: Kagoshima

Over-expression of sectors: Communications; Services incidental to internet; Retail trade (machinery and equipment)

Table 9: **Overexpressions of the level 2 communities**

Index	Size	Rank	Region	Sector	IN	GSCC	OUT
1:1	2,618	6	Hokkaido (0.04); Tohoku (0.10); Chugoku (0.09); Shikoku (0.05); Kyusyu-Okinawa (0.12)	Retail (0.85)	0.03	0.10	0.86
1:2	1,430	16	Kansai (0.32)	Transport (0.08); Retail (0.31)	0.28	0.68	0.03
1:3	1,132	25	Hokkaido (0.04); Tohoku (0.12); Chugoku (0.11); Shikoku (0.05); Kyusyu-Okinawa (0.11)	Retail (0.77)	0.06	0.14	0.79
1:4	863	38	Hokkaido (0.06); Tohoku (0.11); Shikoku (0.04)	Retail (0.82)	0.05	0.12	0.83
1:5	854	39	Hokkaido (0.04); Chugoku (0.10); Kyusyu-Okinawa (0.13)	Retail (0.43); Accommodations (0.03)	0.11	0.59	0.29
2:1	2,474	8	Tokyo (0.18); Chubu (0.23)	Information (0.02); Transport (0.13); Accommodations (0.19); Living-related (0.49); Education (0.009); Services N.E.C. (0.04)	0.17	0.44	0.38
2:2	1,200	23	Tokyo (0.45); Kansai (0.22)	Manufacturing (0.23); Retail (0.55); Scientific research (0.04)	0.53	0.37	0.09
2:3	1,121	27	Kanto (0.20); Chubu (0.24); Kansai (0.18)	Manufacturing (0.26); Retail (0.60)	0.09	0.56	0.35
2:4	1,022	28	Kanto (0.22); Tokyo (0.25), Kansai (0.26)	Retail (0.72)	0.12	0.34	0.53
2:5	1,010	29	Kanto (0.33)	Agriculture (0.13); Retail (0.67)	0.10	0.49	0.40
3:1	2,173	12	Hokkaido (0.02); Tohoku (0.05); Tokyo (0.33); Chubu (0.09); Chugoku (0.04); Shikoku (0.02); Kyusyu-Okinawa (0.05)	Manufacturing (0.05); Scientific research (0.09), Services N.E.C. (0.06)	0.60	0.36	0.05
3:2	834	41	Tokyo (0.55)	Informations (0.02); Real estate(0.73); Scientific research (0.09)	0.30	0.25	0.40
3:3	776	46	Tokyo (0.97)	Scientific research(0.12); Medical (0.32); Services, N.E.C. (0.22); Government (0.02)	0.70	0.12	0.11
3:4	740	53	Tohoku (0.08); Chubu (0.16); Chugoku (0.06); Shikoku (0.02); Kyusyu-Okinawa (0.09)	Construction (0.78)	0.68	0.26	0.06
3:5	547	87	Hokkaido (0.02); Tohoku(0.06); Chubu (0.20); Chugoku (0.08); Shikoku (0.04); Kyusyu-Okinawa (0.12)	Construction (0.84)	0.83	0.12	0.05

The overexpression in terms of regions and sector-divisions in five largest communities at level 2. "Rank" refers to the rank among all the level-2 subcommunities. The fraction of nodes having the particular attribute is also indicated within the parentheses.

Table 10: **Overexpressions of the level 2 communities, continued**

Index	Size	Rank	Region	Sector	IN	GSCC	OUT
4:1	7,843	1	Hokkaido (0.06); Tokyo (0.20); Shikoku(0.04)	Information (0.02); Finance (0.54); Real estate (0.12); Scientific research (0.03)	0.09	0.13	0.77
4:2	2,747	5	Hokkaido (0.05); Kanto (0.27); Chubu (0.23)	Retail (0.87)	0.04	0.34	0.61
4:3	2,249	11	Shikoku (0.05); Kyusyu-Okinawa (0.13)	Retail (0.87)	0.03	0.12	0.85
4:4	1,416	17	Hokkaido (0.06); Shikoku (0.04)	Retail (0.88)	0.03	0.13	0.83
4:5	1,149	24	Kansai (0.96)	Retail(0.46); Services, N.E.C. (0.36)	0.06	0.43	0.51
5:1	2,996	4	Hokkaido (0.05); Tohoku (0.07); Kanto (0.17); Chubu (0.20); Kansai (0.16); Chugoku (0.06); Kyusyu-Okinawa (0.10)	Living-related (0.54)	0.09	0.23	0.62
5:2	1,248	22	Hokkaido (0.05); Tohoku (0.07); Kanto (0.22); Chubu (0.18)	Transport(0.04); Retail(0.62)	0.11	0.22	0.66
5:3	1,127	26	Chubu (0.28)	Manufacturing(0.12); Retail (0.62), Education (0.09)	0.09	0.29	0.61
5:4	832	42	Kansai (0.17)	Retail (0.33); Real estate (0.10); Living-related(0.27)	0.16	0.39	0.43
5:5	750	50	Hokkaido (0.05); Shikoku (0.04)	Information (0.32); Retail (0.39)	0.26	0.37	0.36

The overexpression in terms of regions and sector-divisions in five largest communities at level 2, continued. "Rank" refers to the rank among all the level-2 subcommunities. The fraction of nodes having the particular attribute is also indicated within the parentheses.

Intra-link density for weighted links

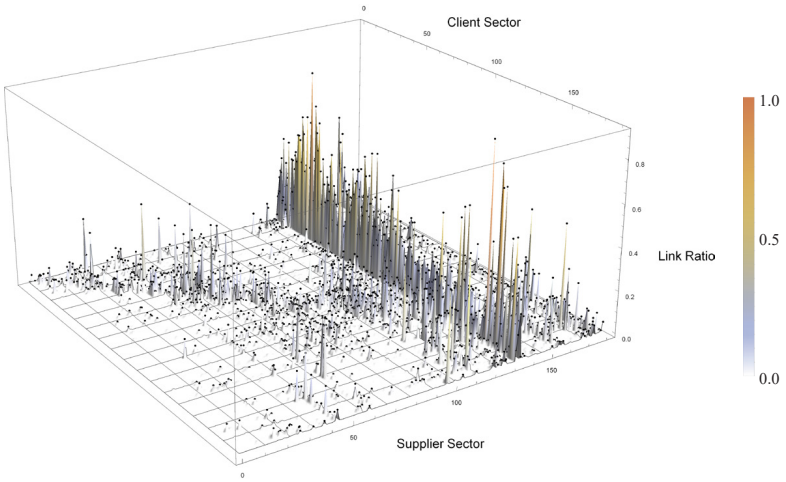
Figure 13 in Comparison with Industrial Sectors Section shows the matrixes that represent the number of inter or intra links between groups. If we add the information of weight, i.e. sales volume, to links and create the same matrixes for the weighted links, the matrixes are helpful to consider the agglomerative behaviour of groups further.

Although the TSR data contain supplier and client relationships, it has no data of sales volume for each relationship. Therefore we artificially add it by the plausible way [34]. Each supplier's sales are proportionally divided into its clients' sales. Here we assume that we ignore supplier's sales to final consumers and the clients volume of purchase from the supplier can be relatively estimated by using the sales of the clients as proxies.

Figure 15 provides the results. The visualization for the industrial sectors has denser connections to wholesale and retaining than the one for the number of links shown in Figure 13. In addition, the visualization for the communities has denser connections to communities in left rows than the one shown in Figure 13. The overall ratio of internal sales volume, i.e., (total volume of intra-group links)/(total volume of all links) are 14.3% for the industrial sectors and 55.9% for the communities.

As a result, we obtain weaker connections of intra-groups for both industrial sectors and communities. The weaker intra-connections of communities is understandable because we do not use the weighted links to detect communities. However, we do not discuss this weighted community detection for simplicity of the discussion of this paper.

(a) Ratio of supplier links by sector



(b) Ratio of supplier links by community

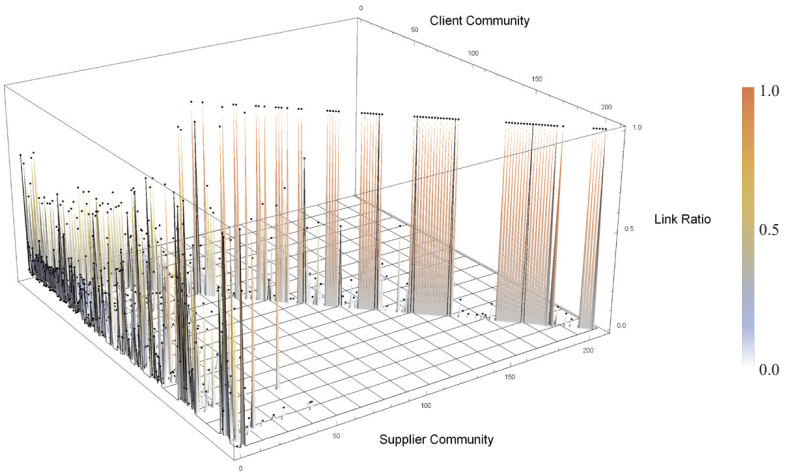


Figure 15: **Density of weighted links over inter and intra-sectors of communities** These figures show how much sales volume inter groups have. The top figure (a) shows the 3D plots for industrial sectors. The bottom figure (b) shows the 3D plots for communities.

1 Comparison of the two community structures

One can quantify similarity between the level-1 community structure obtained with the map equation and that with the modularity in terms of the Rand index [35], a measure of similarity between two data clusterings. The adjusted Rand index, in which coincidental chance that clustering of a pair of nodes is identical is subtracted from the original index, is calculated as 0.325. This large value indicates that the two partitions resemble each other significantly.

In Table 11, more detailed comparison between the two community structures is made, that is, community by community in terms of the Jaccard index [36], a measure of similarity between two sets. We see that there is remarkable one-to-one correspondence between the two partitions. The major communities of the map equation, up to the 7th largest in Table 11, have their counterparts in the partition with modularity.

Table 11: **Jaccard index between major communities obtained with modularity and those with the hierarchical map equation at the top level**

	1	2	3	4	5	6	7	8	9	10
1	0.02	0.64	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
2	0.41	0.01	0.01	0.03	0.01	0.00	0.00	0.05	0.01	0.01
3	0.04	0.01	0.31	0.01	0.01	0.00	0.00	0.03	0.00	0.00
4	0.02	0.03	0.01	0.01	0.11	0.09	0.01	0.12	0.01	0.20
5	0.03	0.00	0.14	0.01	0.00	0.00	0.00	0.02	0.00	0.00
6	0.03	0.03	0.01	0.28	0.01	0.00	0.01	0.03	0.04	0.02
7	0.05	0.01	0.02	0.02	0.32	0.01	0.01	0.01	0.01	0.03
8	0.12	0.01	0.05	0.01	0.02	0.00	0.00	0.02	0.02	0.00
9	0.01	0.03	0.01	0.01	0.01	0.01	0.50	0.02	0.00	0.01
10	0.01	0.01	0.00	0.00	0.01	0.39	0.06	0.01	0.01	0.00

The 10 largest communities of the map equation are aligned in the horizontal direction and those of modularity, in the vertical direction. Note that the Jaccard index takes 1/3 for two sets of equal size with 50% of elements overlapped.