RIETI Discussion Paper Series 18-E-011

How Firms Choose their Partners in the Japanese Supplier-Customer Network? An application of the exponential random graph model

Hazem KRICHENE
University of Hyogo

ARATA Yoshiyuki
RIETI

Abhijit CHAKRABORTY
University of Hyogo

FUJIWARA Yoshi
University of Hyogo

INOUE Hiroyasu
University of Hyogo

RIETI
Research Institute of Economy, Trade & Industry, IAA

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

RIETI Discussion Paper Series 18-E-011
March 2018

How Firms Choose their Partners in the Japanese Supplier-Customer Network? An application of the exponential random graph model*

Hazem KRICHENE[1], ARATA Yoshiyuki [2], Abhijit CHAKRABORTY[1], FUJIWARA Yoshi [1], INOUE Hiroyasu [1]
[1]Graduate School of Simulation Studies, University of Hyogo, Kobe
[2]Research Institute of Economy, Trade and Industry, Tokyo

Abstract

This work aims to explain how firms behave and select their suppliers and customers in the Japanese production network. We study a supplier-customer network of listed firms in Japan (3,198 firms with 20,417 links). In order to specify how firms choose their partners, the so-called exponential random graph model is applied to estimate the ties formation process. For the estimation of such a large-scale network, we employ a recent technique of sampling called the improved fixed density Markov Chain Monte Carlo (MCMC). Our main result shows that all of the effects (social and economic effects) are statistically significant in explaining the ties formation between firms. Social effects such as mutuality and transitivity with common partners in different directional links between suppliers and customers are shown. Moreover, homophily with the same industrial sectors and geographical locations, and disassortative mixing between low-profit firms and high-profit ones are also found. We argue that our method is extended to the spatially heterogeneous structure of communities reflecting industrial sectors and geographical locations and temporal changes of supplier-customer relationships in such a framework of the stochastic actor-oriented model.

Keywords: Supplier-customer network, Exponential random graph model, Improved fixed density MCMC, Social and economic effects
JEL classification: C15, L22, D85

*This study is conducted as a part of the project "Large-scale Simulation and Analysis of Economic Network for Macro Prudential Policy" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This research was supported by MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena, Macroeconomic Simulations), Grant-in-Aid for Scientific Research (KAKENHI) by JSPS Grant Number 17H02041. The authors are grateful for helpful comments and discussions by Hideaki Aoyama (Kyoto Univ.), Thomas Lux (Kiel Univ.), Hiroshi Yoshikawa (Rissho Univ.) and Discussion Paper seminar participants at RIETI.

1 Introduction

After the last global financial crisis, several work as Gatti et al. [2010] and Fagiolo and Roventini [2017], showed limits of DSGE models which were unable to provide effective policy advice in such disastrous period. In fact, they explained how DSGE models are based on the assumption of the rational expectations of economic agents where agents' behaviors are summarized to a problem of inter-temporal optimization. Therefore, these approaches neglect the economic complexity driven by the bounded rationality of economic agents. The relation between crisis and economic agents' irrationality was discussed in Loretan [1996]. Moreover, current literature considers that bounded rationality is the main microscopic ingredient of the business cycle fluctuations as shown in Gabaix [2016]. These fluctuations are the outcomes of suppliers-customers' interactions[1] as discussed in **?**. Accordingly, firms behave as a "*black box*" and decide to establish commercial links with suppliers and customers in the economy.

In Jackson et al. [2017], this "*black box*" is referred to the endogeneity of networks in terms of relationship between the links formations and the firms' behaviors. Following these authors words: "*the symbiotic relationship of social context and behavior complicates empirical analysis, since the relationships among most of the variables of interest are endogenous. It is thus essential for many economic questions to understand how networks form, evolve, and interact with behaviors*". In that way, some empirical works tried to explain this "*black box*" decision of firms in economic networks. Gulati and Gargiulo [1999] considered a database of 166 business organization from different industries (new materials (62), automotive products (52), industrial automation (52)) and different countries Countries (Japan (66), USA (54), Europe (46)). They observed these agents during 9 years. Their main result suggests that the probability of alliance between two organizations increases by their endogenous interdependence as mutual alliance (reciprocity), common third part (transitivity)... Authors highlighted the importance of social attributes in the process of deciding strategic alliance between business organizations. With a more specific approach to explain links formation within a supplier-customer network, Lomi and Pattison [2006] considered a database of 106 firms which belong to the sector of production of the means of transportation in Italy. They applied an Exponential Random Graph Model (ERGM) proposed by Snijders [2002], i.e. P* model. This method is considered as a powerful formulation that incorporates all endogenous inter-dependencies as explained in Jackson et al. [2017]. This approach allows the regression of the observed adjacency matrix with endogenous and exogenous attributes. The first represent the network topology, while the second are refereed to the agent characteristics (economic, geographic or financial variables). Main result of this study shows that ties formation depends on both endogenous and exogenous attributes. They quantified the contribution of each attribute in the micro-level interaction between firms.

In reality, economic systems are much larger and much more complex than those considered in previous works (size around 100 firms). In fact, stochastic network methods which deal with the estimation of ties formation are mainly oriented for social networks which have a small size. These models -as ERGM- suppose that in a small size world all people know each others and can make decision about their relationships. This assumption motivates the computational problem of these models. Indeed, the estimation of ERGM requires a high computational time and the model can degenerate for size more than 200 agents. To deal with this constraint, we aim in this work

[1] Suppliers-customers' interactions are represented by the production network which is the backbone of the economy. Some works present a deep survey of the Japanese production network, see for e.g. Chakraborty et al. [2017].

to construct our own "fast" ERGM algorithm based on new improvements in literature in order to estimate the ties formation process in the Japanese production network.

In a novel contribution, our goal is to quantify the self-organization of firms' behavior in a large-scale economic network. In fact, to overcome with the endogeneity limits of the classic econometric approaches, ERGM is applied on a large-scale economic network to understand how economic behaviors interact with social behaviors as explained in Jackson et al. [2017].

Stivala et al. [2016] exposed a new approach to estimate ERGM for large-scale network. They applied a snowball sampling algorithm, and generated multiple small-medium size networks. They estimated the conditional independent samples and applied a meta-analysis technique to get the whole network estimation. Their technique was applied on random networks with size of 40.000 nodes. However, on real networks as the production network of Japan, the degree distribution does not follow a binomial law as for random networks. In fact, real networks are characterized by high heterogeneity and existence of hubs which may generate serious bias in the obtained snowball samples. To avoid this bias, Byshkin et al. [2016] introduced a new MCMC sampler to estimate ERGM in faster time. This method is designed to decrease the number of allowed network states without worsening the Markov chains, as claimed by the authors. As result, they estimated an ERGM for a real Netscience network of size 1589 without using any sampling technique, as the snowball sampling to generate sub-networks. As mentioned before, other than time consuming problem of ERGM, this approach suffers from the high risk of degeneracy in large-scale network, specially when we deal with a high dimensional model (ERGM with high number of endogenous attributes, as for.e.g. k- stars, k-triangles...). Recently, Thiemichen and Kauermann [2017] introduced a method to deal with the degeneracy problem. In fact, they proposed a subsampling scheme to obtain conditional independent samples. Then, they replaced some network statistics by smooth functions as discussed in the Curved ERGM literature by Hunter [2007]. These two methods allowed them to fit stable model for large networks.

Based on these recent developments, we construct in this work an ERG Model from scratch on Python which uses a stochastic approximation method to estimate the parameters (see Snijders [2002]) with the Improved Fixed Density MCMC sampler introduced in Byshkin et al. [2016]. We apply this algorithm on the production network of Japan based on fourteen endogenous and exogenous attributes related to social behavior and rational economic behavior, respectively. The production network[2] is a directed network between suppliers and customers. It represents the flow of goods and services in the Japanese economy. We note that this network is unweighted. However, the production network has more million firms, and in the current contribution we cannot estimate the entire network without any sampling technique (this point is discussed in the conclusion). Thus, we consider a sub-production network which counts only firms that belong to the Tokyo Stock Exchange[3]. This network is denoted in this paper by the TSE (Tokyo Stock Exchange) production network which has 3,198 firms and 20,417 links. The total sales amount of these firms is 22% of the total sales of the whole production network. Therefore, our considered sample is significant and can approximate the estimation of ties formation in the Japanese economy.

This paper is organized as follows. Section 2 presents the data and some properties of the considered production network. Section 3 discusses the ERGM approach, the proposed algorithm and the considered statistical model. In Section 4, simulations are carried out and results are

[2]This network is based on the 2016 study by Tokyo Shoko Research (TSR).

[3]The Nikkei Digital Media database is used to select the sub-production network of firms belong to the Tokyo Stock Exchange.

exposed and discussed. Finally, conclusion and research perspectives will be given in Section 5.

2 The TSE production network

The TSE production network considers the flows of goods and services between suppliers and customers in 2016. 3,198 listed firms are considered with 20,417 links and 22% of the total sales of the whole production network. Figure 1 illustrates the TSE production network, which shows scale-free behavior. A few firms have a high degree, which reflects their importance in the economy. This result is confirmed in Figure 2 showing the in and out degree distributions with tails' indexes $P(k_{in}) \propto k_{in}^{-2.52}$ and $P(k_{out}) \propto k_{out}^{-3.26}$ highlighting the high heterogeneity of the network. Moreover, the production network shows a disassortative mixing (the assortativity coefficient, $r = -0.21$) confirmed by the decreasing shape of the nearest neighbor degree with the total degree (slope of -0.4) in the Figure 3.b. Figure 3.a the decreasing shape of the clustering coefficient with the total degree (slope of -0.82). This reflect the low probability of having triangle loops for hubs (high degree nodes), which is an indication about the hierarchical structure of the TSE production network.

All the found properties show the non-random behavior of the TSE production network which is an indication about the existence of some social mechanism of partners selection. This observation will be studied through an Exponential Random Graph Model.

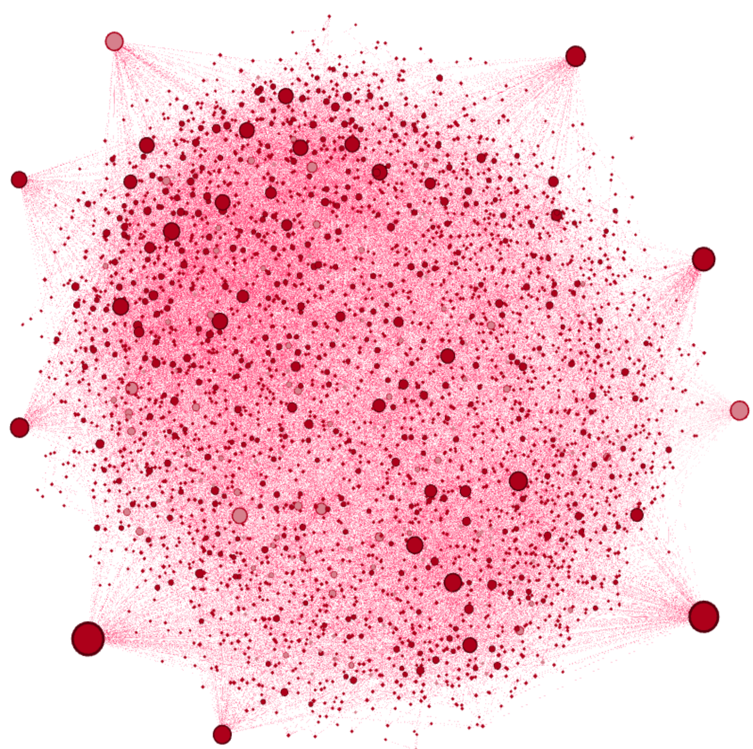

Fig. 1. The TSE supplier-customer network: It reveals the existence of a few large firms with high degrees and many small firms with low degrees.

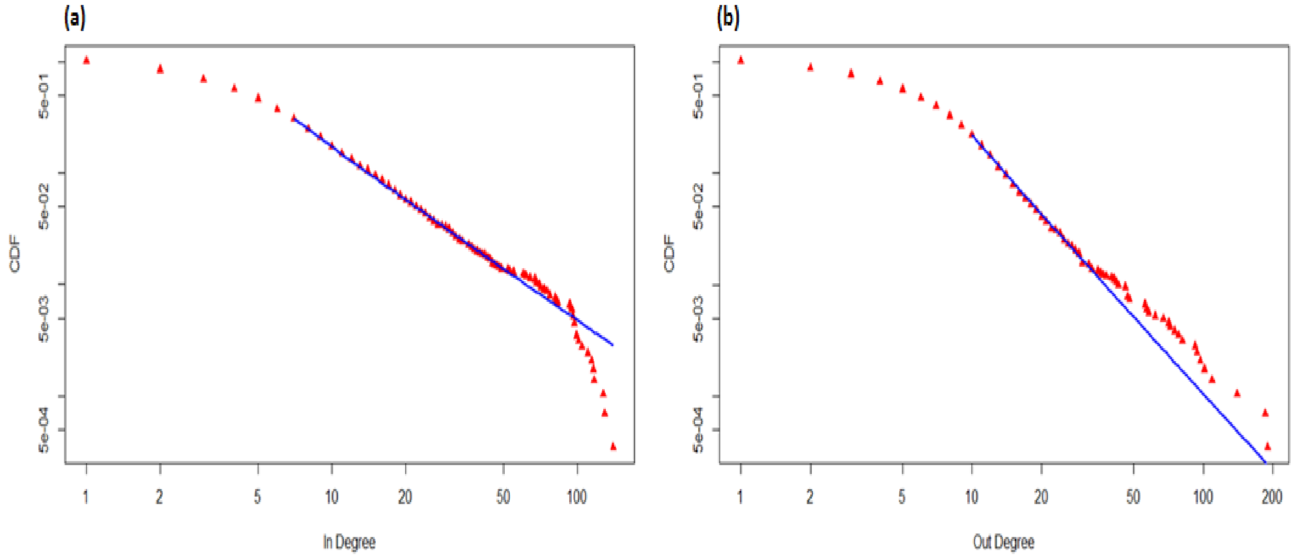

Fig. 2. The degree distributions of the TSE supplier-customer network: Both, the in and out degree distributions show scale-free behavior, with the respective power-law indexes $P(k_{in}) \propto k_{in}^{-2.52\pm0.02}$ and $P(k_{out}) \propto k_{out}^{-3.27\pm0.03}$.

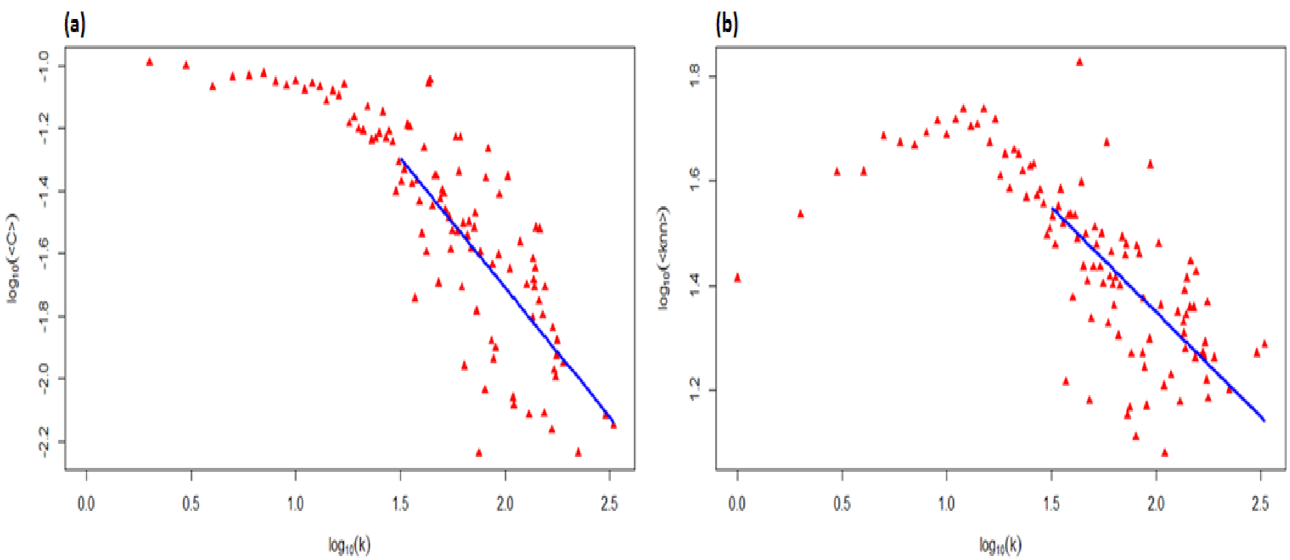

Fig. 3. The relationship between average clustering, average nearest neighbors' degrees and the total degree: Both figures show the disassortative topology of the Japanese production network $C(k) \propto k^{-0.82}$ and $P(k_{nn}) \propto k^{-0.40}$

3 Exponential random graph models

ERGM technique is highly used in the study of social networks, but much less application on economic networks. Up to our knowledge, a few works estimate ERGM of economic networks as Lomi and Pattison [2006] (network of 106 Italian firms of production of the means of transportation), Lomi and Fonti [2012] (network of 75 Italian firms of the production of machines for the manufacturing of ceramic tiles). This size restriction is mainly due to the high computationally intensive Monte Carlo estimation of ERGM parameters. In the following of this section, our statistical model will be presented. Then, we will discuss the use of recent developments of ERGM to build a fast algorithm able to estimate a the large TSE production network.

3.1 ERGM: theory and assumptions about the statistical model

Let's $X = [X_{ij}]$ be the adjacency matrix of an unweighted network. ERGM is simply a regression of X with a set of endogenous attributes z_a and exogenous attributes z_e. It can be considered as a logistic regression. However, due to the endogeneity problem (we explain tie between two nodes based on the existent ties is the network), we cannot apply a logistic regression, see for more details Lusher et al. [2013]. Therefore, the canonical form of ERGM is given as follows:

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

where, x is a realization of X, θ_a and θ_e are parameters of endogenous and exogenous attributes, and κ is a normalizing constant to ensure a proper distribution. Large simulations are carried out in order to estimate the ERGM parameters $\hat{\theta}_a$ and $\hat{\theta}_e$.

3.1.1 Assumptions about the endogenous attributes

A random network as Erdös-Renyi model depends only on a probability p to generate ties, which reflects the network density. To estimate such network, the simplest ERGM configuration is required called, the Bernoulli model. This model depends on one endogenous attribute reflecting the number of edges z_L associated to the parameter θ_L. In case of scale-free network, Bernoulli model cannot fit the real data (see for e.g. Lomi and Pallotti [2013]). However, it is of high importance to consider the number of edges z_L in order to control the degree distribution of the real estimated network. Negative (positive) value of θ_L implies a low (high) network density.

The probability of link from i to j increases if a link from j to i exists. It constitutes the simplest form of the social-based attachment. This fact was shown on different economic network (Gulati and Gargiulo [1999], Lomi and Pattison [2006], Lomi and Fonti [2012]). Thus, the reciprocity statistics z_R is considered in our ERGM associated to the parameter θ_R.

Since the Markov random graphs model (Frank and Strauss [1986]), a multiple graphlets were introduced as endogenous attributes to estimate real networks. The more general model is the social circuit dependence introduced in Snijders et al. [2006], Robins et al. [2007]. As graphlets we can cite some, like: k-stars, k-triangles, 4-cycle, bridging... We can model as much as possible. However, it is better to consider the more significant attributes depending on the research goals. In our case, we will consider some graphlets which have a social meaning in case of economic relationships.

The TSE production network is a scale-free network (see Figures 1 and 2) where hubs exist, i.e. big firms with high in and out degrees. The ties formation process in terms of preferential attachment is explained as follows: more active (higher in-degree) and more popular (higher out-degree) firms will be more active and more popular, which explain the emergence of hubs. To model these patterns, the proposed ERGM considers two additional endogenous attributes: k-in-stars and k-out-stars (see Lusher et al. [2013] for details about these graphlets).

Figure 3a. shows that the TSE production network is characterized by triangles. In fact, it shows that there is a higher probability that two firms establish commercial link if they share a common neighbor. The probability increases if two firms share more than one common neighbor (k-neighbor). In that case, we are talking about k-triangles statistics. In Gulati and Gargiulo [1999], Lomi and Pattison [2006], it was shown that transitivity has an effect on ties formation in economic network. Accordingly, the transitivity is considered in our ERGM through four statistics which depends on the ties' directions: z_{AT-T}, z_{AT-C}, z_{AT-D} and z_{AT-U} (see Lusher et al. [2013] for details about these graphlets).

Figure 3.b shows that the TSE production network is characterized by disassortative mixing. To capture this correlation of ties formation, we introduce in our ERGM model the k-2-path statistics z_P.

Therefore, nine endogenous attributes are considered refereed to the social irrational decision of firms. All the discussed graphlets are exposed graphically in Figure 10 of the appendix A. However, this high number of network statistics with the large size of the TSE production network will cause the degeneracy of the ERGM estimation. To avoid such problem, we follow the approach of Hunter [2007] by modeling: the k-stars as Geometrically Weighted Degree (GWD) function, the k-triangles as Geometrically Weighted Edgewise Shared Partner (GWESP) and the k-2-path as Geometrically Weighted Dyadic Shared Partner (GWDSP).

3.1.2 Assumptions about the exogenous attributes

Several economic attributes can influence the ties formation between suppliers and customers, as: the financial health of the partner, the quality of its offered goods and services, the applicable price, the payment method and delay... In reality, it is very complex. However, due to data constraint, we should introduce some assumptions in order to consider the most significant attributes that are available.

In Chakraborty et al. [2017] it was shown that in the Japanese production network communities are formed based on sector and geographic location of firms. This result can reflect that firms with similar sector of activity and similar location are more likely to establish commercial links. Therefore, firms in the TSE production network will be characterized by their sectors and locations as attributes. We note that 50% of the TSE firms are located in Tokyo[4], other firms have heterogeneous location distribution. Accordingly, the results discussed hereafter cannot be generalized to the whole production network. Sector homophily and location homophily are two exogenous attributes for the ERGM estimation. Homophily between firm i and firm j in terms of their attributes y_i, y_j is given by $\sum_{i,j} x_{ij} y_i y_j$.

In Krichene et al. [2017] it was shown that in the Japanese production network dependence between firms can be explained by their respective profit. Thus, profit will be considered as attribute in the ERGM estimation (total sales are assumed as profit). Consequently, three additional exogenous attributes are considered: profit heterophily, profit sender effect and profit receiver effect. For an attribute y, sender effect, receiver effect and heterophily effect between firm i and firm j are given respectively by: $\sum_{i,j} x_{ij} y_i$, $\sum_{i,j} x_{ij} y_j$ and $\sum_{i,j} x_{ij} |y_i - y_j|$.

[4] TSR data consider the location of the main branch only.

3.2 A "*fast*" ERGM algorithm

The implemented ERGM algorithm is described through the diagram of Figure 4. We use the stochastic approximation approach of Snijders [2002] in order to estimate Θ. Each step is based on a sampled network through a Markov Chain Monte Carlo (MCMC) procedure. Byshkin et al. [2016] introduced a new sampling technique, the Improved Fixed Density MCMC. This technique allows the generation of sampled network with stable number of edges L^s, i.e. $L^s \in [L^{obs}-1, L^{obs}+1]$. This procedure allows the decrease of the number of allowed networks which accelerate the convergence by minimizing the number of required iterations of the stochastic approximation. With the IFD sampler on a Desktop Core i7-6700 CPU 3.40 GHz allows the convergence of the algorithm for a network with 3,198 firms in almost 13 hours.

However, under the constraint of having stable number of edges, the IFD sampler may generate artificially high GoF. Therefore, to have a robust Goodness of Fit procedure, we generate samples based on the Metropolis MCMC sampler with the estimated Θ (see Lusher et al. [2013] for this technique). These samples will be compared to the real data.

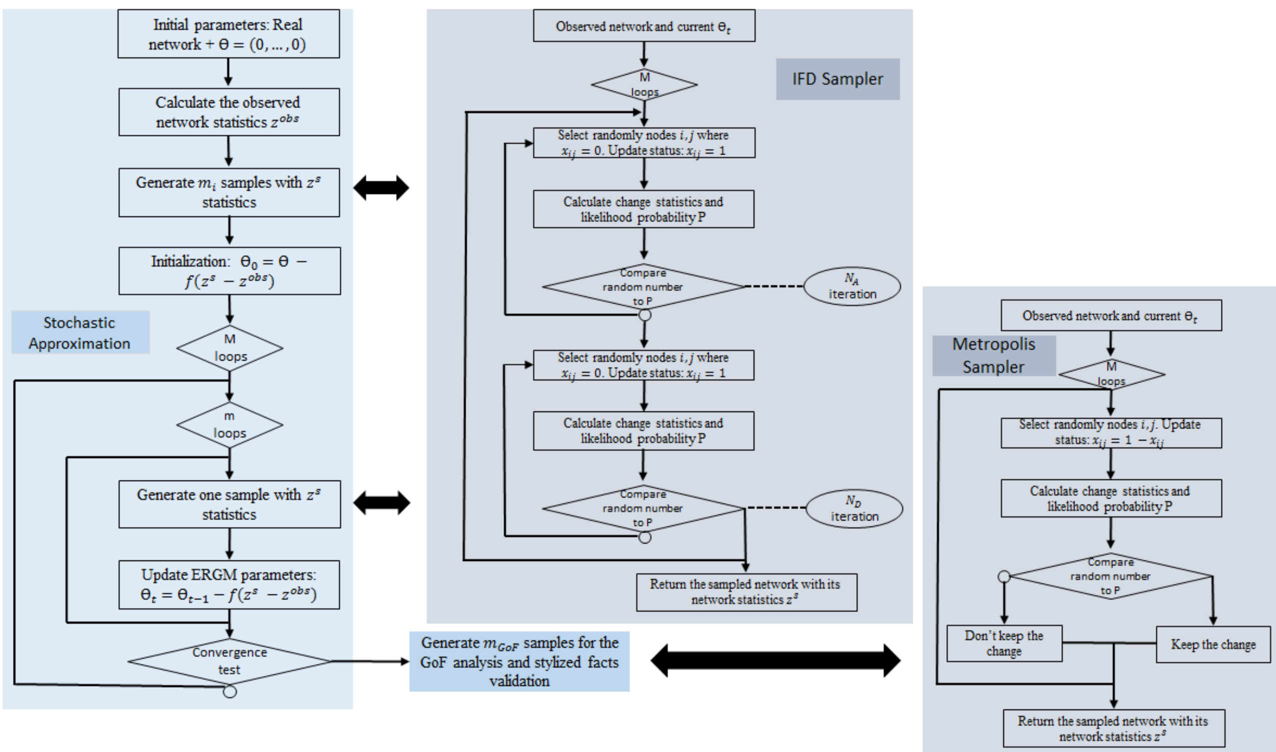

Fig. 4. A detailed diagram about the proposed algorithm. A stochastic approximation method is used to estimate Θ. The IFD sampler is used in estimation to minimize the number of required iterations until the convergence. The Metropolis sampler is used to have a robust Goodness of Fit analysis.

4 Estimation of parameters of the TSE production network

Based on the described data and the implemented ERGM algorithm, we will expose the simulation experiments procedure allowing the estimation of the parameters Θ. The model will be validated based on the GoF analysis. Then, results will be exposed and discussed.

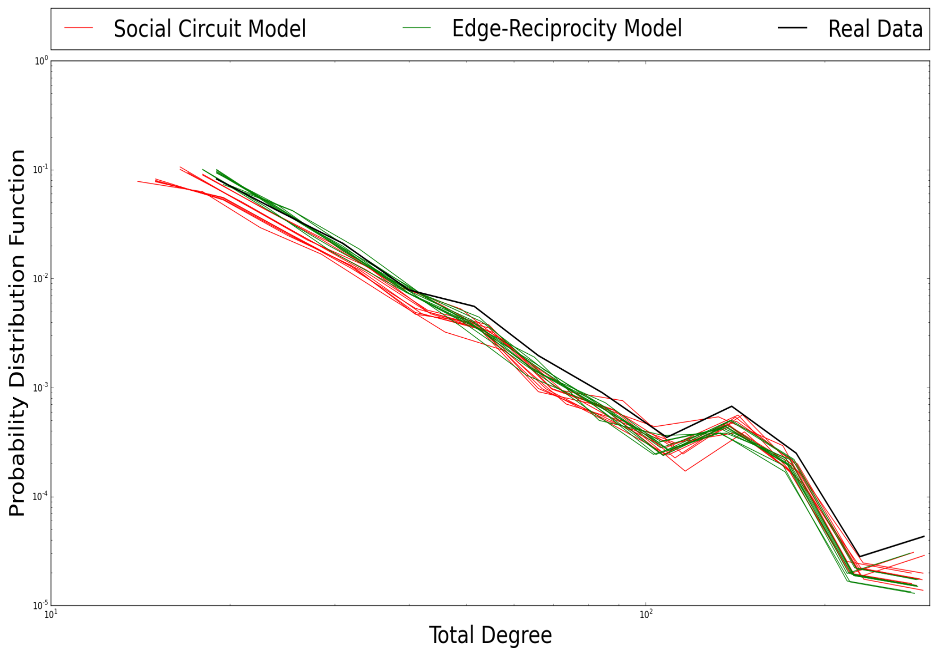

Fig. 5. A comparison of the degree distributions between real network (black), 20 social circuit ERGM networks (red) and 20 Bernoulli ERGM networks (green).

4.1 Estimation and simulation experiments

The estimation procedure depends on the stochastic approximation algorithm iterations. Each iteration is done after several iteration of the IFD sampler. To generate one sample, the IFD sampler is simulated for 100.000 steps. The convergence of parameters[5] is reached after 100 steps of stochastic approximation. In total, 10.000.000 simulation steps were required to get the estimated values $\hat{\Theta}$.

Using $\hat{\Theta}$, 20 samples were generated for the GoF[6] analysis. Each sample is based on 1.000.000 steps simulation of the Metropolis MCMC sampler.

To show the importance of introducing several endogenous attributes in our statistical model, we estimated a simple ERGM for the same data based on two endogenous attributes: degree and reciprocity. This model estimation follows the same procedure described above.

[5] The t-ratio is used as convergence criterion. For all parameters, $|\mathrm{t-ratio}| \leq 0.1$.

[6] The GoF analysis is based on the following test: For each parameters θ_k, the ratio $\frac{|z_k^{obs} - z_k^s|}{SD(z_k^s)}$ is calculated. If this ratio is large, the observed measure is far from what is expected under model. Works consider the value 2 as critical. If the ratio is larger than 2, the GoF for the parameter k is rejected.

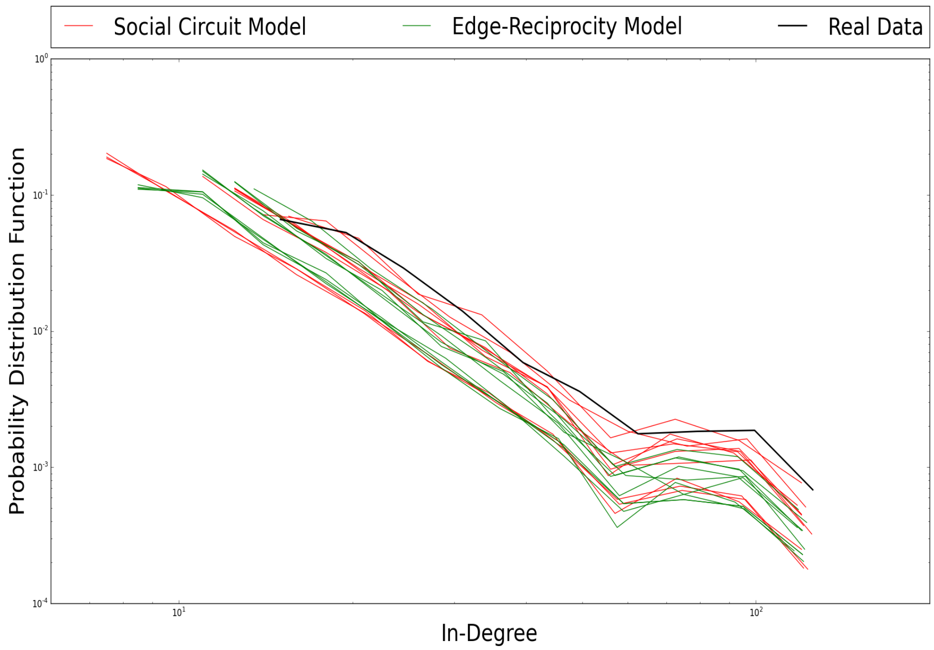

Fig. 6. A comparison of the in-degree distributions between real network (black), 20 social circuit ERGM networks (red) and 20 Bernoulli ERGM networks (green).

4.2 GoF and model validation

Five stylized facts are considered to validate the proposed model. Two models are compared with real data: a degree-reciprocity model and a social circuit model. In terms of degree distribution (see Figures 5, 6 and 7), the degree-reciprocity model shows a good GoF, graphically and numerically, i.e. GoF ratio less than 2. It means that the degree-reciprocity model captures well the real degree distribution of the TSE production network. However, this model fails to reproduce more complex behaviors as shown in Figures 8 and 9 for the distributions of edge-wise shared partners and geodesic distance.

To reproduce the more complex patterns of the TSE production network, a more general ERGM is required. In fact, the GoF of the social circuit model shows a high fitting with real data. This fitting is shown on all Figures 5, 6, 7, 8 and 9, and gave a GoF ratio less than 2 for all the estimated parameters (14 endogenous and exogenous parameters). Therefore, we validate the employed ERGM estimation which will allow the interpretation of the obtained results.

4.3 Numerical results and discussion

Numerical results of the ERGM estimation are exposed in Table 1. The first finding shows that all the considered attributes are highly significant. Consequently, firms' decisions in the TSE production network are based on social and economic behaviors. This result will be discussed in the following parts by commenting the social and economic behaviors and by discussing some possible policies.

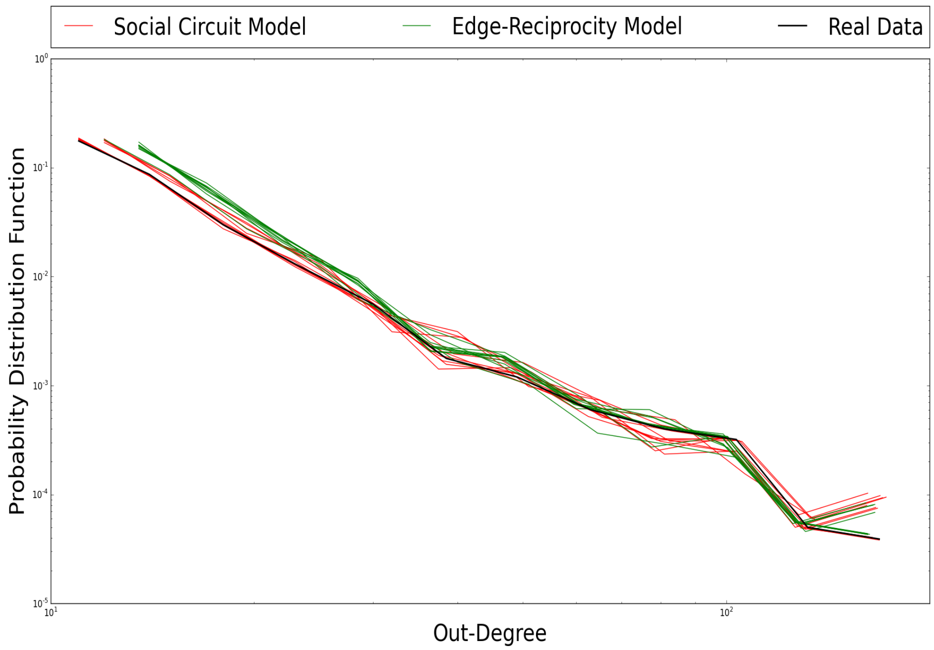

Fig. 7. A comparison of the out-degree distributions between real network (black), 20 social circuit ERGM networks (red) and 20 Bernoulli ERGM networks (green).

4.3.1 Social selection process

As explained in Section 3, degree attribute has no social meaning, it reflects only the network density. The obtained negative parameter of degree (-4.05) reflects the very low density of the TSE production network which is 0.2%.

All other parameters reflect a significant presence of social preferential attachment. Moreover, the likelihood of links formation increases by the reciprocity, the transitivity and the activity and popularity of firms. In fact, based on results of Table 1, the positive parameters Θ of reciprocity, triangles and stars indicate that firms are more likely to establish links in case of mutual trading, common partner or popular and active partner, respectively. Finally, the negative k-2-path shows the negative correlation between in-out degrees which confirms the disassortative mixing of the network (see Figure 3.b). This is an indication that less smaller firms are more likely to be connected to bigger firms.

4.3.2 Economic selection process

All exogenous attributes are significant. Firms choose rationally their partners based on their sector of activity, their geographic location and their financial wealth (profit). Indeed, the likelihood of links formation increases by the sector homophily, the geographic homophily and the profit heterophily.

In terms of sector, suppliers and customers from same sector are more likely to exchange goods and services between them. The economy is considered by a sector-based interaction process. This may make trading easier, i.e. firms with same economic activity have similar interest. However,

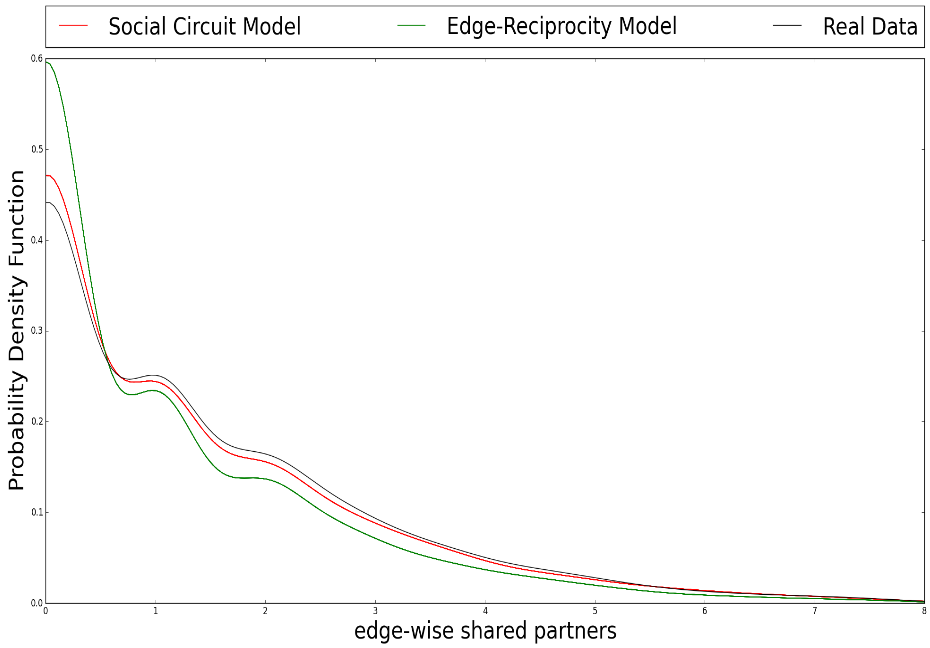

Fig. 8. A comparison of the edge-wise shared partners distributions between real network (black), 20 social circuit ERGM networks (red) and 20 Bernoulli ERGM networks (green). Shared partners are considered without direction.

the economy present systemic sector-based risk. If a disaster happen in one firm of sector A, a high contagion effect could be observed along other firms of that sector.

Moreover, firms choose partners which belong to the same geographic location (prefecture). This choice can be explained by the financial benefit of having geographically close partners (less transportation cost). However, the economy presents a significant systemic geographic-based vulnerability. In case of Japan, this vulnerability could be dramatic especially with the high risk of natural disasters which can induce firms of same prefecture to a cascading avalanche of bankruptcies (see Inoue and Todo [2017] for details about natural disasters in the production network).

Moreover, firms' decisions are affected by the profit. In fact, firms with higher profit are more likely to be active and popular (profit sender and receiver effects). In addition, the positive profit heterophily indicates that firms with low profit tends to be connected to firms with high profit. This mechanism stimulates the emergence of hubs in the production network.

4.3.3 A vulnerable concentrated production network: discussion and policies

According to the discussed economic selection processes, the production network displays high concentration, i.e. small world structure as discussed in Nakajima et al. [2012], Saito [2013], Saito. Saito highlighted the high vulnerability of the supply chain due to this small world structure especially in case of natural disaster as the Great East Japan Earthquake. In this part we discuss some options to reduce the potential vulnerability of the production network.

First concentration effect is induced by the sector-based structure of the production network. These links are naturally created due to the economic activity of firms. One option to reduce the

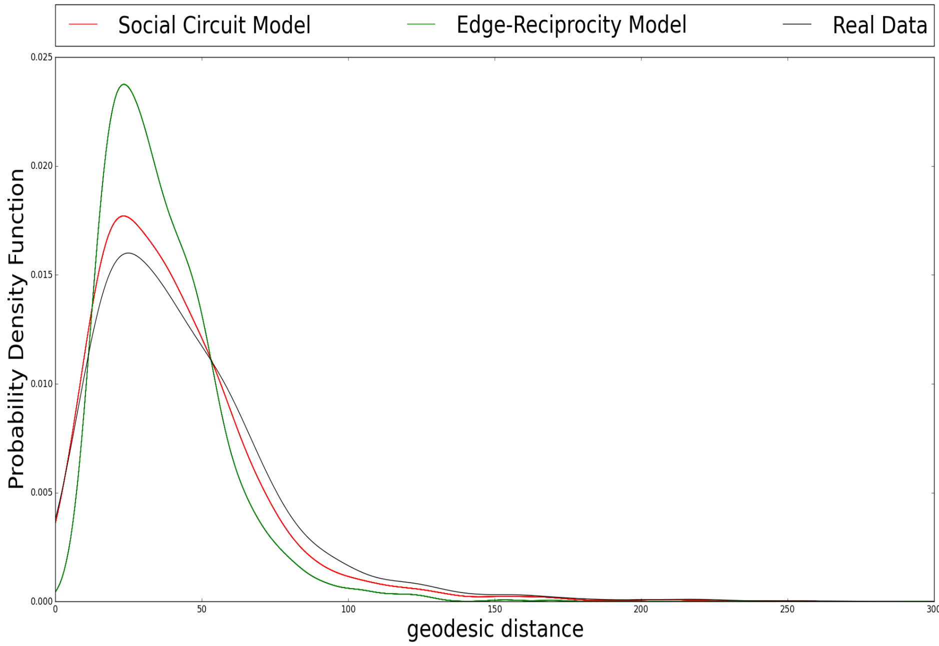

Fig. 9. A comparison of the geodesic distance distributions between real network (black), 20 social circuit ERGM networks (red) and 20 Bernoulli ERGM networks (green).

sector-based risk is to encourage holding companies[7]. By diversifying activities firms can protect their business and improve their productivity over a long-time (see Kawakami [2017]). In addition, in case of sector-based crisis, the holding company can insure easy fund transfer to its affected firms with lower cost which can reduce the crisis spread.

The geographic concentration effect is the key of natural disasters vulnerability. In fact, firms with suppliers and customers in the same affected prefecture will have a difficulties to continue production (see Saito, Inoue and Todo [2017]). This fact will increase the impact on the whole economy: the spread and the persistence of the shock. By boosting firms decentralization, the geographic effect will be reduced. If firms have benefits as tax reducing on far trading, they will be motivated to diversify their partners from all over Japan, which will reduce the geographic dependence and the high productivity disturbance in case of natural disasters.

The hub concentration effect is behind this sector and geographic organization as explained in Saito [2013]. Following words of Song and Schaar [2015]: "*Under incomplete information, the formation process converge to a hub network*". Thus, hub structure is the outcome of an incomplete information. In fact, small firms will have more confidence toward popular and active firms and look to establish links with them. In case of complete information, firms can choose their partners based on the real product quality rather than the popularity influence.

Accordingly, information transparency will help firms to diversify their partners, to able to decentralize their trading relationships especially with some financial motivations like fiscal advantages.

[7] Holding company does not produce goods or services. It owns others companies from different economic sectors.

Table 1
Estimation results of ERGM applied on the TSE production network. All endogenous and exogenous attributes are highly significant based on the p-value.

Attributes	$\hat{\theta}_{MLE}$	$s.e(\hat{\theta}_{MLE})$	p-value	Attributes	$\hat{\theta}_{MLE}$	$s.e(\hat{\theta}_{MLE})$	p-value
Degree	-4.05	0.01	≤ 0.01	Profit Sender	$2.31 \cdot 10^{-07}$	$9.61 \cdot 10^{-10}$	≤ 0.01
Reciprocity	1.15	0.004	≤ 0.01	Profit Receiver	$1.02 \cdot 10^{-07}$	$4.10 \cdot 10^{-10}$	≤ 0.01
In-Stars	2.01	0.004	≤ 0.01	Out-Stars	2.04	0.005	≤ 0.01
Sector Homophily	0.59	0.005	≤ 0.01	Location Homophily	0.16	0.002	≤ 0.01
Two-Path	-0.02	$9.08 \cdot 10^{-05}$	≤ 0.01	Profit Heterophily	$8.37 \cdot 10^{-08}$	$4.43 \cdot 10^{-10}$	≤ 0.01
AT-T	0.06	0.0007	≤ 0.01	AT-U	0.17	0.0005	≤ 0.01
AT-C	0.13	0.002	≤ 0.01	AT-D	0.20	0.0008	≤ 0.01

5 Conclusion

In this paper we presented a "fast" ERGM estimation on a large economic network of the TSE production network in order to quantify the self-organization of firms in the Japanese economy. In a new perspective than classic econometric modeling, this work estimates the system evolution under the endogeneity property reflected by the inter-dependencies between the network links, the network statistics (social behavior) and the economic behaviors.

Results showed the high significance of all endogenous and exogenous attributes. We showed that endogenous attributes are important in the formation process of the production network, and allowed us to validate our modeling assumptions by reproduction of several real stylized facts.

Exogenous attributes allowed us to understand the self-organization of firms in the TSE production network. Indeed, sector, geographic location and profit are essential in the firm decision process. However, these attributes highlight the vulnerabilities of the Japanese economy due to its small world structure. Some alternatives were discussed to reduce such risks especially with the frequent natural disasters that happen in Japan. A decentralized economy with diversified holding companies could be an alternative to reduce the concentration effect of the TSE production network.

ERGM is a powerful model to estimate the TSE production network under the presence of endogeneity. It allowed the quantification of the likelihood of edge formation based on several attributes. However, this model gives only statistical explanation and presents a lack of economic interpretation, as the impact of the variation of parameters on the real economy (GDP, inflation...).

Other works should follow this paper to reach more precise analysis and economic policies proposal. Near future research is looking to quantify the firms' behavior in the whole production network of Japan (more than million firm) using parallel computing on K computer. Moreover, in future researches we will overcome the limits of ERGM (lack of economic interpretation) by merging these results with an agent-based model which reproduce the business cycle dynamics. The idea is to simulate the impact of changing the parameters Θ on the real economy, i.e. for different parameters configuration, simulate different network and analyze the impact on the business cycles generated by the agent-based model.

Acknowledgment

This study is conducted as a part of the Project “Large-scale Simulation and Analysis of Economic Network for Macro Prudential Policy" undertaken at Research Institute of Economy, Trade and Industry (RIETI). This research was supported by MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena, Macroeconomic Simulations), Grant-in-Aid for Scientific Research (KAKENHI) by JSPS Grant Number 17H02041. The authors is grateful for helpful comments and discussions by Hideaki Aoyama (Kyoto Univ.), Thomas Lux (Kiel Univ.), Hiroshi Yoshikawa (Rissho Univ.) and Discussion Paper seminar participants at RIETI.

References

B. Byshkin, A. Stivala, A. Mira, R. Krause, G. Robins, and A. Lomi. Auxiliary parameter mcmc for exponential random graph models. *Journal of Statistical Physics*, 165:740–754, 2016.

A. Chakraborty, H. Krichene, H. Inoue, and Y. Fujiwara. Characterization of the community structure in a large-scale production network in japan. *SSRN, https://dx.doi.org/10.2139/ssrn.2978505*, 2017.

G. Fagiolo and A. Roventini. Macroeconomic policy in dsge and agent-based models redux: New developments and challenges ahead. *Journal of Artificial Societies and Social Simulation*, 20(1), 2017.

O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81: 832–842, 1986.

X. Gabaix. Behavioral macroeconomics via sparse dynamic programming. *NBER Working Paper No. 21848*, 2016.

D. Delli Gatti, E. Gaffeo, and M. Gallegati. Complex agent-based macroeconomics: a manifesto for a new paradigm. *Journal of Economic Interaction and Coordination*, 5(2):111–135, 2010.

R. Gulati and M. Gargiulo. Where do interorganizational networks come from? *Americain Journal of Sociology*, 104(5):1439–1493, 1999.

D.R. Hunter. Curved exponential family models for social networks. *Social Networks*, 29(2):216–230, 2007.

H. Inoue and Y. Todo. Propagation of negative shocks through firm networks: Evidence from simulation on comprehensive supply chain data. *RIETI Discussion Paper*, 17-E-044, 2017.

M.O. Jackson, B.W. Rogers, and Y. Zenou. The economic consequence of social-network structure. *Journal of Economic Literature*, 55(1):49–95, 2017.

A. Kawakami. Diversified firms and productivity in japan. *Public Policy Review*, 13(2), 2017.

H. Krichene, A. Chakraborty, H. Inoue, and Y. Fujiwara. Business cycles’ correlation and systemic risk of the japanese supplier-customer network. *Plos One*, 2017.

A. Lomi and F. Fonti. Networks in markets and the propensity of companies to collaborate: An empirical test of three mechanisms. *Economics Letters*, 114:216–220, 2012.

A. Lomi and F. Pallotti. How to close a hole: Exploring alternative closure mechanisms in interorganizational networks. In *Exponential random graph models for social networks: Theory, methods and applications*, pages 202–212. Cambridge University Press, 2013.

A. Lomi and P. Pattison. Manufacturing relations: An empirical study of the organization of production across multiple networks. *Organization Science*, 17(3):313–332, 2006.

M. Loretan. Economic models of systemic risk in financial systems. *North Americain Journal of Economics and Finance*, 7(2):147–152, 1996.

D. Lusher, J. Koskinen, and G. Robins. *Exponential random graph models for social networks: Theory, methods and applications.* Cambridge University Press, 2013.

K. Nakajima, Y-U. Saito, and I. Uesugi. Localization of interfirm transaction relationships and industry agglomeration. *REITI Discussion Paper Series 12-E-023*, 2012.

G. Robins, T.A.B. Snijders, P. Wang, M. Handcock, and P. Pattison. Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29:192–215, 2007.

Y-U. Saito. Supply-chain vulnerability: An analysis of the impact of earthquakes using micro data. URL www.rieti.go.jp/en/columns/v01_0011.html.

Y-U. Saito. Role of hub firms in geographical transaction network. *RIETI Discussion Paper 13-E-080*, 2013.

T.A.B. Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3:1–40, 2002.

T.A.B. Snijders, P.E. Pattison, G.L. Robins, and M.S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36:99–153, 2006.

Y. Song and M. Van Der Schaar. Dynamic network formation with incomplete information. *Economic Theory*, 59:301–331, 2015.

A.D. Stivala, J.H. Koskinen, D.A. Rolls, P. Wang, and Garry L. Robins. Snowball sampling for estimating exponential random graph models for large networks. *Social Networks*, 47:167–188, 2016.

S. Thiemichen and G. Kauermann. Stable exponenial random graph models with non-parametric components for large dense networks. *Social Networks*, 49:67–80, 2017.

Appendix A: The endogenous network statistics

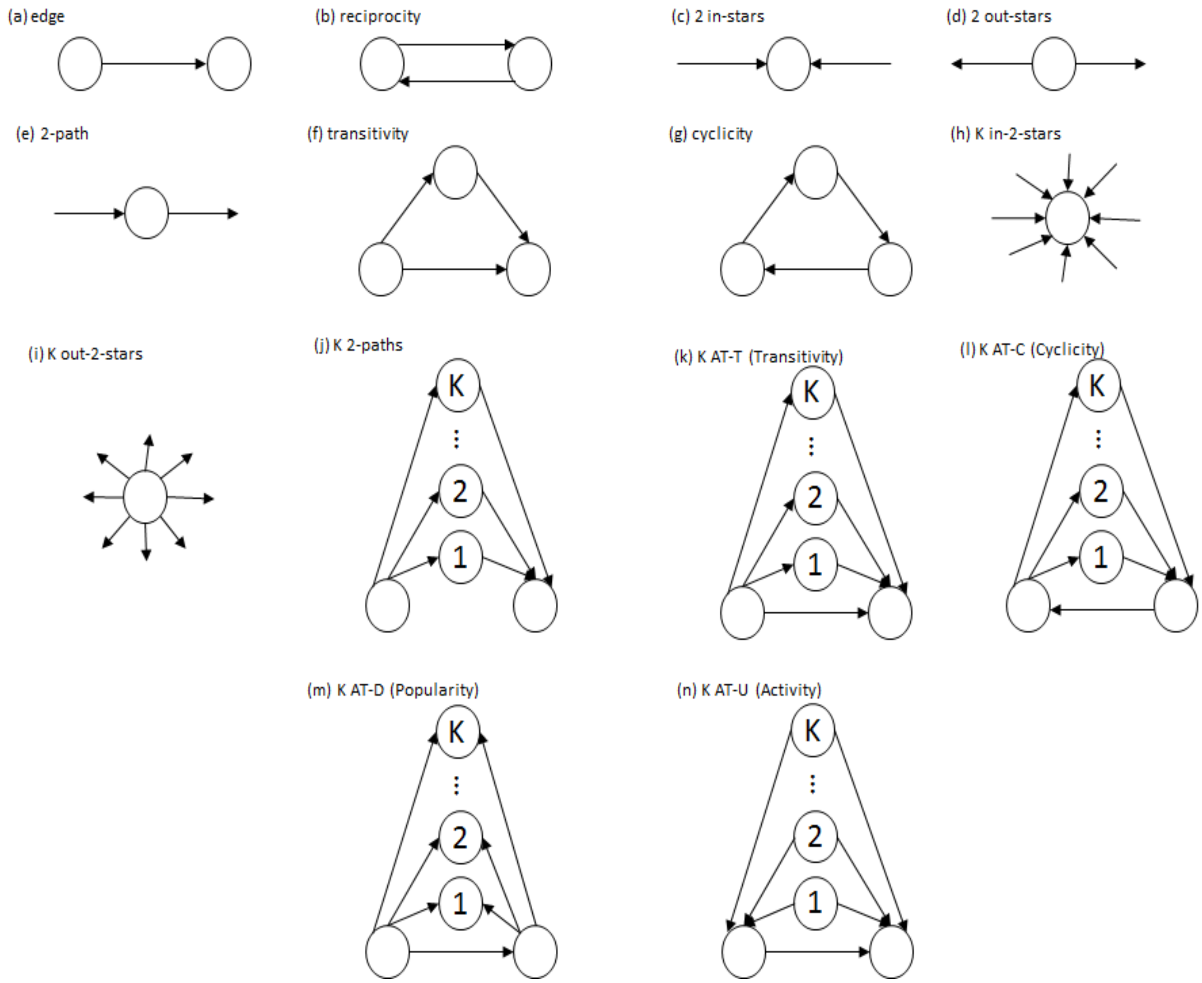

Fig. 10. The different considered network statistics. These endogenous attributes represent the Bernoulli model, the Markov graph model and the social circuit model. Deep explanations could be found in Lusher et al. [2013]