

RESEARCH

Open Access



Tie-formation process within the communities of the Japanese production network: application of an exponential random graph model

Hazem Krichene^{1*} , Abhijit Chakraborty¹, Yoshi Fujiwara¹, Hiroyasu Inoue¹ and Masaaki Terai²

*Correspondence:

krichene.hazem@gmail.com

¹Graduate School of Simulation Studies, University of Hyogo, Kobe, Japan

Full list of author information is available at the end of the article

Abstract

This paper studies the driving forces behind the formation of ties within the major communities in the Japanese nationwide network of production, which contains one million firms and five million links between suppliers ("upstream" firms) and customers ("downstream" firms). We apply the Infomap algorithm to reveal the hierarchical structure of the production network. At the second level of the hierarchy, we find a reasonable community resolution, where the community size distribution follows a power law decay. Then, we estimate the tie formation within 100 communities of different sizes. The studied model considers a large set of attributes, including both endogenous attributes (network motifs, e.g., stars and triangles) and exogenous attributes (economic variables, e.g., net sales and firm size). The estimation results show that the considered model converges and presents a high goodness of fit (GoF) for all communities. Moreover, it is found that the forces explaining the formation of links between suppliers and customers differ among communities. Some attributes, such as reciprocity, popularity, activity, location homophily, bank homophily and sales statistics, are common drivers of internal link formation for most of the studied communities. However, transitivity is rejected as a significant influencing factor for most communities, reflecting an absence of a sense of trust and reliability between firms with a common partner. Finally, we show that sector homophily does not serve as an obvious mechanism of partnership at the community level in the production network.

Keywords: ERGM, Japanese production network, Tie-formation process, Infomap algorithm, Network communities

Introduction

Recent economic phenomena, such as multiple unexpected global crises, have motivated scientists to consider the economy as a complex system. Agents (households, banks, firms, etc.) interact throughout economic networks, thereby determining and contributing to the emergence of macrofluctuations in the economy. The core of an economy is the production network, in which firms exchange goods and services (intermediate goods) and produce goods for consumption by final consumers (households or others). Such networks were considered by Gabaix (2011) and Acemoglu et al. (2012), who showed how idiosyncratic shocks at the microlevel lead to aggregated business cycle fluctuations.

Beginning in the last decade (see, for example, (Axtell 2001)), many works on production networks have emerged. Some works have looked at the topological characteristics of production networks, while others have analyzed the dynamics of link formation between “*upstream*” and “*downstream*” firms.

Several econophysics studies have characterized the topology of production networks, such as the works of Axtell (2001) and Fujiwara and Aoyama (2010), who studied the U.S. and Japanese production networks, respectively. The main findings in the literature include disassortative mixing in addition to the scale-free nature of the degree distribution and the Zipf distribution of the firm size. From another perspective, (Iino and Iyetomi 2015) specified the topology of the supplier-customer network of Japan in terms of its community structure. This analysis was extended by Chakraborty et al. (2018, 2019) who showed that the topology of the Japanese production network is better fitted by a walnut structure than by a bow-tie structure. This research was based on an analysis of the hierarchical structure of the communities in the nationwide Japanese production network.

Having identified the major empirical characteristics of production networks, researchers have been faced with the challenge of understanding how these properties emerge¹. Addressing such research questions requires the estimation of link formation, in which challenges arise due to the peer effect. These problems were enumerated by Jackson et al. (2017); the major concerns are related to the identification problem² and the endogenous network problem³. Frank and Strauss (1986) proposed a flexible and powerful class of models to deal with these problems: so-called exponential random graph models (ERGMs).

ERGMs have the ability to incorporate multiple choice-based variables representing endogenous attributes (network-based variables) and exogenous attributes (characteristic-based variables). Therefore, they can account for the interdependencies arising from the peer effect problem. ERGMs have recently been applied to a wide range of networks, such as brain networks (Simpson et al. 2011), disease networks (Rolls et al. 2013), and a climate change hyperlink network (Haussler 2018). In the context of production networks, most previous works have been limited to very small networks. For example, Lomi and Pattison (2006) considered a network of 106 firms in the Italian transportation sector, Lomi and Fonti (2012) studied a production network of 75 Italian companies involved in the production of machines for the manufacturing of ceramic tiles and Molina-Morales et al. (2015) modeled a business network containing 36 Spanish firms. All of these works demonstrated the effects of network-related attributes (such as transitivity and mutuality) in explaining partnerships between suppliers and customers. Krichene et al. (2018) were the first to study a large-scale production network, namely, the Tokyo Stock Exchange production network (a production network composed of 3189 listed firms). They identified the roles of various endogenous and exogenous attributes in the formation of ties between suppliers and customers.

In this paper, as a new contribution to this area, we propose a complementary study building on previous works on the application of ERGMs to production networks. As discussed previously, Chakraborty et al. (2018) extracted the hierarchical structure of communities in the Japanese production network using the Infomap method introduced by Rosvall and Bergstrom (2008). In this paper, we analyze the Japanese production network at the community level. Most irreducible communities are found to belong to the

second level of the hierarchy. This level provides a reasonable community resolution and exhibits a power law decay in the community size distribution. The 100 largest communities, whose sizes vary between 100 and 2347 firms, are considered for the estimation of the emergence mechanisms of their links by means of an ERGM.

Previous works have considered production networks based on industrial sectors to model the tie-formation process (see, for example, Lomi and Pattison (2006), Lomi and Fonti (2012) and Molina-Morales et al. (2015)). Chakraborty et al. (2018, 2019) showed that it is not sufficient to classify firms in the production network of Japan according to either the sector or the geography. Alternatively, the community structure, which is specified by complex properties based on the sector, geography and other economic attributes, constitutes the most suitable way to represent clusters of suppliers and customers. In this manuscript, our contribution is to understand the forces behind the formation of ties based on the community structure rather than on sectoral or geographical clusters as in the previous literature. Our main finding is that some attributes, such as reciprocity, location homophily and bank homophily, are common drivers of the formation of ties within communities, whereas the effects of some other attributes, such as transitivity and sector homophily, are heterogeneous among different communities.

This paper is organized as follows. First, we present the data and the properties of the hierarchical community structure of the Japanese production network. Then, we briefly introduce the ERGM and describe the considered statistical model. Subsequently, simulations are reported, and the estimation results are discussed. Finally, our conclusions and research perspectives are discussed.

The hierarchical structure of the Japanese production network

The Japanese production network data set is commercially available from Tokyo Shoko Research (TSR), Inc., one of the leading credit research agencies in Japan. TSR collects information about the 24 major suppliers and clients of each firm through questionnaires. However, the number of suppliers and customers of each firm is not limited to 24 because large firms are designated by many other firms as suppliers or customers. The resulting production network (consisting of 1,247,521 firms and 5,488,484 links) is an unweighted network representing the flow of goods and services from suppliers to customers. The data set contains, for each firm, precise information about its geographic location, its sectorwise classification, its sales figures, its number of employees and its major bank. Each geographic location is specified as one of the 47 Japanese prefectures, and the industrial sectors are hierarchically categorized into 20 divisions, 99 major groups, 529 minor groups and 1455 industries (Japan Standard Industrial Classification, November 2007, Revision 12). Prior to performing community detection, the data must be treated. Accordingly, following the exclusion of inactive and failed firms and the elimination of self-loops and parallel edges, the weakly connected giant component consisting of 1,066,037 firms and 4,974,802 edges is considered to be the final production network (see (Chakraborty et al. 2018)).

In the following subsections, we briefly present the community detection method and the results obtained, which are described in more detail in Chakraborty et al. (2018). Then, we present a descriptive analysis of the 100 communities selected to estimate the formation of ties using an ERGM. The considered 100 communities contain the largest Japanese firms in terms of sales, and represent all Japanese listed companies.

The community detection method

The map equation method introduced by Rosvall and Bergstrom (2008), popularly known as “Infomap”, is one of the best-performing algorithms (Lancichinetti and Fortunato 2009) for detecting communities in a large-scale network. Operating within the framework of information theory, it generates a map to describe the dynamics across the links and nodes of the network. The links in the network represent information flows between nodes. The Infomap method provides an efficient coarse-grained description of the information flows in the network and thus reveals the communities in the network by providing a compressed description of the flows. The algorithm uses a random walk as a proxy for information flow in the network. With this method, a subset of nodes in the network in which the random walker spends a relatively long time can be identified as a well-connected community.

The objective of the map equation method is to find an efficient code for minimizing the length of the description of a random walk to generate a partition \mathcal{C} dividing n nodes into c communities. The average single-step description length is defined as

$$L(\mathcal{C}) = q_{\sim} H(\mathcal{Q}) + \sum_{i=1}^c p_{\circ}^i H(\mathcal{P}^i).$$

The first term in the above equation represents the entropy of intercommunity movement, and the second term represents the entropy of intracommunity movement. The notation q_{\sim} represents the intercommunity switching probability of the random walker for any given step, and $H(\mathcal{Q})$ is the entropy of the codewords for the communities. The weight p_{\circ}^i is the fraction of intracommunity movement in community i , including the probability of exiting from the community, and $H(\mathcal{P}^i)$ is the entropy of intracommunity movement, including the exit code for community i .

The map equation method described above gives a two-level partition of the network, i.e. the original network and its related community structure. This two-level map equation method has a resolution limit problem. In fact, it is well known that there is no two-level community detection method which is resolution limit free when applied to a multi-level modular network, see Kawamoto and Rosvall (2015). Furthermore, it has been extended to a hierarchical map equation method (Rosvall and Bergstrom 2011), which can decompose a network into communities, subcommunities, sub-subcommunities and so on⁴.

The hierarchy of communities

We have employed the hierarchical map equation method to reveal the communities in the Japanese production network at different levels; the results are given in Table 1. Most of the irreducible communities are found at the second level, and we further observe that the community size distribution at this level is best fitted with a power law decay (Chakraborty et al. 2018). The partition we considered is based on the sizes of the detected communities. The first level contains very large communities which can be decomposed further. The third, fourth and fifth levels contains many small communities, i.e. several communities contain one firm, and then they are non-significant. In addition, most of the communities cannot be decomposed beyond the second level. Moreover, most of the subcommunities and 94% of the firms belong to the 2nd level communities. Because the

Table 1 The numbers of communities identified at different levels of the Japanese production network using the Infomap method

Level	c	c_r	n_c
1	209	106	830
2	65,303	60,603	998,267
3	18,271	17,834	61,748
4	1,544	1,539	5,168
5	10	10	24
Total	85,337	80,092	1,066,037

c denotes the total number of communities. The number of irreducible communities, which are communities that do not contain any subcommunities, is denoted by c_r . n_c denotes the number of firms in irreducible communities

sizes of the communities at this level also reflect a reasonable partition resolution, we investigate the network structure at the second level of the hierarchy using an ERGM.

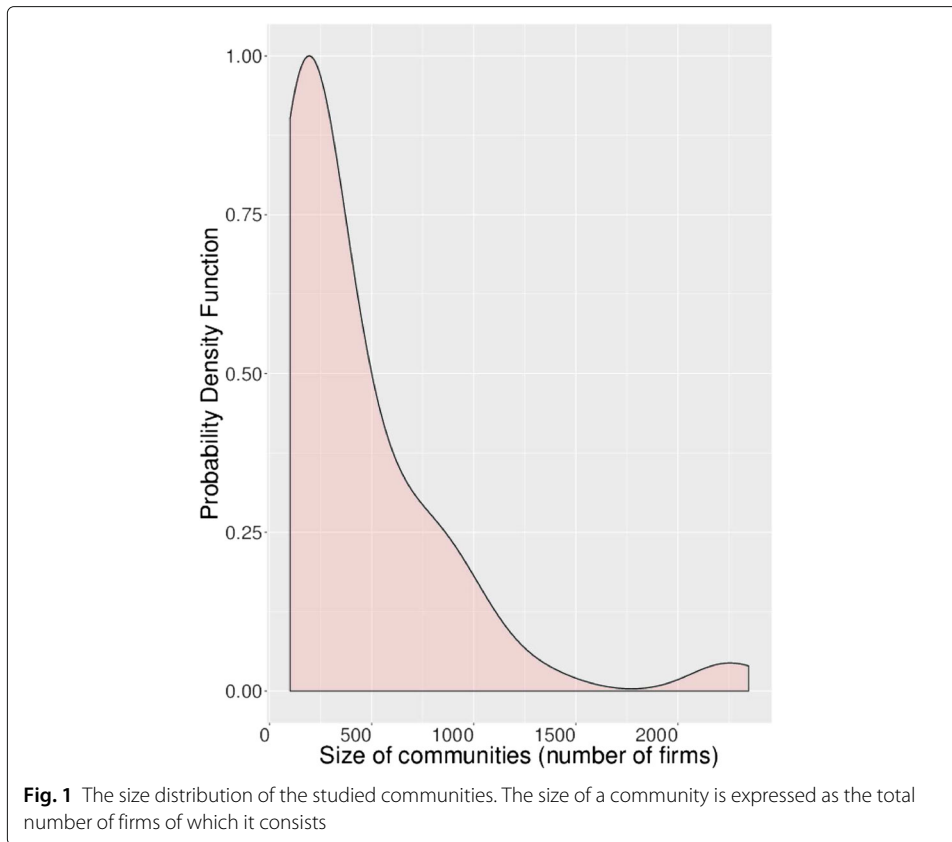
Properties of the considered communities

The second level of the hierarchy contains 65,303 communities. Approximately 1000 of these communities have a size of more than 100 firms. Therefore, communities consisting of fewer than 100 firms are not considered here because they are formed by small firms in terms of sales and number of employee. From among the communities of more than 100 firms, we selected the 100 most important communities in terms of total sales to obtain the best representation of the Japanese production network. Moreover, the selected communities contain all the listed Japanese firms in the Tokyo Stock Exchange which improve the economic representativeness of our sample. The selected communities represent 16% of the total sales in the production network and contain all listed firms. By referring to Gabaix (2011), the big firms in the economy explain the business cycle fluctuation. Figure 1 shows the heterogeneity of the considered communities in terms of size. On average, the community size is approximately 420.1 with a standard deviation of 441.8 (see Table 2 for details). We selected three statistics to describe the communities shown in Table 2. The first statistic is the fraction of firms from the same industrial sector, i.e., the number of pairs of firms from the same industrial sector divided by the community size. The second statistic is the fraction of firms with the same geographic location, i.e., the number of pairs of firms in the same prefecture divided by the community size. The third statistic is the fraction of firms with the same major bank, i.e., the number of pairs of firms with the same major bank divided by the community size.

The results in Table 2 suggest that these communities are characterized by different levels of similarity in terms of their industrial sectors, geographic locations or banking connections.

The exponential random graph model

An ERGM is a tie-based regression model that explains how links are formed between nodes. For a network $X = [x_{ij}]$, an ERGM regresses the adjacency matrix with a set of endogenous attributes z_a (network statistics) and exogenous attributes z_e (node characteristics). The canonical form of the ERGM is as follows:



$$\Pr_{\Theta}(X = x) = \frac{1}{\kappa(\Theta)} \exp \left(\sum_a \theta_a \cdot z_a(x) + \sum_e \theta_e \cdot z_e(x) \right), \tag{1}$$

where x is a realization of X , $\Theta = (\theta_a, \theta_e)$ is a vector of parameters of endogenous and exogenous attributes, and κ is a normalizing constant that ensures a proper distribution. Normalization is performed with respect to all possible network realizations, as follows:

$$\kappa(\Theta) \equiv \sum_{y \in X} \exp \left(\sum_a \theta_a \cdot z_a(y) + \sum_e \theta_e \cdot z_e(y) \right). \tag{2}$$

It is technically impossible to explicitly determine Eq. 2 due to the large number of possible network realizations, which increases exponentially with the number

Table 2 Summary of the properties of the considered communities

	Mean	Max	Min	std	Total
Community size	420.1	2,347	100	441.8	42,011
Fraction from the same sector	73.6%	95.5%	27.2%	14.5%	-
Fraction in the same location	89.7%	99.5%	62.3%	8.7%	-
Fraction with the same bank	27.7%	84.5%	2.8%	24.3%	-

of nodes. For a directed network of n nodes, one would need to determine all $4^{\binom{n}{2}}$ possible networks to calculate Eq. 2 and the true generation probability of the network ties. Consequently, the use of Markov chain Monte Carlo (MCMC) sampling techniques has been introduced in the literature (Snijders 2002). Because of the high level of computational resources required for the Monte Carlo simulation to estimate the parameters, we have implemented an ERGM estimation method based on a fixed-density MCMC sampling approach (discussed in Hunter et al. (2008)).

Accordingly, by estimating the parameters θ_a and θ_e , the normalization quantity $\kappa(\Theta)$ in Eq. 2 can be approximated by the MCMC sampling approach, and the probability of link formation $\Pr_{\Theta}(X = x)$ in Eq. 1 can be calculated. Appendix A presents the pseudo-code of the implemented ERGM.

The ERGM algorithm with fixed-density MCMC sampling

The idea is to estimate the parameters θ_a and θ_e such that the probability $\Pr_{\Theta}(X = x)$ defined in Eq. 1 generates networks X that are consistent with the observed network. We wish to solve the moment equation $E_{\theta}(z(X)) - z(x_{obs}) = 0$, where $z(X)$ represents the statistics of interest for a network X sampled with the MCMC approach and $z(x_{obs})$ represents the observed statistics for the real network.

Our algorithm is based on the stochastic approximation method proposed by Snijders (2002), which uses the Robbins-Monro algorithm for the maximum likelihood estimation (MLE) of the ERGM. The algorithm is composed of three phases: initialization, optimization and convergence (details are given in (Lusher et al. 2013)). The algorithm can be summarized into two major steps that are repeated until convergence is reached ($E_{\theta}(z(X)) - z(x_{obs}) \rightarrow 0$):

- 1 Use Θ to generate a network X via MCMC sampling.
- 2 Update Θ to minimize the moment equation.

In the first step, it would be highly time consuming to approximate all possible network realizations. Alternatively, in our code, we adopt the fixed-density MCMC sampler discussed by Snijders (2002) and Hunter et al. (2008). The fixed-density MCMC sampler randomly selects two dyads, namely, one null dyad ($x_{ij} = 0$) and one non-null dyad ($x_{ij} = 1$). Then, with the Hastings probability, these dyads are simultaneously toggled. Thus, the fixed-density MCMC sampler reduces the number of possible networks considered by keeping the global number of edges constant ($L = L_{obs}$).

Assumptions about the statistical model

In the considered ERGM, we consider different endogenous or network-dependent attributes and exogenous or economic attributes. Attributes are selected by investigating the properties of the Japanese production network and by referring to the economic properties of production networks highlighted in the literature as Gulati and Gargiulo (1999), Lomi and Pattison (2006) and Lomi and Fonti (2012). In total, we consider 8 endogenous attributes and 7 exogenous attributes.

The endogenous attributes

Endogenous attributes are social-based attributes related to the network structure (see Snijders et al. (2006) and Robins et al. (2007) for details about network motifs). The motifs considered in our statistical model are given in Eq. 3. The z_r statistic represents reciprocal links. It is expected that the probability of the emergence of a tie from supplier i to customer j will increase if j is already a supplier of i . This has been shown for several economic networks; see, for example, Gulati and Gargiulo (1999), Lomi and Pattison (2006) and Lomi and Fonti (2012). The Japanese production network is characterized by hubs; see Fujiwara and Aoyama (2010). Thus, ties are more probable for firms with higher in-degrees and out-degrees, i.e., so-called popular and active firms, respectively. This phenomenon was described in Lomi and Fonti (2012) as the trustworthiness of firms, i.e., other firms will have more confidence in more active and popular firms. This structure is modeled by the k -out-star (activity) and k -in-star (popularity) motifs (see z_{stars} in Eq. 3).

In addition, in a production network, there may be a correlation between a firm's activity and its popularity. A supplier with a larger number of customers (out-degree, or popularity) will require more intermediate goods and thus may have a larger number of suppliers (in-degree, or activity). Thus, to capture the popularity-activity correlation, the k -two-path motif is considered in our model (see z_{path} in Eq. 3).

Transitivity is a common property of social networks. In a production network, transitivity implies a higher level of trust between firms with a common partner; see Gulati and Gargiulo (1999). Accordingly, in the current model, we include four statistics of transitivity (k -triangles) based on the directionality of the edges: cyclic closure (AT-C), popularity closure (AT-D), path closure (AT-T) and activity closure (AT-U). All these statistics follow the $z_{triangles}$ form given in Eq. 3.

$$\left\{ \begin{array}{l} z_r = \sum_{i,j:x_{ij}=x_{ji}=1} x_{ij} \\ z_{stars} = \sum_{k=2}^{N-1} (-1)^k \cdot \frac{S_k}{\lambda^{k-2}} \\ z_{path} = P_1 - 2 \cdot \frac{P_2}{\lambda} \sum_{k=3}^{N-2} \left(\frac{-1}{\lambda}\right)^{k-1} \cdot P_k \\ z_{triangles} = 3 \cdot T_1 + \sum_{k=1}^{N-3} (-1)^k \cdot \frac{T_{k+1}}{\lambda^k} \end{array} \right. \quad (3)$$

In Eq. 3, S_k is the number of stars (either in- or out-stars) of order k , P_k is the number of two-paths of order k , and T_k is the number of triangles (AT-C, AT-T, AT-U or AT-D) of order k . The functional forms of these statistics were discussed in Snijders et al. (2006) as an alternative to the Markov assumption for an ERGM to ensure convergence. Economically, the use of these geometric forms decreases the impact of higher-order motifs on the partnering decisions of firms. With regard to k -stars (z_{stars}), we suppose that firms cannot have complete information about the numbers of suppliers and customers of all other firms. Thus, even if one supplier has 100 clients, a new potential client is

mainly influenced by a set of only a few clients. The same supposition holds for transitivity (k -triangles). For the k -two-path motif, we suppose that the correlation between the in-degree and out-degree has a certain saturation. In fact, when a firm establishes a contract with a new customer, it will not necessarily also look for a new supplier. Instead, the most probable case is that the firm will base its trade expansion strategy on its inventory.

Figure 4 in Appendix B presents a graphical illustration of the endogenous attributes considered in our model.

The exogenous attributes

As discussed previously, the process of tie formation between suppliers and customers is very complex and can depend on attributes other than the network motifs. The financial situation of the firm, the prices of the intermediate goods, and the reliability of the potential partner are some of the multiple economic attributes that can encourage two firms to become partners. Due to data limitations, some assumptions are required to select the most significant attributes for the Japanese production network.

Table 1 shows the similarities within the production network communities in terms of industrial sectors, geographic locations and major banks. At the community level, the formation of links between suppliers and customers can be stimulated by the homophily of these attributes; for example, the probability of link emergence between firms from the same community increases if they also have a common major bank. Thus, the sector homophily, geographic homophily and bank homophily are all considered in our model as follows: homophily = $\sum_{i,j} x_{ij} I(y_i = y_j)$, where $I(y_i = y_j)$ is an indicator function for the similarity between two attributes y_i and y_j . For sector homophily, the industrial level is considered, and for geographic homophily, the prefecture in which the head office of the firm is located is considered.

Moreover, firms are expected to choose partners based on wealth and size. The wealth of a firm is approximated as its total sales, while its size is represented by its number of employees, which is more stable over time (we note that large firms may realize poor profits or sales). These statistics are considered in terms of heterophily (sales heterophily and size heterophily) to see whether firms of similar wealth or size are more likely to be connected. The heterophily statistics are calculated as $\sum_{i,j} x_{ij} |y_i - y_j|$. In addition, the activity and popularity attributes, introduced as endogenous attributes, can exert complementary sales sender/receiver effects. These statistics reflect the potential of a firm to gain more clients (sender effect) or more suppliers (receiver effect) as its sales increase. The sales sender/receiver effects are expressed as $\sum_{i,j} x_{ij} y_i$ and $\sum_{i,j} x_{ij} y_j$, respectively.

ERGM simulations and estimated results

Simulations of the ERGM were carried out in parallel for all communities using the K computer⁵. For each community i , 100 simulations were performed to test the robustness and significance of the estimated parameters $\hat{\Theta}_i$. Based on the obtained estimates, for each community i , 100 networks were sampled to validate our model in terms of the goodness of fit (GoF) ratio proposed by Hunter et al. (2008).

Only the results for the three largest communities are presented in Table 3 for model validation because of the impossibility of displaying the GoFs for all 100 communities. High GoF values were also achieved for all other communities with the proposed statistical model.

Mechanisms of tie formation within the communities of the Japanese production network

The estimation results are presented in Figs. 2 and 3 and Table 4. The results are heterogeneous among communities, indicating the existence of different tie-formation mechanisms in the supply chain network of Japan at the community level. Figures 2 and 3 show the probability density functions of the estimated endogenous and exogenous parameters, respectively, across the 100 considered communities in the Japanese production network. Table 4 summarizes these results by means of seven columns presenting the average, maximum and minimum values and standard deviations of the parameters and the percentages of nonsignificant, significant positive, and significant negative effects. We note that the average values shown in column two of Table 4 are not considered to

Table 3 GoF analysis: Comparison between real and simulated networks

Attributes	Community 1		Community 2		Community 3	
	Real	Sim.	Real	Sim.	Real	Sim.
Endogenous attributes:						
Reciprocity	156.0 -	136.5 (1.89)	89.0 -	81.2 (1.19)	74.0 -	63.4 (1.96)
Popularity (<i>k</i> -in-stars)	2068.0 -	1895.3 (1.89)	2053.0 -	1978.3 (1.88)	1968.6 -	1934.1 (0.61)
Activity (<i>k</i> -out-stars)	359.0 -	343.6 (1.54)	527.0 -	497.3 (1.60)	414.0 -	381.6 (1.63)
<i>k</i> -two-paths	5000.8 -	4708.7 (1.64)	608.3 -	591.9 (0.53)	5773.1 -	5606.3 (1.94)
Cyclic closure (AT-C)	228.2 -	241.1 (-0.64)	89.9 -	85.2 (0.37)	104.1 -	119.6 (-1.11)
Path closure (AT-T)	4772.6 -	4467.5 (1.71)	518.4 -	506.7 (0.42)	5669.0 -	5486.7 (1.97)
Activity closure (AT-U)	4818.3 -	4547.7 (1.69)	473.5 -	487.1 (-0.51)	5584.6 -	5473.8 (1.18)
Popularity closure (AT-D)	600.6 -	737.4 (-1.93)	286.2 -	354.6 (-1.81)	579.1 -	673.7 (-1.17)
Exogenous attributes:						
Sector homophily	600.0 -	651.8 (-1.17)	1,073.0 -	1119.7 (-1.30)	322.0 -	295.8 (1.19)
Location homophily	849.0 -	787.2 (1.67)	703.0 -	688.7 (0.30)	2218.0 -	2096.7 (1.46)
Bank homophily	60.0 -	53.1 (1.68)	39.0 -	35.0 (1.40)	84.0 -	82.2 (0.41)
Size heterophily	$7.76 \cdot 10^8$ -	$7.99 \cdot 10^8$ (-0.10)	$7.28 \cdot 10^6$ -	$8.19 \cdot 10^6$ (-0.26)	$4.25 \cdot 10^7$ -	$4.69 \cdot 10^7$ (-0.90)
Sales heterophily	$2.73 \cdot 10^{11}$ -	$2.65 \cdot 10^{11}$ (1.10)	$1.48 \cdot 10^{11}$ -	$1.43 \cdot 10^{11}$ (1.78)	$4.41 \cdot 10^{12}$ -	$4.29 \cdot 10^{12}$ (1.46)
Sales receiver effect	$7.56 \cdot 10^9$ -	$8.65 \cdot 10^9$ (-1.59)	$2.00 \cdot 10^9$ -	$2.56 \cdot 10^9$ (-1.30)	$4.91 \cdot 10^{10}$ -	$5.30 \cdot 10^{10}$ (-0.83)
Sales sender effect	$2.74 \cdot 10^{11}$ -	$2.64 \cdot 10^{11}$ (1.37)	$1.48 \cdot 10^{11}$ -	$1.44 \cdot 10^{11}$ (1.44)	$4.41 \cdot 10^{12}$ -	$4.33 \cdot 10^{12}$ (0.98)

The networks represented here are the three largest communities considered from the second level of the hierarchical structure of the Japanese production network. The sizes of communities 1, 2 and 3 are 2,347, 2,249 and 2,173, respectively. The values given in parentheses represent the GoF ratio calculated as described by Hunter et al. (2008)

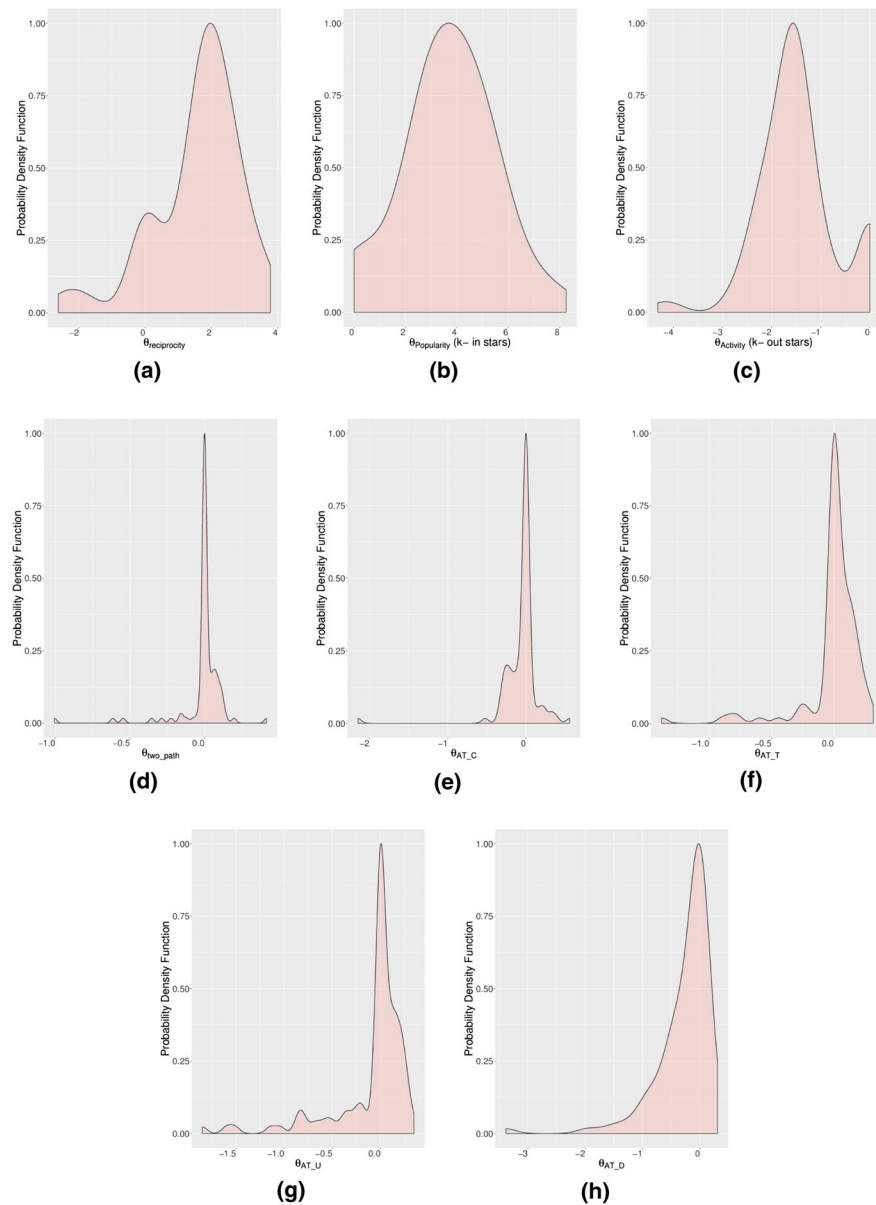
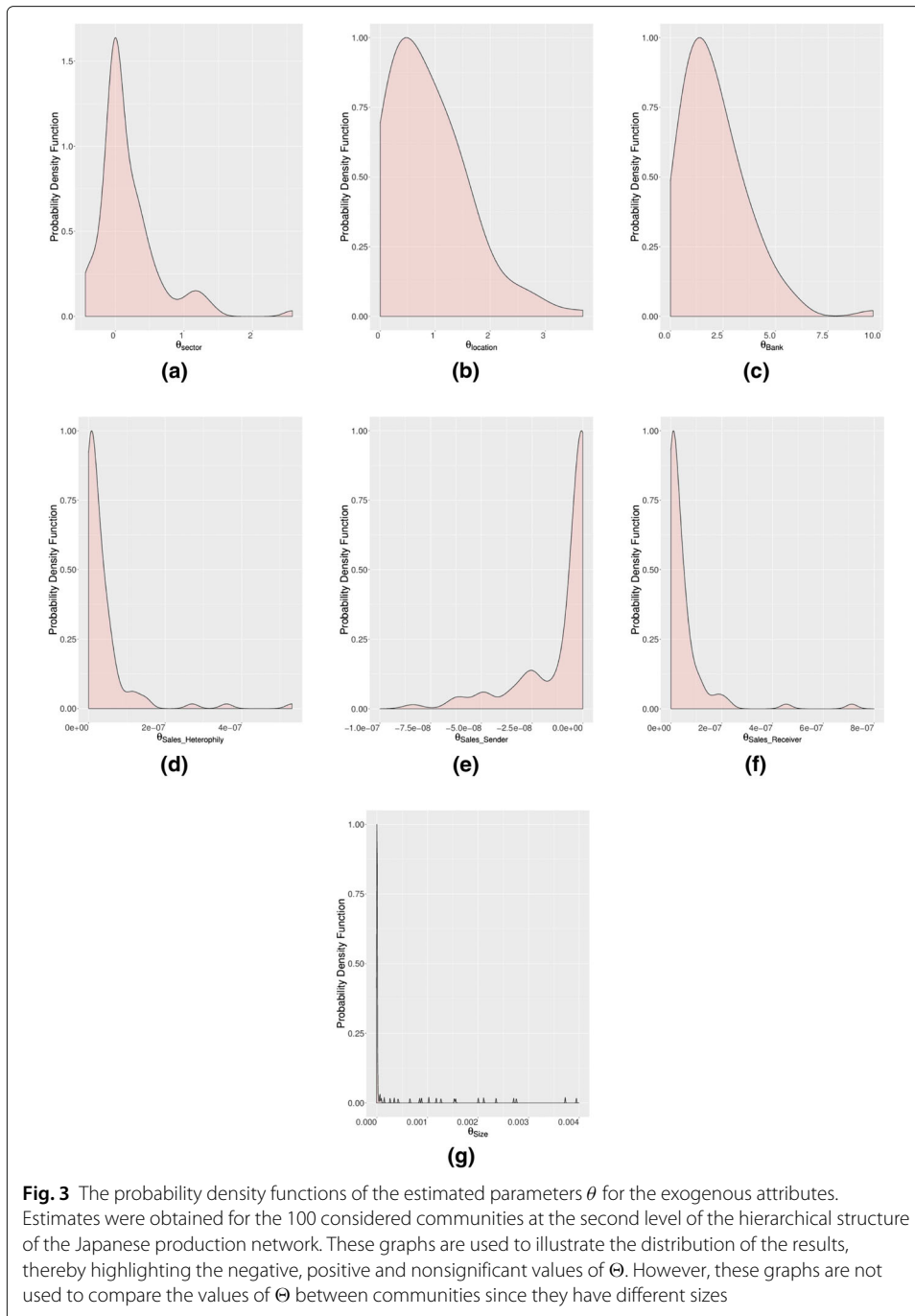


Fig. 2 The probability density functions of the estimated parameters Θ for the endogenous attributes. Estimates were obtained for the 100 considered communities at the second level of the hierarchical structure of the Japanese production network. These graphs are used to illustrate the distribution of the results, thereby highlighting the negative, positive and nonsignificant values of Θ . However, these graphs are not used to compare the values of Θ between communities since they have different sizes

be estimates for the global production network; they are given only for illustration. The significance was estimated using the Wald test, for which a Wald ratio of ≥ 2 indicates a significant parameter (see Lusher et al. (2013)). The significance tests were applied independently for each community.

The effects of endogenous attributes

Based on Table 4, some endogenous attributes have the same effect on all communities. In particular, reciprocity has a significant positive effect on 80% of our sample (in 15%,



reciprocity has no effect on the emergence of new relations between suppliers and customers). Thus, in 80% of the communities, there is a high chance of the emergence of a link from a supplier to a customer if the reverse relation exists. A few communities (5%) show a negative effect of reciprocity on the appearance of new links between firms. These five communities are characterized by an absence of reciprocal links; three of them have 2 reciprocal links, one has 0 reciprocal links, and one has 4 reciprocal links. Thus, the observed negative effect can be interpreted as an absence of mutual partnerships.

Table 4 A summary of the estimation results for the parameters Θ

Attributes	$\bar{\Theta}_{MLE}$	Θ_{MLE}^{max}	Θ_{MLE}^{min}	s.d. (Θ_{MLE})	nonsig.	positive	negative
Endogenous attributes:							
Reciprocity	1.59	3.78	-2.62	1.31	15%	80%	5%
Popularity (<i>k</i> -in-stars)	3.68	8.25	0.90	1.79	8%	92%	0%
Activity (<i>k</i> -out-stars)	-1.48	-0.66	-4.32	0.81	14%	0%	86%
<i>k</i> -two-paths	-0.01	0.41	-1.00	0.15	60%	28%	12%
Cyclic closure (AT-C)	-0.06	0.56	-2.14	0.26	59%	10%	31%
Path closure (AT-T)	0.00	0.31	-1.38	0.24	48%	40%	12%
Activity closure (AT-U)	-0.09	0.35	-1.58	0.40	39%	37%	24%
Popularity closure (AT-D)	-0.28	0.31	-3.38	0.51	50%	6%	44%
Exogenous attributes:							
Sector homophily	0.19	2.58	-0.43	0.45	41%	42%	17%
Location homophily	0.86	3.68	0.13	0.72	9%	91%	0%
Bank homophily	0.44	0.14	3.12	0.002	15%	80%	5%
Size heterophily	$3.04 \cdot 10^{-4}$	$9.02 \cdot 10^{-3}$	$-6.15 \cdot 10^{-3}$	$1.12 \cdot 10^{-3}$	48%	33%	19%
Sales heterophily	$3.59 \cdot 10^{-8}$	$5.31 \cdot 10^{-7}$	$4.31 \cdot 10^{-10}$	$7.26 \cdot 10^{-8}$	21%	79%	0%
Sales receiver effect	$5.80 \cdot 10^{-8}$	$1.22 \cdot 10^{-6}$	$4.73 \cdot 10^{-10}$	$1.50 \cdot 10^{-7}$	22%	78%	0%
Sales sender effect	$-1.31 \cdot 10^{-8}$	$-1.18 \cdot 10^{-10}$	$-3.37 \cdot 10^{-7}$	$3.71 \cdot 10^{-8}$	52%	0%	48%

The results for each parameter are given based on the 100 considered communities in the Japanese production network. A parameter is considered significant for a Wald ratio of ≥ 2 . Significance is rejected for a Wald ratio of < 2 . The entry in the significance column (sig.) is '+' if the parameter is significant and '-' otherwise

The popularity (k -in-stars) is the attribute with the most significant endogenous effect considered in our sample, with a significant positive effect on the emergence of new links in 92% of the communities. Thus, popular firms (firms with high in-degrees) are likely to become even more popular, i.e., suppliers have higher confidence in popular customers. Similarly, the activity (k -out-stars) has a significant negative effect in 86% of the communities. This finding indicates the existence of an upper bound on the number of customers that a supplier can have. This bound may be related to the production capacity of the supplier in terms of intermediate goods, i.e., a supplier cannot have an unlimited number of customers.

On the other hand, k -triangles are only weakly present. In fact, transitive cooperation is not common in the considered communities of the Japanese production network. Popularity closure (AT-D), cyclic closure (AT-C), activity closure (AT-U) and path closure (AT-T) all show nonsignificant or significant negative effects in 94%, 90%, 63% and 60% of the communities, respectively. Thus, aside from a few special cases, transitive cooperation is not a property of the studied communities. Production competition between firms may reduce their trust level and thus their inclination toward business cooperation. Similarly, Fujiwara and Aoyama (2010) showed an absence of significant clustering in the Japanese production network by comparing the real clustering to the level of clustering that would be generated by chance.

The k -two-path statistic is nonsignificant for 60% of the considered communities. In some communities (28%), there is a positive correlation between in-degree and out-degree. Accordingly, more popular (active) firms are more likely to be more active (popular). Thus, we can expect the formation of communities with hubs that simultaneously have many suppliers and many customers. In other communities (12%), there is a negative correlation between in-degree and out-degree. Economically, such a scenario could be explained as a community consisting of firms producing raw capital goods (firms at the top of the upstream channel) and firms producing goods for final consumption (firms at the bottom of the downstream channel, selling to the household market).

The effects of exogenous attributes

Location homophily is the most important exogenous factor in explaining the emergence of links at the community level. Table 4 shows that 91% of the considered communities exhibit a significant positive location homophily. Thus, the factor of distance is very important to strategic partnership decisions among Japanese firms. Moreover, bank homophily is the second most important factor. In 80% of the considered communities, the existence of a common major bank increases the probability that two firms will be connected. Surprisingly, sector homophily has only a limited influence on the emergence of links at the community level. In 41% of cases, a common industrial sector has no significant effect on the existence of partnerships between suppliers and customers, whereas 42% of the communities show significant sector-based selection in the formation of partnerships between Japanese firms. However, 17% of the communities show sector heterophily. This sector heterophily could be related to two possible scenarios: communities with highly diversified activities (firms need heterogeneous intermediate goods for production) or communities with sector homophily saturation (there is an upper bound on the formation of new links with firms from the same sector).

Although the firm size (number of employees) has a clearly nonsignificant effect on tie emergence, the sales volume may explain, in some cases, the formation of connections between suppliers and customers. In 79% of the considered communities, there is a significant positive sales heterophily, which implies that firms with lower production activity are more likely to be connected to firms with higher production activity. The sales receiver effect has a significant positive presence in 78% of the communities, in line with our finding concerning the popularity effect (*k*-in-stars). Thus, firms with higher production activity are more likely to receive intermediate goods from multiple suppliers. By contrast, the sales sender effect has a significant negative presence in 48% of the communities, in line with the previously discussed upper bound on the number of customers that a supplier can have, as indicated by the results for the activity statistic (*k*-out-stars).

Analysis of several special cases

We focus on the three largest communities at the second level of the hierarchical structure of the Japanese production network. Figure 5 in Appendix C shows the sectoral and geographical properties of these communities. The estimation results are given in Table 5, which shows that reciprocity, popularity, activity, location homophily, bank homophily and sales statistics are the common attributes that motivate the emergence of partnerships between suppliers and customers at the community level.

By contrast, the effect of transitivity depends on the properties of the community. In Table 5, community 2 shows a significant positive cyclic closure effect, which reflects an economy based on the exchange of general goods. However, community 3 shows a significant negative cyclic closure effect, and community 1 displays a nonsignificant effect. Communities 1 and 3 are also characterized by a significant positive effect of activity closure (AT-U), which means that two suppliers of the same firm, who might be assumed

Table 5 Estimation results for the three largest communities considered from the second level of the hierarchical structure of the Japanese production network

Attributes	Community 1		Community 2		Community 3	
Endogenous attributes:	Θ_{MLE}	sig.	Θ_{MLE}	sig.	Θ_{MLE}	sig.
Reciprocity	2.50	+	3.52	+	3.31	+
Popularity (<i>k</i> -in-stars)	8.25	+	6.23	+	7.80	+
Activity (<i>k</i> -out-stars)	-1.71	+	-1.69	+	-1.45	+
<i>k</i> -two-paths	0.07	+	0.08	-	0.07	+
Cyclic closure (AT-C)	-0.06	-	0.39	+	-0.19	+
Path closure (AT-T)	0.07	+	0.07	-	0.08	+
Activity closure (AT-U)	0.05	+	-0.17	+	0.05	+
Popularity closure (AT-D)	-0.55	+	-0.87	+	-0.47	+
Exogenous attributes:	Θ_{MLE}	sig.	Θ_{MLE}	sig.	Θ_{MLE}	sig.
Sector homophily	-0.42	+	-0.33	+	0.41	+
Location homophily	0.65	+	1.27	+	0.37	+
Bank homophily	1.88	+	3.26	+	1.92	+
Size heterophily	$1.73 \cdot 10^{-09}$	-	$-4.18 \cdot 10^{-08}$	-	$-1.56 \cdot 10^{-07}$	+
Sales heterophily	$2.58 \cdot 10^{-08}$	+	$5.27 \cdot 10^{-08}$	+	$1.56 \cdot 10^{-09}$	+
Sales receiver effect	$2.81 \cdot 10^{-08}$	+	$5.61 \cdot 10^{-08}$	+	$1.76 \cdot 10^{-09}$	+
Sales sender effect	$-6.23 \cdot 10^{-09}$	+	$-1.86 \cdot 10^{-08}$	+	$-5.76 \cdot 10^{-10}$	+

The sizes of communities 1, 2 and 3 are 2,347, 2,249 and 2,173, respectively. A parameter is considered significant for a Wald ratio of ≥ 2 . Significance is rejected for a Wald ratio of < 2 . The entry in the significance column (sig.) is '+' if the parameter is significant and '-' otherwise

to be competitors, are more likely to be partners. The economy in such a community may be regarded as a cooperative economy. By contrast, community 2 presents a significant negative estimate of the activity closure effect (-0.17), which implies that the firms are in competition and that suppliers of the same firm cannot be partners.

Community 3 shows a significant positive sector homophily (0.41), indicating that firms from the same sector are more likely to be connected. However, in communities 1 and 3, a significant negative sector homophily is observed. As seen from Table 3, community 1 has 600 links between firms from the same sector, community 2 has 1073 links between firms from the same sector, and community 3 has 322 links between firms from the same sector. These statistics represent densities of sector homophily of 10%, 25% and 5%, respectively, in each community. Thus, we cannot conclude that sector heterophily exists in communities 1 and 2. However, link saturation may be present, as explained in the previous section, which would decrease the probability of the emergence of new links between suppliers and customers from the same industrial sector. By contrast, in some other communities with a negative sector homophily, we find a low density of links between firms from the same industrial sector (fewer than 0.5%). In these cases, we can confirm the presence of sector heterophily.

Discussion and concluding remarks

This paper has presented a comparative analysis among communities in the Japanese nationwide production network. The communities show heterogeneous rules driving the formation of their internal ties. It has been shown that reciprocity, popularity, activity, location homophily, bank homophily and sales statistics constitute the common forces driving the formation of internal links within most of the studied communities. By contrast, transitivity is rejected as a motivation for connections between suppliers and customers in most communities. Accordingly, the phenomena of trustworthiness and reliability related to common partners that have been found in other studies of production networks, such as those of Gulati and Gargiulo (1999), Lomi and Pattison (2006), Lomi and Fonti (2012) and Krichene et al. (2018), cannot be confirmed at the community level. Moreover, it was expected that sector homophily would be one of the main driving forces of tie formation at the community level, as shown for the TSE production network by Krichene et al. (2018). However, through the ERGM estimation at the community level, it has been shown that sector homophily is not always a significant factor for tie formation in communities of the Japanese production network.

The tie-formation process based on the geographic location and sales confirms the well-known gravity model, as is discussed in the ERGM application by (Koskinen and Lomi 2013). However, we have shown that more complex features can explain the emergence of partnerships between suppliers and customers. In addition, the introduction of the network topology into the model captures the social-based effects not reflected by the economic attributes; this is shown by the mutual partnerships and by the high number of partners connected to hub companies.

Our presented results help policymakers to understand the organization between suppliers and customers in the production network of Japan. In fact, the way of how firms are organized reflects the origins of the economic dependencies between firms as it was discussed in Krichene et al. (2017), which is in the origin of multiple systemic economic risks.

Although this work contributes to research on production networks, some limitations must be discussed to pave the way for future work. First of all, the results may depend on the community detection technique applied, i.e., the community structure may change depending on the applied algorithm, which may affect the results of ERGM estimation. However, our choice of the Infomap algorithm was based on previous discussions such as those presented by (Rosvall and Bergstrom 2008), who showed that Infomap is suitable for networks with flows between nodes (flows of goods and services, in the case of production networks), and by (Lancichinetti and Fortunato 2009), who considered Infomap to be one of the best-performing algorithms on large-scale networks.

Another limitation of this work concerns the neglect of intercommunity links. Indeed, the entire production network of Japan contains more than one million firms. An ERGM cannot be used for the estimation of such an enormous network due to major limitations of computational feasibility. In addition, a network of such size can result in serious problems of degeneracy. Stivala et al. (2016) used the snowball sampling technique to estimate a large-scale network with an ERGM. This technique consists of sampling multiple sub-networks of moderate size, estimating their ERGMs and then performing estimation for the whole network via meta-analysis. Stivala et al. (2016) successfully applied this algorithm to a random network of 40,000 nodes. However, this technique is not suitable for the estimation of a real network such as the Japanese nationwide production network because of the scale-free topology of this network (see Fujiwara and Aoyama (2010)) and the multiple hubs it contains, which would cause the sampling results to be biased. Accordingly, focusing on the community level is an efficient way to begin to investigate the driving forces behind the formation of supplier-customer relationships. Future research will follow the recent work of Byshkin et al. (2018), who are working on speeding up MCMC sampling (see Byshkin et al. (2016)). In their recent work, these authors used an ERGM to perform estimation for a large-scale network of 104,103 nodes.

Endnotes

¹ Although our interest is focused on production networks, these topics have also been studied in the context of other networks, such as brain networks, WWW networks, and biological networks.

² In an interactive influence system, what behaviors should be specified in the estimation model?

³ The presence of a link between two agents could be due to unobservable behaviors.

⁴ The hierarchical map equation method from <http://www.mapequation.org/> is used in this study to reveal the hierarchical communities in the large-scale Japanese production network.

⁵ The K computer is the first 10-petaflop supercomputer; it was developed by RIKEN and Fujitsu under a Japanese national project. The system includes 82,944 compute nodes connected by Tofu high-speed interconnects. For more details, see (Yamamoto et al. 2014).

⁶ <https://github.com/hazem2410/SEPNET>

Appendix A: Pseudo-code for the ERGM algorithm

The C++ code is publicly available on Github⁶. It can be used for any production network following our defined statistical model. Other attributes can be added by the user in order to estimate another statistical model.

Algorithm 1 The fixed-density Markov chain Monte Carlo sampler: This algorithm allows sampling of networks based on the parameters Θ s. It is called in the following algorithm

```

1:  $\theta_a, \theta_e$  updated by the stochastic approximation Algorithm 2
2: || for  $i = 1 : M_{max}$  do
3:   → while(False) do
4:     ⇒ Select one null dyad:  $x_{ij} = 0$ 
5:     ⇒ Calculate the change probability of adding a link between  $i, j : x_{ij} = 1$ :
        $P_+ = \exp(\theta_a \cdot z_a^+ + \theta_e \cdot z_e^+)$ 
6:     ⇒ #Comment:  $z_a^+, z_e^+$  are the change statistics
7:     ⇒ #Comment: e.g, if  $z_a$  is the total number of edges;  $z_a^+ = +1$ 
8:     ⇒ With the probability  $P_+$ , accept the change, then: True
9:   → repeat until (True)
10:  → while(False) do
11:    ⇒ Select one non-null dyad:  $x_{ij} = 1$ 
12:    ⇒ Calculate the change probability deleting a link between  $i$  and  $j : x_{ij} = 0$ :
        $P_- = \exp(\theta_a \cdot z_a^- + \theta_e \cdot z_e^-)$ 
13:    ⇒ #Comment:  $z_a^-, z_e^-$  are the change statistics
14:    ⇒ #Comment: e.g, if  $z_e$  is the total sector homophily and  $i$  and  $j$  belong to the
       same sector,  $z_e^- = -1$ 
15:    ⇒ With the probability  $P_-$ , accept the change, then: True
16:  → repeat until (True)
17: || End for

```

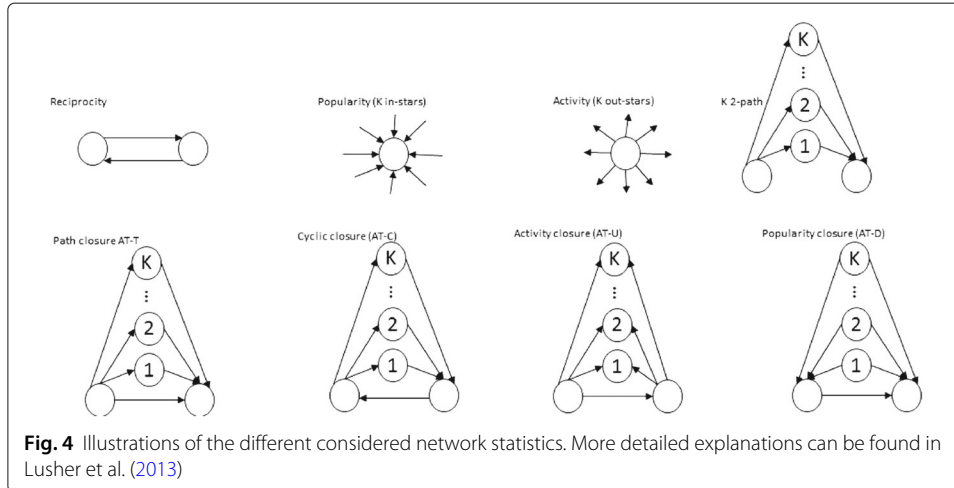
Algorithm 2 The stochastic approximation algorithm: This algorithm updates the value of the parameters Θ s until the convergence

```

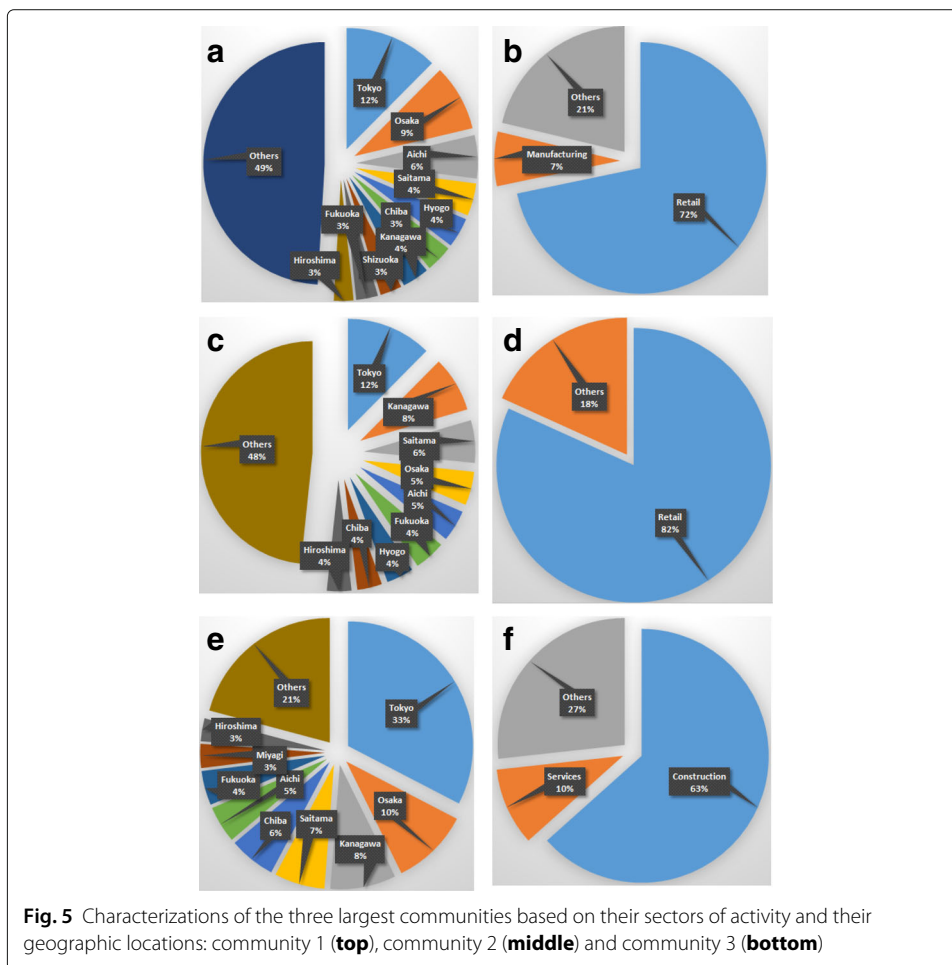
1: Initialize  $\theta_0 = (\theta_a; \theta_e) = (0; 0)$ 
2: || for  $i = 1 : M_0$  do
3:   → Generate  $M_0$  networks from Algorithm 1
4: || End for
5:  $Z_{obs}$  is the vector of real network statistics
6: #Comment: Total number of edges, total number of triangles, etc.
7:  $E$  is the vector of the average of the  $M_0$  simulated network statistics
8:  $D$  is the covariance matrix of the  $M_0$  simulated network statistics
9: || for  $i = 1 : M_{max}$  do
10:  → Generate one network from Algorithm 1 with the values of  $\theta_{t-1}$ 
11:  → Update  $\theta_t = \theta_{t-1} - a \times D_0^{-1} \times (Z - Z_{obs})$ 
12:  → #Comment:  $Z$  is the vector of simulated network statistics
13: || repeat until convergence or End For #Comment: If "End For" is reached without
       convergence, the algorithm is considered as non-converging, and further specifica-
       tion of the statistical model are required (add or discard endogenous and exogenous
       attributes).

```

Appendix B: The endogenous network statistics



Appendix C: Characterizations of the three largest communities



Abbreviations

ERGM: Exponential random graph model; GoF: Godness of fit; MCMC: Markov chain monte carlo; TSR: Tokyo Shoko research

Acknowledgements

The authors are grateful to Hideaki Aoyama (Kyoto Univ.), Thomas Lux (Kiel Univ.), Hiroshi Yoshikawa (Rissho Univ.), Yoshiyuki Arata (RIETI) and Alex Stivala (Univ. of Melbourne) for helpful comments and discussions. This research used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID: hp170242).

Funding

This study was conducted as 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 supported by MEXT under two Exploratory Challenges on the Post-K Computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena and Macroeconomic Simulations), by a Grant-in-Aid for Scientific Research (KAKENHI), and by JSPS Grant Number 17H02041.

Availability of data and materials

The data that support the findings of this study are available from the Tokyo Shoko Research (website: <http://www.tsr-net.co.jp/>) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. TSR offers various services/ data in accordance with customers' demand, such as TSR Report, TSR Company Profile Data File, financial statement data, Internet Service "tsr-van2", viability scores, etc.

Authors' contributions

HK implemented the code, analyzed and interpreted the data regarding the estimation of the Japanese production network. AC applied the Infomap detection of communities of the Japanese production network. YF and HI prepared the

database, performed an examination of the paper and were major contributors in writing the manuscript. MT examined the implementation of the MPI and the execution of the code on the K computer. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Graduate School of Simulation Studies, University of Hyogo, Kobe, Japan. ²RIKEN Advanced Institute for Computational Science, Kobe, Japan.

Received: 27 August 2018 Accepted: 10 January 2019

Published online: 31 January 2019

References

- Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A (2012) The network origins of aggregate fluctuations. *Econometrica* 80(5):1977–2016
- Axtell RL (2001) Zipf distribution of u.s. firm sizes. *Science* 293:1818–1820
- Byshkin B, Stivala A, Mira A, Krause R, Robins G, Lomi A (2016) Auxiliary parameter mcmc for exponential random graph models. *J Stat Phys* 165:740–754
- Byshkin M, Stivala A, Mira A, Lomi A (2018) Fast Maximum Likelihood estimation via Equilibrium Expectation for Large Network Data. <https://arxiv.org/abs/1802.10311>
- Chakraborty A, Kichikawa Y, Iyetomi H, Iino T, Inoue H, Fujiwara Y, Aoyama H (2018) Hierarchical communities in walnut structure of japanese production network. *PLoS ONE* 13(8):0202739
- Chakraborty A, Krichene H, Inoue H, Fujiwara Y (2019) Characterization of the community structure in a large-scale production network in japan. *Physica A: Stat Mech Appl* 513:210–221
- Frank O, Strauss D (1986) Markov graphs. *J Am Stat Assoc* 81:832–842
- Fujiwara Y, Aoyama H (2010) Large-scale structure of a nation-wide production network. *The European Physical Journal B-Condensed Matter and Complex Systems* 77(4):565–580
- Gabaix X (2011) The granular origins of aggregate fluctuations. *Econometrica* 79(3):733–772
- Gulati R, Gargiulo M (1999) Where do interorganizational networks come from?. *Am J Sociol* 104(5):1439–1493
- Haussler T (2018) Heating up the debate? measuring fragmentation and polarisation in a german climate change hyperlink network. *Soc Networks* 54:303–313
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M (2008) ergm: a package to fit, simulate and diagnose exponential-family models for networks. *J Stat Softw* 24(3):54860
- Iino T, Iyetomi H (2015) Community Structure of a Large-Scale Production Network in Japan. In: *The Economics of Interfirm Networks*. Springer, pp 39–65
- Jackson MO, Rogers BW, Zenou Y (2017) The economic consequence of social-network structure. *J Econ Lit* 55(1):49–95
- Kawamoto T, Rosvall M (2015) Estimating the resolution limit of the map equation in community detection. *Phys Rev E* 91(1):012809
- Koskinen J, Lomi A (2013) The local structure of globalization: The network dynamics of foreign direct investments in the international electricity industry. *J Stat Phys* 150(5)
- Krichene H, Chakraborty A, Inoue H, Fujiwara Y (2017) Business cycles' correlation and systemic risk of the japanese supplier-customer network. *PLoS ONE* 12(10):0186467
- Krichene H, Arata Y, Chakraborty A, Fujiwara Y, Inoue H (2018) How Firms Choose their Partners in the Japanese Supplier-Customer Network? An application of the exponential random graph model
- Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. *Phys rev E* 80(5):056117
- Lomi A, Fonti F (2012) Networks in markets and the propensity of companies to collaborate: An empirical test of three mechanisms. *Econ Lett* 114:216–220
- Lomi A, Pattison P (2006) Manufacturing relations: An empirical study of the organization of production across multiple networks. *Organ Sci* 17(3):313–332
- Lusher D, Koskinen J, Robins G (2013) *Exponential Random Graph Models for Social Networks: Theory, Methods and Applications*. Cambridge University Press
- Molina-Morales FX, Belso-Martinez JA, Mas-Verdu F, Martinez-Chafer L (2015) Formation and dissolution of inter-firm linkages in lengthy and stable networks in clusters. *J Bus Res* 68(7):1557–1562
- Robins G, Snijders TAB, Wang P, Handcock M, Pattison P (2007) Recent developments in exponential random graph (p*) models for social networks. *Soc Netw* 29:192–215
- Rolls DA, Wang P, Jenkinson R, Pattison PE, Robins GL, Sacks-Davis R, Daraganova G, Hellard M, McBryde E (2013) Modelling a disease-relevant contact network of people who inject drugs. *Soc Netw* 35(4):699–710
- Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci* 105(4):1118–1123

- Rosvall, M, Bergstrom CT (2011) Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6(4):18209
- Simpson SL, Hayasaka S, Laurienti PJ (2011) Exponential random graph modeling for complex brain networks. *PLoS ONE* 6(5):20039
- Snijders TAB (2002) Markov chain monte carlo estimation of exponential random graph models. *J Soc Struct* 3:1–40
- Snijders TAB, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. *Sociol Methodol* 36:99–153
- Stivala AD, Koskinen JH, Rolls DA, Wang P, Robins GL (2016) Snowball sampling for estimating exponential random graph models for large networks. *Soc Netw* 47:167–188
- Yamamoto K, Uno A, Murai H, Tsukamoto T, Shoji F, Matsui S, Sekizawa R, Sueyasu F, Uchiyama H, Okamoto M, Ohgushi N, Takashina K, Wakabayashi D, Taguchi Y, Yokokawa M (2014) The k computer operations: Experiences and statistics. *Procedia Comp Sci* 29:576–585

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
