Practicality evaluation of stochastic model on networks for the real economy simulation[*]

Hiromitsu Goto[1], Wataru Souma[1], and Yuji Fujita[1,2]

[1] College of Science and Technology, Nihon University, Japan
hiromitsu.goto.phys@gmail.com
[2] Turnstone Research Institute, Japan

Abstract. We investigate the practicality of the Bouchaud-Mézard (BM) model for the real economy simulation using an information of Japanese production network, which contains basic financial information for more than a million firms and several million links of supplier-customer links. We found that the BM model can fit the power-law tail of the sales distribution and partially predict the sales for the individual firm.

Keywords: Stochastic model · Real economy simulation · Japanese production network.

1 Introduction

One of the most important but challenging issue is to realize the simulation of the real economy in real time. This is because the real economy results from the aggregation of the dynamics of micro level numerous agents such as consumers, firms and financial institutions; they form the multi-layered network roughly estimated in Japan hundreds of millions of people, millions of firms and hundreds of banks. Inevitably, considerable computation should be involved to simulate whole system. In addition, it must be time-consuming to take into account interaction with non-economic condition such as social media. Although it might be achieved in coming decades by the supercomputer developments and the application of machine learning, there remains a need for an efficient model that can save the computational cost and provide a real understanding at micro level in terms of network structure.

Initial attempts focused on identifying the cause of macro phenomenon with power law by using stochastic process of micro units. While they can help us obtain a deep understanding of the phenomenon for the system, over-simplistic assumption tends to make this approach an impractical option in cases such as the real economy simulation. However, the practicality of the stochastic model using the real economic data at micro level remains unclear.

The purpose of this study is to revisit the stochastic model on networks and examine its practicality for the real economy simulation. The Bouchaud-Mézard (BM) model is one of the simplest stochastic model on networks, which was introduced to explain the wealth distribution with a power-low tail [1]. Although the application for complex networks have been studied in [2, 3], there is no research to apply the BM model for actual economic networks. Therefore, we investigate the practicality of the BM model for the real economy simulation using an information of firm-level Japanese production network, which contains basic financial information for more than a million firms and several million links of supplier-customer links. The approach we use in this study aims to simulate the firm size such as sales in terms of only network structure.

This paper is organized as follows: first, we introduce the firm-level data of Japanese production network and review the BM model briefly. After fitting the power-law tail of the sales size distribution using the BM model, we check the accuracy of our simulation comparing with the real data for each firm. Finally, we summarize our results and conclude.

2   Method

Since it is difficult to acquire the data of material flow in general, we discuss the flow of money using the information of production network. Therefore, we simulate the distribution process of sales size with the BM model. The BM model includes the interactions of firms which depends on the weighting. In order to investigate the significance of the network structure to the simulation, our weighting is based on only the degree of firm.

2.1   Data: Japanese production network

The data was collected by Tokyo Shoko Research (TSR) on July 2016, which consists two datasets of 'TSR Kigyo Jouhou' (firm information) and 'TSR Kigyo Soukan Jouhou' (firm correlation information). They contain basic financial information for 1,546,583 firms and 5,943,073 links of supplier-customer and ownership links. Although TSR collected the data based on questionnaires for firms, whose the maximum number of suppliers and customer reported is limited to 24, it does not mean that the number of links is limited to 24 because the links of large firms can be complemented by those suppliers or customers. For the sake of the coherence, we followed the manner to define the Japanese production network from TSR datasets used by the previous study [4].

Let us denote a money flow of supplier-customer link as $i \rightarrow j$, where firm j is a supplier to another firm i, and consider Japanese production network as directed graph using TSR datasets. In order to exclude inactive and failed firms, we used an indicator flag for them in the firm information. Note that we eliminated self loop and parallel links which is duplicated links recorded in the data. The network has the giant weakly connected component (GWCC), which

Table 1. Japanese production network structure: sizes of different components.

Component	# of firms (with sales data)	Ratio (%)
GSCC	530,174 (521,419)	49.7
IN	278,880 (274,070)	26.2
OUT	219,927 (216,809)	20.6
TE	37,056 (36,312)	3.5
GWCC	1,066,037 (1,048,610)	100.0
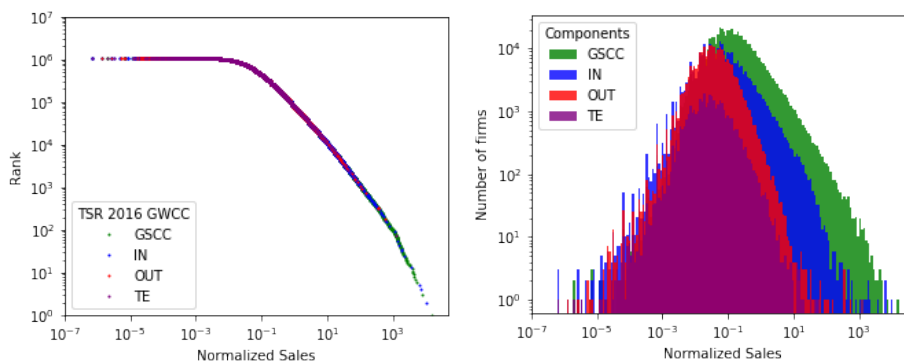
Fig. 1. Rank plot of the normalized sales and the distributions of each component.

corresponds to the largest connected component when viewed as an undirected graph. The GWCC can be decomposed into the parts defined as follows:

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

where the giant strongly connected component (GSCC) is the largest connected one when viewed as a directed graph and the firms which can reach to and from the GSCC via a directed path are named IN and OUT components respectively. The rest of the GWCC are named as tendrils (TE). We show the number of the firms each components in Table 1. It has been known that Japanese production network has the walnut structure [4].

As shown in Fig. 1, the distribution of sales normalized by its mean has power-low tail mainly composed of the GSCC, whose tail exponent can be computed as $\mu = 1.0223 \pm 0.0150$, where $p(x) \propto x^{-(1+\mu)}$ in large x. In general, the simulation on a directed network requires some extension to deal with a dead end as can seen in PageRank [5]. Note that from aspect of the money flow the whole production network is also incomplete for the lack of the workers and consumers. Hence, we only focus on the GSCC hereafter. We show the cumulative distribution function for the in- and out-degree of the GSCC in Fig. 2 (left). As shown in Fig. 2 (right), it is apparent that there have correlation between the degree and the sales size of firm. Since the simulation do not always reflect these feature, we simulate using the weighting based on network structure to confirm.
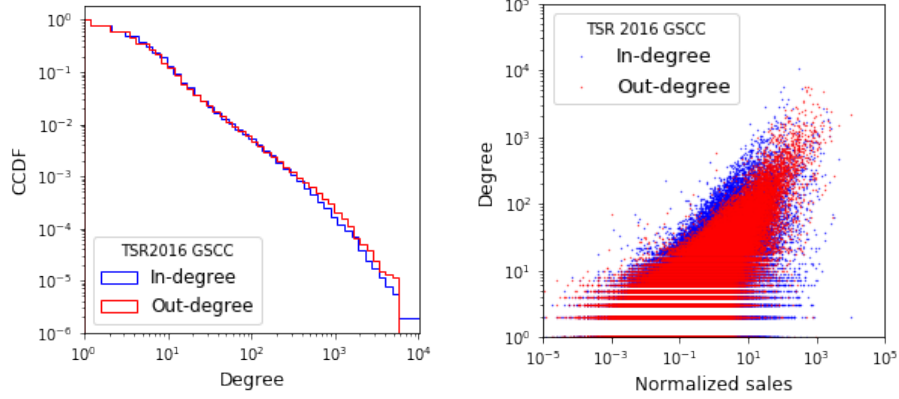
Fig. 2. Left: Complementary cumulative distribution function (CCDF) for the in- and out-degree of the GSCC. Right: Correlation diagram between the normalized sales and degree for the GSCC. The blue and red corresponds to in- and out-degree respectively.

2.2   Simulation with Bouchaud-Mézard model

The BM model consists of the multiplicative stochastic process and the network effect from interactions of firms as follow:

$$X_i(t+1) = a_i(t)X_i(t) + \sum_{j(\neq i)}^{N} \left(J_{ij}X_j(t) - J_{ji}X_i(t) \right) \ , \tag{2}$$

where $X_i(t)$ is the sales size of the i-th firm and a_i is a Gaussian random variable $N(m, \sigma^2)$. The interaction coefficients J_{ij} constitute a dealing matrix from j-th firm to i-th firm. In this work, we assume a dealing matrix as

$$J_{ij} = JA_{ij}/k_j^{\text{out}} \ , \tag{3}$$

where k_j^{out} and A_{ij} is the out-degree of j-th firm and the adjacency matrix, respectively. In our weighting, Eq. (2) can be rewritten as

$$X_i(t+1) = a_i(t)X_i(t) + J\left(\overline{X}_i(t) - X_i(t) \right) \ , \tag{4}$$

where $\overline{X}_i$ is a local mean field of i-th firm, which corresponds to the additive noise from network effects, defined as

$$\overline{X}_i(t) = \sum_{j(\neq i)}^{N} A_{ij}X_j(t)/k_j^{\text{out}} \ . \tag{5}$$

Hence, the all firms distribute $JX_i(t)$ equally to supplier. Moreover, it has been known that the multiplicative stochastic process with additive noise obeys the power-law distribution [6]. Therefore, the additive network effect can be seen as

additive noise in the previous literature to explains the power-low tail. Note that the additive term in our weighting is proportional to its in-degree; $\overline{X}_i(t) \propto k_i^{\text{in}}$.

In order to have a stationary distribution, we use the normalized sales by the mean $x_i(t) = X_i(t)/\overline{X}(t)$ where $\overline{X}(t) = N^{-1} \sum_k^N X_k(t)$. Under the mean approximation in large N limit, which is the case ($k_i^{\text{out}} = N$), the probability density function is known as

$$p(x) = \frac{(\mu - 1)^\mu}{\Gamma[\mu]} \exp\left\{-\frac{\mu - 1}{x}\right\} x^{-(1+\mu)}, \quad \mu = 1 + \frac{J}{\sigma^2}, \tag{6}$$

where the exponent μ is the Pareto index in economics. In general, therefore, the tail exponent of the normalized sales distribution can be fitted by only the parameter σ^2 and J in the BM model. After fitting the power-law tail, we compute the ratio between the simulation result of the BM model and the data for each firm, $r_i = x_i/x_i^{\text{data}}$ to estimate the simulation accuracy.

3   Simulation Results and Discussion

The main purpose of this work was to test the practicality of the simulation of the real economy using the simplified model. As mentioned previously, we sought to explain the sales size of firms by using only the BM model and the information of a real network. Numerical simulation were carried out for the GSCC of Japanese production network and repeated 10 times using Eq. (2) with an initial condition $x_i(0) = 1$, a Gaussian random variable $N(1.01, \sigma^2)$ and various value of J. The distribution at $t = 10^4$ was taken as the stationary distribution.

As shown in Fig. 3, we found that the BM model can fit the power-law tail of the normalized sales distribution with $J = 0.003$, $\sigma^2 = 0.02$ and the ratio between the simulation result of the BM model and the data seems to follow the log-normal distribution, whose central value locates at $r_i = 1$. Comparing the simulated distribution with the TSR data as shown in Fig. 3 (left), it appears that the BM model underestimated the firms located in middle range of sales as lower. Moreover, it is clear that the BM model unfits for the purpose of the simulation required high precision as can be seen from Fig. 3 (right). Although this is not ideal result, it is surprising that the firms simulated with $0.5 < r_i < 2$ accuracy account for 19.8% of whole firms in spite of the stochastic model only used the information of network.

In order to improve our results, it is important to understand what kind of firms the BM model can describe accurately in our weighting. In Fig. 4, we visualized the ratio r_i in terms of the network structure. The white dot corresponds to $r_i = 1$. And it changes to blue (red) as the ratio r_i becomes larger (smaller) than one. As shown in Fig. 4 (left), the ratio increase in a downward direction. It should be noted that the IN and OUT component of the GWCC are ignored in our simulation, which locate above and below of the GSCC. Therefore, this results indicates the possibility of improvement by extension to whole network. As shown in Fig. 4 (right), we found that the simulation with the BM model is better suited for the firms which have larger in-degree. Nevertheless, it does not
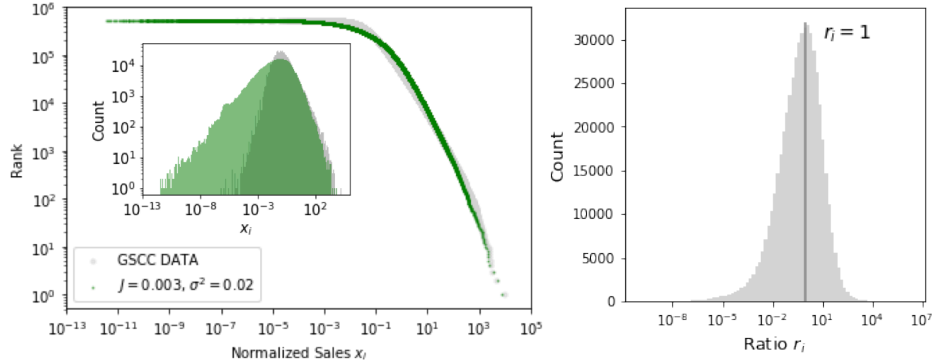
Fig. 3. Left: Comparison of BM model (green) and data (gray) with respect to the distribution of the sales. Right: Histogram of the ratio $r_i = x_i/x_i^{\mathrm{data}}$.

necessarily correspond to the sales. In future, the reasonable results for specific firms might be obtained by choosing the appropriate weight.

Finally, we carried out the same simulation using the randomized network, whose in- and out-degree were conserved. In addition, to compare the network effects from actual and randomized network, we used same random variable $a_i(t)$ for the simulations. As can be seen in Fig. 5 (left), distributions are almost identical in both cases. This is because the additive noise from network effect is determined by the nearest-neighbor interaction in our simulation, which means each additive noise was proportional to own in-degree. It could be inferred therefore that the correlation between the sales and in-degree shown in Fig. 2 (right) may have affected in our simulation directly. We show the correlation between the results from actual and randomized network in Fig. 5 (right), which is gradationally colored by the difference of in-degree. It is evident that the results obtained from both networks have good agreement at large in-degree region. This trend is in line with the previously discussed feature of our simulation.

4   Conclusion

We have investigated the practicality of the simulation of the real economy by using the BM model for the GSCC of firm-level Japanese production network. We found that the BM model can fit the power-law tail of the distribution of sales size and those accuracy defined by the ratio between the result and the real data seems to follow the log-normal distribution, whose central value locates at one. Moreover, we confirmed that the accuracy of our simulation results reflect the correlation between the degree and the sales size. To the knowledge of the authors, this is the first result to evaluate the practicality of the stochastic model by using the real data. Although they seem to unfit for the purpose of the simulation required high precision, this results have potential to improve by changing our weighting.
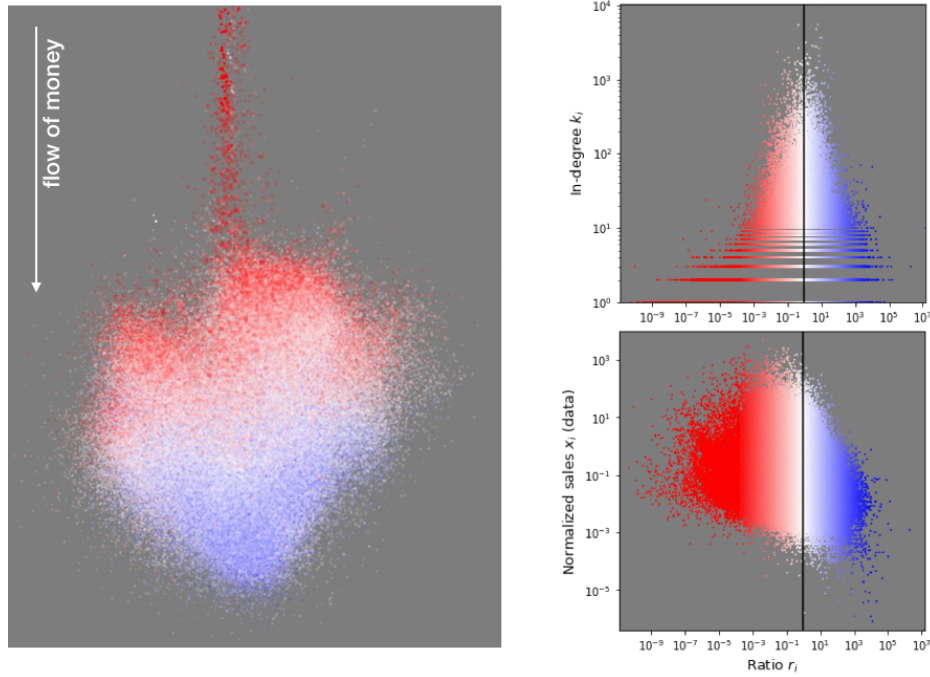
Fig. 4. Visualization of simulation accuracy. The white dots and the black solid lines correspond to $r_i = 1$ and they change to blue (red) as the ratio $r_i = x_i/x_i^{\text{data}}$ becomes larger (smaller) than one.

The real economy simulation in real time can be achieved in coming decades by the development of high performance computing. In order to implement the model on multi-layered network overlapped each others, it is important to partially approximate the system in accordance with intended use; from the central bank's point of view, the interactions of firms can be replaced by the stochastic model. Therefore, our study can provide support for the modeling of the whole system to simulate the real economy.

References

1. J-P. Bouchaud and M. Mézard, Wealth condensation in a simple model of economy, Physica A: Statistical Mechanics and its Applications, Vol. 282, No. 3, pp. 536-545, 2000
2. W. Souma, Y. Fujiwara and H. Aoyama, Small-World Effects in Wealth Distribution, eprint arXiv:cond-mat/0108482, 2001
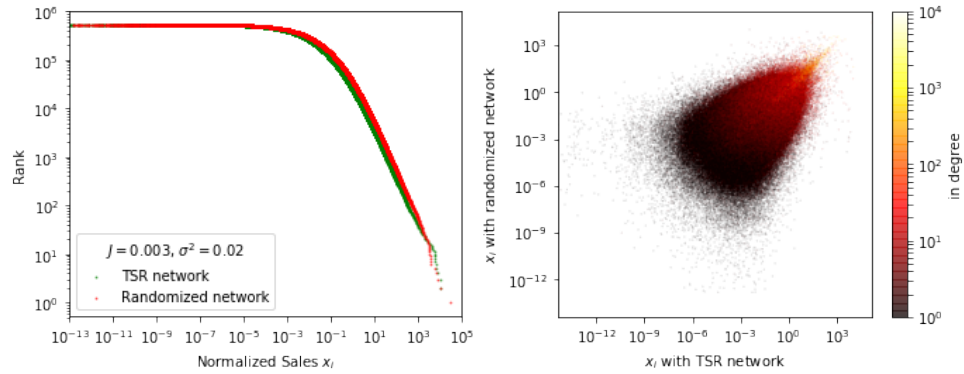3. T. Ichinomiya, Wealth distribution on complex networks, Physical Review E86 , 066115 (2012)

Fig. 5. Comparison of the simulation result between actual and randomized network. Left: Rank plot of the normalized sales simulated by using the BM model with actual (green) and randomized (red) network. Right: Correlation between the results from actual and randomized network gradationally colored by the difference of in-degree.

4. A. Chakraborty, Y. Kichikawa, T. Iino, H. Iyetomi, H. Inoue, Y. Fujiwara and H. Aoyama, Hierarchical Communities in Walnut Structure of Japanese Production Network, RIETI Discussion Paper Series 18-E-026
5. L. Page, S. Brin, R. Motwani and T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web., Stanford InfoLab, 1998
6. H. Takayasu, A. Sato, and M. Takayasu, Stable Infinite Variance Fluctuations in Randomly Amplified Langevin Systems, Phys. Rev. Lett. 79, 966, 1997