Bow-tie structure and community identification of global supply chain network

Abhijit Chakraborty1,* , Yuichi Ikeda2

1 Graduate School of Simulation Studies, The University of Hyogo, Kobe 650-0047, Japan
2 Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Kyoto 606-8306, Japan

* abhiphyiitg@gmail.com ikeda.yuichi.2w@kyoto-u.ac.jp

Abstract

We study on topological properties of global supply chain network in terms of degree distribution, hierarchical structure, and degree-degree correlation in the global supply chain network. The global supply chain data is constructed by collecting various company data from the web site of Standard & Poor’s Capital IQ platform in 2018. The in- and out-degree distributions are characterized by a power law of the form \( P(k) \propto k^{-\gamma} \) with \( \gamma_{in} = 2.42 \) and \( \gamma_{out} = 2.11 \). The clustering coefficient decays \( \langle C(k) \rangle \sim k^{-\beta_k} \) with an exponent \( \beta_k = 0.46 \). The nodal degree-degree correlation \( \langle k_{nn}(k) \rangle \) indicates the absence of assortativity. The Bow-tie structure of GWCC reveals that the OUT component is the largest and it consists 41.1% of total firms. The GSCC component comprises 16.4% of total firms. We observe that the firms in the upstream or downstream sides are mostly located a few steps away from the GSCC. Furthermore, we uncover the community structure of the network and characterize them according to their location and industry classification. We observe that the largest community consists of consumer discretionary sector mainly based in the US. These firms belong to the OUT component in the bow-tie structure of the global supply chain network. Finally, we confirm the validity for propositions S1 (short path length), S2 (power-law degree distribution), S3 (high clustering coefficient), S4 (“fit-gets-richer” growth mechanism), S5 (truncation of power-law degree distribution), and S7 (community structure with overlapping boundaries) in the global supply chain network.

Introduction

National economies are linked by international trade and consequently economic globalization forms a giant economic complex network with strong links, i.e., interactions due to increasing trade. Especially if we view the globalized world economy with high resolution or microscopic view, we might notice that the giant economic network is a global supply chain consisting of a huge number of firms. On the other hand, it has been known that various collective motions exist in natural and social phenomena. The collective motions are due to strong interactions between constituent elements. Thus, it is expected that various collective motions will emerge in the globalized world economy under trade liberalization.

In the study of supply chain network, a several review papers have been published. M. A. Bellamy et. al. categorized the study into three themes: network structure...
(i.e., system architecture), network dynamics (i.e., system behavior), and network strategy (system policy and control). They listed important factors to characterize supply chain network. For instance, factors of network structure are node-level property, network-level property, and link-level property. The factors of network dynamics are stimuli, phenomenon, and sustainability. The factors of network strategy are scope, intent, and governance. S. Perera et. al. [2] surveyed the methodologies to model topology and robustness. They pointed out the limitation of preferential attachment growth model to explain characteristics of the supply chain network and stressed the importance of fitness based growth models [3] to explain the observed topological characteristics. Notable phenomena on the supply chain networks are not only resilience against random failure and targeted attack but also collective motion such as cascading failure or chain bankruptcy. Y. Fujiwara studied the chain bankruptcy by analyzing supply chain and bankruptcy data, and Y. Ikeda developed a agent-based model and ran realistic simulation of the chain bankruptcy caused by a failure of a single firm [4]. K. J. Mizgier et. al. [5] studied the dynamics of default process in supply chain network using a agent-based model. Based on the simulation, they discussed implication in risk management and policy making. L. Tang et. al. [6] developed a theoretical cascading failure model considering interdependence of firms in supply chain network. They observed a sudden collapse of the interdependence of supply chain network. T. Mizuno et. al. [7] have analyzed a large set of global supply chain data. They have investigated three different types of networks: a customer-supplier network, a licensee-licensor network and a strategic alliance network. The degree distributions of all these three networks show scale free properties characterized by a power law tail. They also observed that all three network shows average path length around six. They have further studied the community structure of undirected versions of the networks using modularity maximization technique [8].

In addition to these studies, E. J.S. Hearnshaw et. al. [9] have studied the supply chain network in terms of complex network approach and have proposed the following nine propositions:

- **S1**: Efficient supply chain systems show a short characteristic path length
- **S2**: The nodal degree distribution of efficient supply chain systems follows a power law as indicated by the presence of hub firms
- **S3**: Efficient supply chain systems demonstrate a high clustering coefficient
- **S4**: The growth of efficient supply chain systems follows “fit-gets-richer” mechanism
- **S5**: The power law degree distribution of efficient supply chain systems is truncated
- **S6**: The link weight distribution of efficient supply chain systems follows a power law
- **S7**: Efficient supply chain systems demonstrate a pronounced community structure with overlapping boundaries
- **S8**: The fitness of hub firms determines the resilience of supply chain systems against both random disturbances and targeted attacks
- **S9**: Resilient supply chain systems demonstrate a power law distribution for link-weights

The nine propositions are related to path length, power-law degree distribution, clustering coefficient, preferential attachment growth mechanism, truncated power-law connectivity distribution, power-law distribution of node strength, community structure with overlapping boundaries, resilience against random failure and targeted attack, core-periphery structure, respectively. They tried to explain various functions of the supply chain by the structural characteristics of supply chain network.
In order to understand the globalized world economy and to make effective policy recommendations, it is indispensable to study global supply chain, international trade, business cycle, and economic growth by analyzing global big data using network scientific methodology. In this paper, we focus on topological properties of the global supply chain network. The study on topological properties of the global supply chain network is the first step to understand the globalized world economy with a microscopic view. We study a degree distribution, hierarchical structure, and the degree-degree correlation in the global supply chain network. We uncover the community structure of the network using map equation method and characterized them according to their location and industry classification. Furthermore, the composition of communities in terms of the bow-tie components is analyzed. Finally, we investigate the validity of the nine propositions on the supply chain network [9] based on the obtained results on the topological properties of global supply chain network.

Our paper is organized as follows. In section Data, we briefly describe the global supply chain network data used in this study. The data was collected from Standard & Poor's Capital IQ platform in 2018. In section Methods, methodologies for the identification of bow-tie structure, the community detection, and the over-expression of bow-tie components are explained. In section Results, the obtained results from the analysis of the global supply chain network data: basic structural properties, bow-tie structure, community structure, and over-expression of bow-tie components are explained using figures and tables. Finally, we investigate the validity of the nine propositions based on the obtained results on the topological properties of global supply chain network. To close, this paper concludes in section Conclusions.

Data

The global supply chain data was constructed by collecting various firm data from the web site of Standard & Poor’s (S&P) Capital IQ platform in 2018. The data include firm ID, firm name, country and location of firm, primary industry, and sector as node information. The industrial classification is based on the Global Industry Classification Standard (GICS) which is developed by Morgan Stanley Capital International and S&P. We have 206 countries as the location of firms, 11 different sectors of firms, 158 primary industries as listed in Table S1-S3 of Appendix S1.

The data also include types of business relationship between supplier and customer as link information. Although the various types of business relationships that come under suppliers are supplier, creditor, franchisor, licensor, landlord, lessor, auditor, transfer agent, investor relations firm, and vendor, the majority of the relationship types are supplier and creditor. Here the supplier indicates a firm providing the products or services and the creditor indicates a private, public or institutional entity which makes funds available to others to borrow.

In Table 1, types of business relationship for all firms are summarized. We note that the links in the data set are dominated by the business relationship of supply chain. In Table 2, types of supplier for all firms are summarized. We note that the suppliers are dominated by private firms and public firms. Therefore, the entire characteristics of the data set is reflecting the nature of the global supply chain network.

The total number of firms and directed links are 437, 453 and 948, 247, respectively. Number of firms, total revenue of firms for each country is listed in Table S1 of Appendix S1. Firm distribution for different sectors are listed in Table S2 of Appendix S1. The aggregated revenue is compared with Gross Domestic Product (GDP) for each country as shown in Fig 1. This statistics provides an evidence for goodness of data coverage of our global supply chain data. The GDP data was collected from https://data.worldbank.org/ which is in public domain.
Table 1. Types of business relationship for all firms

| Relationship | #Links  | Ratio (%) |
|--------------|---------|-----------|
| Supplier     | 849,223 | 59.0      |
| Creditor     | 465,412 | 32.3      |
| Landlord     | 76,908  | 5.3       |
| Licensor     | 47,558  | 3.3       |
| **Total links** | **1,439,101** | **100.0** |

Table 2. Types of supplier for all firms

| firm type               | #firms  | Ratio (%) |
|-------------------------|---------|-----------|
| Private firm            | 830,915 | 58.6      |
| Private fund            | 32,180  | 2.3       |
| Private investment firm | 20,164  | 1.4       |
| Public firm             | 533,910 | 37.7      |
| **Total firms**         | 1,417,169 | 100.0    |

Fig 1. Total revenue generated by the firms within a country is plotted with Country’s GDP in current US $ for the year 2017.

**Methods**

**Identification of bow-tie structure**

The bow-tie structure [10] is uncovered from the GWCC based on the flow of goods and services (money flows in the opposite direction) along the directed links. The definitions of the different regions of the bow-tie structure are given as follows:

- The Giant strongly connected component (GSCC): The largest region where any two nodes are reachable through directed path.
- IN components: The nodes from which GSCC is reachable through directed paths.
- OUT components: The nodes that are reachable from the GSCC through directed paths.
- Tendrils (TE): The rest of the nodes in the GWCC.

We use breadth-first search algorithm to detect different components of bow-tie structure.
Community detection

Empirical networks are generally non-homogeneous with a high local link density. Community detection captures highly connected groups of nodes as modules. It provides a coarse-grained description of very large scale networks. Modularity maximization \[11\] is one of the popular method to detect communities. In this method, one maximizes the modularity index. Modularity is defined as the fraction of intra-community links with a subtraction of the expected fraction given a random distribution. However, this method suffers from resolution limit problem \[12\] when applied to large networks. This indicate modularity optimization fails to detect well defined small communities in large scale networks. Moreover, this technique provides similar type of partition for both undirected and directed version of a network. It can not capture the dynamic behaviours of the network.

The map equation method \[13\] detects communities using the flow dynamics of the network. We use map equation method for our analysis as it is a directed network of suppliers and customers where link represents flow of goods. This method is one of the best performing community detection techniques to detect communities in a network \[14\]. It minimizes per step average description length \[L(C)\] of a random walker on the network as defined below

\[
L(C) = q_{\rightarrow} H(C) + \sum_{i=1}^{m} p_{i} \rightarrow H(P^i).
\]  

\(q_{\rightarrow}\) and \(H(C)\) are the probability and Shannon entropy for inter community movement of the random walker respectively. \(p_{i} \rightarrow\) is the probability that the random walker leaves the node \(i\), and \(H(P^i)\) is the entropy for intra community movement.

Over-expression of node attributes within communities

Communities are ubiquitous in empirical networks. These communities are formed based on the similarities in some attributes of nodes. For examples, locations and sectors are key attributes for the formation of communities in Japanese supply chain network \[15,16\], in protein-protein interaction networks, biological functions form the basis of community structure \[17\].

To measure the over-expression of attributes in a community we follow the method of Tumminello et. al. \[18\]. In this method, the probability that \(X\) randomly selected nodes in a community \(C\) of size \(N_C\) has the attribute \(A\) is calculated by the following hyper-geometric distribution

\[
H(X|N, N_C, N_A) = \frac{\binom{N_C}{X} \binom{N-N_C}{N-N_C-X}}{\binom{N}{N}}.
\]

where \(N_A\) is the total number of nodes in the network with attribute \(A\). The \(p\)-value \(p(N_C,A)\) for the \(N_C,A\) nodes with attribute \(A\) in the community \(C\) can be obtained from the following expression:

\[
p(N_C,A) = 1 - \sum_{X=0}^{N_C,A-1} H(X|N, N_C, N_A).
\]

The attribute \(A\) is over-expressed when \(p(N_C,A)\) is lower than the some threshold value \(p_c\). As it is a multiple-hypothesis test, one has to choose the \(p_c\) appropriately to exclude false positive. We set \(p_c = 0.01/N_A\) as used in \[18\], which takes care of the Bonferroni correction \[19\]. Here, \(N_A\) indicates total number of distinct attributes for all the nodes of the network.
Results

Basic structural properties

As the supply chain network is directed in nature, one can define in and out degrees for the nodes. The nodal in-degree is defined as the number of incoming links to a node and out-degree is the total number of outgoing links from that node. We observe probability density distributions for both nodal in and out degree’s have a heavy tail nature where the tail of the distributions is characterized by a power law of the form $P(k_{in/out}) \sim k^{-\gamma_{in/out}}$ with $\gamma_{in} = 2.42$ and $\gamma_{out} = 2.11$ respectively as shown in Fig. 2 (a-b). The power law tail of the degree distribution is also observed in past investigations of empirical supply chain network data [7, 20–22]. The degree distribution plays pivotal role in shock propagation among nodes. The high asymmetry in degree distribution can result in system wide aggregate fluctuation due to idiosyncratic shocks to large firms [23]. It has been argued in the literature that such heavy tail distribution of nodal degrees arises due to rich-get-richer mechanism [24,25]. Similar to the rich-get-richer principle, here large firms have more customers and suppliers than small firms.

![Fig 2. Structural properties of the global supply chain network.](image)

Probability density distributions $P$ of (a) the nodal in-degrees $k_{in}$ and (b) the nodal out-degrees $k_{out}$. Variation of (c) the clustering coefficient $C(k)$ as a function of degree $k$ and (d) the average nearest neighbor degree $\langle k_{nn}(k) \rangle$ as a function of degree $k$. Logarithmic binning of the horizontal axis is used in (a) and (b). Red lines represent the best power-law fit to the data. Blue lines in (c) and (d) represent the results for degree preserved random network where average is taken over 100 such uncorrelated networks.

Clustering coefficient, a measure of three-point correlation, reflects cliquishness among the neighbours of a nodes. For most of the real world network, average clustering coefficient is a decaying function of degree having a form $\langle C(k) \rangle \sim k^{-\beta_k}$ with $\beta_k \leq 1.0$. We observe the clustering coefficient in the supply chain network decays with an exponent $\beta_k = 0.46$ as shown in Fig. 2 (c) indicates the presence of a hierarchical
The average degree of the neighbors of a node $i$ which capture the nodal degree-degree correlation is defined as $k_{nn,i} = \sum_j k_j/k_i$ where the $j$ runs over all $k_i$ neighbors of $i$. For the nodes with degree $k$, \( \langle k_{nn}(k) \rangle = \sum_{k_i=k} k_{nn,i}/N_k = \sum_{k_i} k_1 P(k_1|k) \) where $N_k$ is the number of nodes having degree $k$. The $\langle k_{nn}(k) \rangle$ increases with $k$ for an assorative network and decreases for a disassortative network. In the absence of nodal degree-degree correlation $\langle k_{nn}(k) \rangle$ remain constant. As can be seen from Fig. 2 (d), $\langle k_{nn}(k) \rangle$ does not depend on $k$ and remain more or less in constant with $k$, indicating the absence of nodal degree-degree correlation. Further, the statistical significance of these results are tested by comparing it with results of the randomized degree preserving network [26]. The clustering coefficients of randomized network shows $C(k) \sim$ constant as expected. The variation of $\langle k_{nn}(k) \rangle$ with $k$ matches nicely with the case of degree preserving randomized network, which further supports the absence of nodal degree-degree correlation in the empirical network.

We study the connected components when the network is viewed as an undirected network. The largest connected component of the network is known as the Giant weakly connected component (GWCC). As can be seen from Fig. 3, the network consists of a very large GWCC with $N = 407,527$ nodes and $L = 927,316$ links. Using a breadth-first search, we calculate the average path length in the GWCC, by calculating the shortest paths between all pairs of nodes. The average path length is found to be 5.370 reflecting the small world nature of the global supply chain network. While the GWCC contains 93.16% of nodes of the network, rest of the components are very small. In the subsequent sections, we investigate only the GWCC of the network.

**Fig 3. Distribution $n_x$ of the component sizes $x$ in the network. The largest weakly connected component contains $\sim 99\%$ of the nodes in the entire network.**

### Bow-tie structure

We detect the bow-tie components in GWCC of the global supply chain network. The number of firms in each component is shown in Table 3. The OUT component, consists of nodes from which GSCC is reachable through directed paths toward downstream, is the largest and it consists 41.1% of total firms. GSCC (any two nodes are reachable through directed path), IN (nodes from which GSCC is reachable through directed paths toward upstream), and TE (The rest of the nodes in the GWCC) are approximately similar in size and comprise 16.4%, 22.3%, and 20.2% of total firms.
respectively. For Japanese supply chain network, the fraction of each component of the OUT, GSCC, IN, and TE is 26.2%, 49.7%, 20.6%, and 3.5%, respectively [15]. The GSCC in the Japanese supply chain network occupies half of the system, meaning that most firms are interconnected by the small geodesic distances or the shortest-path lengths in the economy. This shows a good contrast to the result of the global supply chain network observed in our study. However, by examining the shortest-path lengths from GSCC to IN and OUT as shown in Table 4, one can observe that the firms in the upstream or downstream sides are mostly located a few step away from the GSCC. This feature of the economic network is different from the bow-tie structure of many other complex networks [27].

Table 3. Bow-tie structure: Sizes of different components

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

| Component | #firms | Ratio (%) |
|-----------|--------|-----------|
| GSCC      | 66,798 | 16.4      |
| IN        | 90,992 | 22.3      |
| OUT       | 167,509| 41.1      |
| TE        | 82,228 | 20.2      |
| Total     | 407,527| 100       |

Table 4. Shortest distances from GSCC to IN/OUT

| Distance | IN to GSCC | Ratio (%) | OUT to GSCC | Ratio (%) |
|----------|------------|-----------|-------------|-----------|
| 1        | 82,761     | 90.954    | 1           | 153,755   | 91.789 |
| 2        | 7,430      | 8.165     | 2           | 11,885    | 7.095  |
| 3        | 665        | 0.731     | 3           | 1,582     | 0.944  |
| 4        | 104        | 0.114     | 4           | 250       | 0.149  |
| 5        | 17         | 0.019     | 5           | 26        | 0.015  |
| 6        | 10         | 0.011     | 6           | 10        | 0.006  |
| 7        | 5          | 0.005     | 7           | 1         | 0.001  |
| Total    | 90,992     | 100       | Total       | 167,509   | 100    |

Community structure

Communities are detected in the largest weakly connected component of the network. We employ the map equation method [13] to uncover the communities in the GWCC of the global supply chain network. The detected communities are found in various sizes. The probability density distributions $D(s)$ of community sizes $s$ for the empirical network and its degree preserving randomized network are shown in Fig. 4 (a). The distribution for the empirical network is more wider than it is for the randomized network.

The biased in the direction of flow between a pair of communities is measured by the polarization ratio defined by $P_{ij} = |w_{ij} - w_{ji}|/(w_{ij} + w_{ji})$, where $w_{ij}$ is the total number of links from $i$-th community to $j$-th community. $P_{ij} = 1$ if the flow is totally biased from one community to the other and $P_{ij} = 0$ if the flow is evenly balanced between the communities. The total flow between a pair of communities is $L_{ij} = (w_{ij} + w_{ji})$. If we assume that there is no bias in the flow direction between any pair of communities, according to a null hypothesis, the values of $P_{ij}$ will fluctuates
around 0 with the standard deviation \( \sigma = 1/\sqrt{L_{ij}} \). As can be seen from Fig. 4 (b), most of the values for the polarizability ratio \( P_{ij} \) are significantly higher than the 2\( \sigma \) level which is indicated by the dashed curve.

![Fig 4. Community sizes and polarizability](image)

**Overexpression within communities**

We study the significant overexpression of different attributes such as primary industry, sectors, firm’s location, bow-tie components within the communities. We have shown the detail overexpression results within 10 largest communities in Table 5. Various interesting features can be observed from the results of attribute overexpression. The largest community comprises of consumer discretionary sector based in the US. Further analysis shows these are private firms mainly from automotive retail, which belong to the OUT component in the bow-tie structure of the global supply chain network. In the second largest community, we observe of consumer discretionary sector based in China, UK, France, Germany, Japan, Malaysia and New Zealand. These firms belong to the IN component of bow-tie structure. The firms of third largest communities are from consumer discretionary, industrials, and materials sectors which are mainly based in Japan, China and Thailand. These firms are mostly belonged to TE component of bow-tie structure.

We construct a weighted and undirected network of countries from their overexpression in communities with size larger than 100 to show the inter-relation between countries. A link of weight 1 is placed between two countries if they over-express simultaneously within a community. Furthermore, we visualize community structure of this network as shown in Fig 5. It shows each community is formed by geographically closely located countries.

Similarly, we also constructed a weighted undirected network of over-expressed primary industries, where a link of weight \( w \) is present between two primary industries, if they are over-expressed simultaneously in \( w \) communities. As can be seen from Fig A and Fig S1 of Appendix S1, the clusters among primary industries are formed based on their sector classification.

We show the frequency of over-expression of the different components within the communities in bow-tie structure in Fig. 7. Here, we selected communities which size of communities is at least 10 firms. G-I indicates both GCC and IN components are overexpressed in the communities. Similarly, G-O, G-T, I-O, I-T and O-T represent overexpression of GCC-OUT, GCC-TE, IN-OUT, IN-TE, and OUT-TE respectively. It reflects that most of the communities are solely composed of a particular component of the bow-tie structure. We also observe there are reasonable number of communities
Table 5. Brief summary on the over-expression of sectors, countries and bow-tie components in the ten largest communities

| Rank | Size  | Sector                      | Country                        | Bow-tie components |
|------|-------|-----------------------------|--------------------------------|--------------------|
| 1    | 1,687 | Consumer Discretionary      | US (77.8%)                     | OUT (99.7%)        |
| 2    | 1,632 | Consumer Discretionary      | China (9.1%), UK (4.0%),      | IN (32.2%)         |
|      |       |                             | France (3.7%), Germany (6.2%),|                    |
|      |       |                             | Japan (9.1%), Malaysia (3.6%),|                    |
|      |       |                             | New Zealand (1.3%)            |                    |
| 3    | 1,179 | Consumer Discretionary      | China (12.4%), Japan (53.1%), | TE (99.6%)         |
|      |       | (20.0%), Industrials (23.1%),| Thailand (9.3%)               |                    |
|      |       | Materials (15.8%)           |                                |                    |
| 4    | 1,027 | Communication Services      | Cambodia (0.6%), Indonesia (66.7%), | GSCC (20.3%), IN (55.1%) |
|      |       |                             | Singapore (3.9%)               |                    |
| 5    | 968   | Financials (62.1%)          | Bahrain (0.7%), Bangladesh (1.8%), | GSCC (22.7%), IN (63.1%) |
|      |       |                             | Hong Kong (5.0%), Hungary (1.0%), |                    |
|      |       |                             | Italy (4.0%), Nepal (1.0%),   |                    |
|      |       |                             | Pakistan (14.0%), Singapore (3.8%), |                    |
|      |       |                             | UAE (2.4%), Yemen (0.4%)      |                    |
| 6    | 933   | Industrials (59.1%)         | US (93.3%)                     | IN (82.3%)         |
| 7    | 867   | Information Technology      | Korea South (3.1%), Singapore (4.4%), Taiwan (30.8%) | GSCC (23.9%), IN (26.6%) |
| 8    | 824   | Consumer Discretionary      | US (52.4%)                     | GSCC (29.5%), IN (28.9%) |
|      |       | (20.0%), Industrials (38.4%)|                                |                    |
| 9    | 723   | Energy (12.3%), Industrials (17.6%), Materials (8.4%), Utilities (7.5%) | India (62.8%) | GSCC (28.5%) |
| 10   | 711   | -                           | Botswana (1%), South Africa (69.5%) | OUT (98.5%) |

are composed by the combination of GSCC and IN (G-I component), which is also observed in Japanese supply chain network [15]. This indicate the flow of goods in the supply chain network is more often confined within the GSCC and IN component compared to any other combination of the components of bow-tie structure. Surprisingly, a large fraction of communities are located in TE component. The firms in the communities located in TE components not only supply but also procure any products and services from GSCC components.

Discussion on the nine propositions

E. J.S. Hearnshaw et. al. [9] have studied the supply chain network in terms of complex network approach and have proposed the nine propositions. In this section, we investigate the validity of the nine propositions based on the obtained results on the topological properties of global supply chain network.

Proposition S1

S1: Efficient supply chain systems demonstrate a short characteristic path length

The average path length in the GWCC of the global supply chain was found to be 5.370. The average path length in the small world network $L_s$ is known to be similar to the average path length in the random graph $L_r$. The average path length in the random graph $L_r$ is approximately calculated by $L_r = \log N / \log <k> = 6.77$. Here, the number of nodes in the GWCC is $N = 407527$, the average degree is $<k> = (<k_{in}> + <k_{out}>)/2 = 6.74$. By assuming the degree distributions to be a power-law distributions in the entire range of the degree with $\gamma_{in} = 2.42$ and $\gamma_{out} = 2.11$. The average in-degree and the average out-degree are calculated by
Fig 5. Overexpression network of countries. Different node color indicates different communities of the network. Here, communities are detected using modularity maximization technique.

\[
<k_{\text{in}} > = k_{\text{min}}^{\text{in}} (\gamma_{\text{in}} - 1)/(\gamma_{\text{in}} - 2) = 3.38, \quad \text{and} \quad <k_{\text{out}} > = k_{\text{min}}^{\text{out}} (\gamma_{\text{out}} - 1)/(\gamma_{\text{out}} - 2) = 10.0, \quad \text{where} \quad k_{\text{min}}^{\text{in}} = 1 \quad \text{and} \quad k_{\text{min}}^{\text{out}} = 1.
\]

The estimated value of \(L_r = 6.77\) is close to the observed value 5.370. This is reflecting the small world nature of the global supply chain network. Therefore the estimation of the average path length validate the proposition S1.

**Propositions S2**

**S2: The nodal degree distribution of efficient supply chain systems follows a power law as indicated by the presence of hub firms**

We observe probability density distributions for both nodal in and out degree’s have a heavy tail nature where the tail of the distributions is characterized by a power law of the form \(P(k_{\text{in/out}}) \sim k^{-\gamma_{\text{in/out}}}\) with \(\gamma_{\text{in}} = 2.42\) and \(\gamma_{\text{out}} = 2.11\) respectively as shown in Fig. 2 (a-b). The network whose degree distribution is characterized by a power law possess hub firms. The hub firms are known as channel leader firms which are said to control performance and provide system-wide coordination of the supply chain [28,29]. The channel leader firms can exert their influence and provide opportunities and motivation for other firms to align themselves with their own specific objectives [30].
Fig 6. Overexpression network of primary industries. Different node color indicates different communities of the network. Here, communities are detected using modularity maximization technique. IDs of the nodes are given in Table S3 of Appendix S1.

Fig 7. Frequency of overexpression of bow-tie components within the communities with size at least 10. G-I indicates both GSCC and IN components are overexpressed in the communities. Similarly, G-O, G-T, I-O, I-T and O-T represent overexpression of GSCC-OUT, GSCC-TE, IN-OUT, IN-TE, and OUT-TE respectively.
The power law distributions characterized with $\gamma_{in} = 2.42$ and $\gamma_{out} = 2.11$ validate the proposition S2.

Propositions S3

S3: Efficient supply chain systems demonstrate a high clustering coefficient

Clustering coefficient, a measure of three-point correlation, reflects cliquishness among the neighbours of a node. For most of the real world network, average clustering coefficient is a decaying function of degree having a form $\langle C(k) \rangle \sim k^{-\beta_k}$ with $\beta_k \leq 1.0$. We observe the clustering coefficient in the supply chain network decays with an exponent $\beta_k = 0.46$ as shown in Fig. 2 (c) indicates the presence of a hierarchical structure. The observed moderate clustering coefficient indicates that the proposition S3: It has a high clustering coefficient is weakly valid.

Propositions S4

S4: The growth of efficient supply chain systems follows “fit-gets-richer” mechanism

It has been argued in the literature that such heavy tail distribution of nodal degrees arises due to rich-get-richer mechanism [24][25]. Similar to the rich-get-richer principle, here large firms have more customers and suppliers than small firms. Preferential attachment in rich-get-richer mechanism assumes that the acquisition of new links by a firm is determined solely by the number of its existing links. This assumption leads to the number of links being proportional to their duration in the supply chain. However, one can often observe that older firms have been outstripped by new entrant firms. There is a need therefore, to include the “fitness” of the firms to account for new entrants that can quickly dominate supply chains. By introducing “fit-gets-richer” mechanism [31], the fitter nodes have a greater acquisition rate for links and therefore, resulting network possess a scale-free property. The heavy tail distribution of nodal degrees and overtaking of older firms by new entrant firms validate the proposition S4.

Propositions S5

S5: The power law degree distribution of efficient supply chain system is truncated

The power law distributions $P(k_{in/out}) \sim k^{-\gamma_{in/out}}$ with $\gamma_{in} = 2.42$ and $\gamma_{out} = 2.11$ respectively are observed in the middle region of the distributions as shown in Fig. 2 (a-b). The tail region of both distributions seem like truncated or exponentially cut-off. Especially this tendency is evident for $P(k_{out})$. This phenomenon is said to be caused by four reasons [9]. First, the finite size of marketplaces generates a truncated power law degree distribution. Second, there are practical reasons in the operation of firms that limit the ability of firms to indefinitely form and maintain exchange relationships. Third, when new links are to be formed with a hub firms, incomplete information generates uncertainty which might costs higher than transaction costs. If these costs are unacceptable, the firms will scrap the deal with the hub firms. Finally, the aging and depreciation of firms limits their growth. The observed truncation or cut-off in the tail region of the degree distribution validates the proposition S5.

Propositions S7

S7: Efficient supply chain systems demonstrate a pronounced community structure with overlapping boundaries

We employ the map equation method [13] to uncover the communities in the GWCC of the global supply chain network. The detected communities are found in various sizes. The probability density distributions $D(s)$ of community sizes $s$ for the empirical
network and its degree preserving randomized network are shown in Fig 4 (a). The distribution for the empirical network is more wider than it is for the randomized network. In Table 5 the over-expression of sectors and countries in the ten largest communities is shown. Communities in a supply chain are bound together in clusters predominantly connected by horizontal relationships among firms with similar interests and functions. However, we empirically observed that all firms within a community are not entirely cooperative as shown in Table 5. Therefore, community formation in supply chain possess overlapping boundaries. These results validates the proposition S7.

Remaining Propositions

The supply chain data has no weight on links. Therefore the following two hypotheses:
S6: The link weight distribution of efficient supply chain systems follows a power law,
S9: Resilient supply chain systems demonstrate a power law distribution for link-weights
are not applicable in the analyses of this paper. In addition, we concentrated on the topological properties of the supply chain network and therefore, the resilience of the system: S8: The fitness of hub firms determines the resilience of supply chain systems against both random disturbances and targeted attacks is out of scope of our current study.

Conclusions

We studied on topological properties of global supply chain network in terms of degree distribution, hierarchical structure, and degree-degree correlation in the global supply chain network. The global supply chain data was constructed by collecting various company data from the website of Standard & Poor’s Capital IQ platform in 2018. The total number of firms and directed links in our data were 437, 453 and 948, 247, respectively.

The degree distributions is characterized by a power law of the form with $\gamma_{in} = 2.42$ and $\gamma_{out} = 2.11$. The clustering coefficient decays $\langle C(k) \rangle \sim k^{-\beta_k}$ with an exponent $\beta_k = 0.46$. This indicates the presence of a hierarchical structure of the supply chain network. We observed that $\langle k_{nn}(k) \rangle$ does not depend on $k$ and remain more or less in constant with $k$, indicating the absence of nodal degree-degree correlation. The Bow-tie structure of GWCC revealed that the OUT component was the largest and it consists 41.1% of total firms. The GSCC component comprised 16.4% of total firms. We observed that the firms in the upstream or downstream sides were mostly located a few step away from the GSCC.

Furthermore, we uncovered the community structure of the network using map equation method and characterized them according to their location and industry classification. We observed that the largest community comprises of private firms mainly from automotive retail based in the US. These firms are belong to the OUT component in the bow-tie structure of the global supply chain network. It indicates the retail firms are generally belong to the OUT component of bow-ties structure.

Finally, we investigated the validity of the nine propositions on the supply chain network based on the obtained results on the topological properties. We confirmed the validity of propositions S1 (short path length), S2 (power-law degree distribution), S3 (high clustering coefficient), S4 (“fit-gets-richer” growth mechanism), S5 (truncation of power-law degree distribution), and S7 (community structure with overlapping boundaries) in the global supply chain network. However, the propositions related to link weight and resilient nature of the network were not confirmed due to the limitation of our data and the scope of our current study. This will be left for future study.
Our study provides a detailed topological characterization of the global supply chain network. These topological properties are utmost important to understand the international trade dynamics. It is well-known that community structure plays an important role in spreading phenomena. Our characterization of community structure will be helpful to understand the wide-spread economic crisis. The study further shows the inter-relationships among the countries and among the industrial sectors.

Acknowledgments

We are grateful to Y. Fujiwara, H. Aoyama, H. Iyetomi, W. Souma, and H. Yoshikawa for their insightful comments and encouragement. The present study was supported by the Ministry of Education, Science, Sports, and Culture, Grants-in-Aid for Scientific Research (B), Grant no. 17KT0034 (2017-2019) and Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena).

References

1. Bellamy MA, Basole RC. Network analysis of supply chain systems: A systematic review and future research. Systems Engineering. 2012;16(2):235–249. doi:10.1002/sys.21238.

2. Perera S, Bell MGH, Bliemer MCJ. Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective. Applied Network Science. 2017;2(1). doi:10.1007/s41109-017-0053-0.

3. Bell M, Perera S, Piraveenan M, Bliemer M, Latty T, Reid C. Network growth models: A behavioural basis for attachment proportional to fitness. Scientific Reports. 2017;7(1). doi:10.1038/srep42431.

4. Fujiwara Y, Ikeda Y. 6. In: Econophysics and Companies -Statistical Life and Death in Complex Business Networks-. Cambridge CB2 8RU, UK: Cambridge University Press; 2010. p. 184–220.

5. Mizgier KJ, Wagner SM, Holyst JA. Modeling defaults of companies in multi-stage supply chain networks. International Journal of Production Economics. 2012;135(1):14–23. doi:10.1016/j.ijpe.2010.09.022.

6. Tang L, Jing K, He J, Stanley HE. Complex interdependent supply chain networks: Cascading failure and robustness. Physica A: Statistical Mechanics and its Applications. 2016;443:58–69. doi:10.1016/j.physa.2015.09.082.

7. Mizuno T, Ohnishi T, Watanabe T. Structure of global buyer-supplier networks and its implications for conflict minerals regulations. EPJ Data Science. 2016;5(1):2.

8. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Physical review E. 2004;70(6):066111.

9. Hearnshaw EJS, Wilson MMJ. A complex network approach to supply chain network theory. International Journal of Operations & Production Management. 2013;33(4):442–469. doi:10.1108/0143571311307343.

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

12. Fortunato S, Barthelemy M. Resolution limit in community detection. Proceedings of the national academy of sciences. 2007;104(1):36–41.

13. 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.

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

15. Chakraborty A, Kichikawa Y, Iino T, Iyetomi H, Inoue H, Fujiwara Y, et al. Hierarchical communities in the walnut structure of the Japanese production network. PloS one. 2018;13(8).

16. Chakraborty A, Krichene H, Inoue H, Fujiwara Y. Characterization of the community structure in a large-scale production network in Japan. Physica A: Statistical Mechanics and its Applications. 2019;513:210–221.

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

18. Tumminello M, Micciche 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.

19. Miller RG. Normal univariate techniques. In: Simultaneous statistical inference. Springer; 1981. p. 37–108.

20. Sun H, Wu J. Scale-free characteristics of supply chain distribution networks. Modern physics letters B. 2005;19(17):841–848.

21. Fujiwara Y, Aoyama H. Large-scale structure of a nation-wide production network. The European Physical Journal B. 2010;77(4):565–580.

22. Basole RC. Topological analysis and visualization of interfirm collaboration networks in the electronics industry. Decision Support Systems. 2016;83:22–31.

23. Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. The network origins of aggregate fluctuations. Econometrica. 2012;80(5):1977–2016.

24. Barabási AL, Albert R. Emergence of scaling in random networks. Science. 1999;286(5439):509–512.

25. Chakraborty A, Manna S. Weighted trade network in a model of preferential bipartite transactions. Physical Review E. 2010;81(1):016111.

26. Newman M. Networks: an introduction. Oxford university press; 2010.

27. 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.

28. Jarillo C. On strategic networks. Strategic Management Journal. 1988;9(1):31–41.

29. Human SE, Provan KG. Legitimacy building in the evolution of small firm multilateral networks: a comparative study of success and demise. Administrative Science Quarterly. 2000;45(2):327–65.
30. Cowan R, Jonard N. Structural holes, innovation and the distribution of ideas. Journal of Economic Interaction and Coordination. 2007;2(2):93–110.

31. Bianconi G, Barabási AL. Competition and multiscaling in evolving networks. EPL (Europhysics Letters). 2001;54(4):436.

Supporting information

S1 Appendix. Appendix to the manuscript.
Table A represents the code of countries, firm distribution in different countries, total revenue of the firms in each country and countries’ GDP (current US$) for the year 2017.

Table B represents the industrial sectors and firm distribution.

Table C shows the primary industry and sector classification.
Table A. Countries and firm distribution. "-" in some fields indicates missing information.

| Sl. no. | Name of the sector                  | No. of firms |
|---------|-------------------------------------|--------------|
| 1       | Communication Services              | 17491        |
| 2       | Consumer Discretionary              | 35283        |
| 3       | Consumer Staples                    | 13044        |
| 4       | Energy                              | 9542         |
| 5       | Financials                          | 25953        |
| 6       | Health Care                         | 20979        |
| 7       | Industrials                         | 54819        |
| 8       | Information Technology              | 34648        |
| 9       | Materials                           | 18019        |
| 10      | Real Estate                         | 7555         |
| 11      | Utilities                           | 7017         |
| 12      | Unknown                             | 193103       |

Table B. Sectors and firm distribution

| Sl. no. | Name of the sector                  | No. of firms |
|---------|-------------------------------------|--------------|
| 1       | Communication Services              | 17491        |
| 2       | Consumer Discretionary              | 35283        |
| 3       | Consumer Staples                    | 13044        |
| 4       | Energy                              | 9542         |
| 5       | Financials                          | 25953        |
| 6       | Health Care                         | 20979        |
| 7       | Industrials                         | 54819        |
| 8       | Information Technology              | 34648        |
| 9       | Materials                           | 18019        |
| 10      | Real Estate                         | 7555         |
| 11      | Utilities                           | 7017         |
| 12      | Unknown                             | 193103       |
| ID | Primary industry                        | Sector                      |
|----|----------------------------------------|-----------------------------|
| 1  | Advertising                            | Communication Services      |
| 2  | Aerospace and Defense                  | Industrials                 |
| 3  | Agricultural Products                  | Consumer Staples            |
| 4  | Agricultural and Farm Machinery        | Industrials                 |
| 5  | Air Freight and Logistics              | Industrials                 |
| 6  | Airlines                               | Industrials                 |
| 7  | Airport Services                       | Communication Services      |
| 8  | Alternative Carriers                   | Materials                   |
| 9  | Aluminum                               | Consumer Discretionary      |
| 10 | Apparel Retail                         | Consumer Discretionary      |
| 11 | Apparel, Accessories and Luxury Goods  | Consumer Discretionary      |
| 12 | Application Software                   | Information Technology      |
| 13 | Asset Management and Custody Banks     | Financials                 |
| 14 | Auto Parts and Equipment               | Consumer Discretionary      |
| 15 | Automobile Manufacturers               | Consumer Discretionary      |
| 16 | Automotive Retail                      | Consumer Discretionary      |
| 17 | Biotechnology                          | Health Care                 |
| 18 | Brewers                                | Consumer Staples            |
| 19 | Broadcasting                           | Communication Services      |
| 20 | Building Products                      | Industrials                 |
| 21 | Cable and Satellite                    | Communication Services      |
| 22 | Casinos and Gaming                     | Consumer Discretionary      |
| 23 | Coal and Consumable Fuels              | Energy                      |
| 24 | Commercial Printing                    | Industrials                 |
| 25 | Commodity Chemicals                    | Materials                   |
| 26 | Communications Equipment               | Information Technology      |
| 27 | Computer and Electronics Retail        | Consumer Discretionary      |
| 28 | Construction Machinery and Heavy Trucks| Industrials                 |
| 29 | Construction Materials                 | Materials                   |
| 30 | Construction and Engineering           | Industrials                 |
| 31 | Consumer Electronics                   | Consumer Discretionary      |
| 32 | Consumer Finance                       | Financials                 |
| 33 | Copper                                 | Materials                   |
| 34 | Data Processing and Outsourced Services| Information Technology      |
| 35 | Department Stores                      | Consumer Discretionary      |
| 36 | Distillers and Vintners                | Consumer Staples            |
| 37 | Distributors                           | Consumer Discretionary      |
| 38 | Diversified Banks                      | Financials                 |
| 39 | Diversified Capital Markets            | Financials                 |
| 40 | Diversified Chemicals                  | Materials                   |
| 41 | Diversified Metals and Mining          | Materials                   |
| 42 | Diversified REITs                      | Real Estate                 |
| 43 | Diversified Real Estate Activities     | Real Estate                 |
| 44 | Diversified Support Services           | Industrials                 |
| 45 | Drug Retail                            | Consumer Staples            |
| 46 | Education Services                     | Consumer Discretionary      |
| 47 | Electric Utilities                     | Utilities                   |
| 48 | Electrical Components and Equipment    | Industrials                 |
| 49 | Electronic Components                  | Information Technology      |
| 50 | Electronic Equipment and Instruments   | Information Technology      |
| 51 | Electronic Manufacturing Services      | Information Technology      |
| Number | Sector 1                                      | Sector 2          |
|--------|---------------------------------------------|-------------------|
| 52     | Environmental and Facilities Services       | Industrials       |
| 53     | Fertilizers and Agricultural Chemicals      | Materials         |
| 54     | Financial Exchanges and Data                | Financials        |
| 55     | Food Distributors                           | Consumer Staples  |
| 56     | Food Retail                                 | Consumer Staples  |
| 57     | Footwear                                    | Consumer Discretionary |
| 58     | Forest Products                             | Materials         |
| 59     | Gas Utilities                               | Utilities         |
| 60     | General Merchandise Stores                  | Consumer Discretionary |
| 61     | Gold                                        | Materials         |
| 62     | Health Care Distributors                    | Health Care       |
| 63     | Health Care Equipment                       | Health Care       |
| 64     | Health Care Facilities                      | Health Care       |
| 65     | Health Care REITs                           | Real Estate       |
| 66     | Health Care Services                        | Health Care       |
| 67     | Health Care Supplies                        | Health Care       |
| 68     | Health Care Technology                      | Health Care       |
| 69     | Heavy Electrical Equipment                  | Industrials       |
| 70     | Highways and Railtracks                     | Industrials       |
| 71     | Home Furnishings                            | Consumer Discretionary |
| 72     | Home Improvement Retail                     | Consumer Discretionary |
| 73     | Homebuilding                                | Consumer Discretionary |
| 74     | Homefurnishing Retail                       | Consumer Discretionary |
| 75     | Hotel and Resort REITs                      | Real Estate       |
| 76     | Hotels, Resorts and Cruise Lines            | Consumer Discretionary |
| 77     | Household Appliances                        | Consumer Discretionary |
| 78     | Household Products                          | Consumer Staples  |
| 79     | Housewares and Specialties                  | Consumer Discretionary |
| 80     | Human Resource and Employment Services      | Industrials       |
| 81     | Hypermarkets and Super Centers              | Consumer Staples  |
| 82     | IT Consulting and Other Services            | Information Technology |
| 83     | Independent Power Producers and Energy Traders | Utilities         |
| 84     | Industrial Conglomerates                    | Industrials       |
| 85     | Industrial Gases                            | Materials         |
| 86     | Industrial Machinery                        | Industrials       |
| 87     | Industrial REITs                            | Real Estate       |
| 88     | Insurance Brokers                           | Financials        |
| 89     | Integrated Oil and Gas                      | Energy            |
| 90     | Integrated Telecommunication Services       | Communication Services |
| 91     | Interactive Home Entertainment               | Communication Services |
| 92     | Interactive Media and Services              | Communication Services |
| 93     | Internet Services and Infrastructure        | Information Technology |
| 94     | Internet and Direct Marketing Retail        | Consumer Discretionary |
| 95     | Investment Banking and Brokerage            | Financials        |
| 96     | Leisure Facilities                          | Consumer Discretionary |
| 97     | Leisure Products                            | Consumer Discretionary |
| 98     | Life Sciences Tools and Services            | Health Care       |
| 99     | Life and Health Insurance                   | Financials        |
| 100    | Managed Health Care                         | Health Care       |
| 101    | Marine                                      | Industrials       |
| 102    | Marine Ports and Services                   | Industrials       |
| 103    | Metal and Glass Containers                  | Materials         |

March 6, 2020
| 104 | Mortgage REITs | Financials |
|-----|----------------|------------|
| 105 | Motorcycle Manufacturers | Consumer Discretionary |
| 106 | Movies and Entertainment | Communication Services |
| 107 | Multi-Sector Holdings | Financials |
| 108 | Multi-Utilities | Utilities |
| 109 | Multi-line Insurance | Real Estate |
| 110 | Office REITs | Industrials |
| 111 | Office Services and Supplies | Energy |
| 112 | Oil and Gas Drilling | Energy |
| 113 | Oil and Gas Equipment and Services | Energy |
| 114 | Oil and Gas Exploration and Production | Energy |
| 115 | Oil and Gas Refining and Marketing | Energy |
| 116 | Oil and Gas Storage and Transportation | Financials |
| 117 | Other Diversified Financial Services | Consumer Staples |
| 118 | Packaged Foods and Meats | Industrials |
| 119 | Paper Packaging | Real Estate |
| 120 | Paper Products | Real Estate |
| 121 | Personal Products | Real Estate |
| 122 | Pharmaceuticals | Financials |
| 123 | Precious Metals and Minerals | Utilities |
| 124 | Property and Casualty Insurance | Real Estate |
| 125 | Publishing | Consumer Discretionary |
| 126 | Railroads | Real Estate |
| 127 | Real Estate Development | Real Estate |
| 128 | Real Estate Operating Companies | Industrials |
| 129 | Real Estate Services | Information Technology |
| 130 | Regional Banks | Information Technology |
| 131 | Reinsurance | Materials |
| 132 | Renewable Electricity | Consumer Staples |
| 133 | Research and Consulting Services | Consumer Discretionary |
| 134 | Residential REITs | Financials |
| 135 | Restaurants | Real Estate |
| 136 | Retail REITs | Real Estate |
| 137 | Security and Alarm Services | Industrials |
| 138 | Semiconductor Equipment | Information Technology |
| 139 | Semiconductors | Information Technology |
| 140 | Silver | Materials |
| 141 | Soft Drinks | Consumer Staples |
| 142 | Specialized Consumer Services | Consumer Discretionary |
| 143 | Specialized Finance | Financials |
| 144 | Specialized REITs | Real Estate |
| 145 | Specialty Chemicals | Materials |
| 146 | Specialty Stores | Consumer Discretionary |
| 147 | Steel | Materials |
| 148 | Systems Software | Information Technology |
| 149 | Technology Distributors | Information Technology |
| 150 | Technology Hardware, Storage and Peripherals | Information Technology |
| 151 | Textiles | Consumer Discretionary |
| 152 | Thrifts and Mortgage Finance | Financials |
| 153 | Tires and Rubber | Consumer Discretionary |
| 154 | Tobacco | Consumer Staples |
| 155 | Trading Companies and Distributors | Industrials |
Table C. Primary industry and sector classification

We show a different color code of the nodes for the overexpression network of primary industries, which is shown in Fig 6 in main text. Here we use the node color according to their sector classification. From Fig 6 of main text and Fig A, we observe the clustering among primary industries are formed based their sectors.

Fig A. Overexpression network of primary industries. Different node color indicates different sectors. IDs of the nodes are given in Table C.