

Analysis of Network Robustness for a Japanese Business Relation Network by Percolation Simulation

Hirokazu Kawamoto, Hideki Takayasu, and Misako Takayasu

Abstract This paper describes the application of percolation theory to a Japanese business relation network composed of approximately 3,000,000 links. In this network, we examined the process in which links are randomly removed. At the percolation transition point, we calculate the survival rate for each node as an indicator of its global network connectivity. The basic properties of each node are determined in connection with the values characterising these complex networks, such as the link number and job category. We confirm that this index has strong correlation with degree and shell number, also has significant correlation with sales and number of employee. Finally, we define the network robustness for each prefecture in Japan by using this new indicator.

Hirokazu Kawamoto

Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan.
e-mail: kawamoto@smp.dis.titech.ac.jp

Hideki Takayasu

Sony Computer Science Laboratories, Shinagawa-ku, Tokyo, Japan.

Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, Nakano-ku, Tokyo, Japan.

Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan.
e-mail: takayasu@csl.sony.co.jp

Misako Takayasu

Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan.
e-mail: takayasu@dis.titech.ac.jp

1 Introduction

Percolation theory has been studied in the fields of physics and mathematics. Especially, many interesting properties have been revealed about the percolation transition point at which macroscopic connectivity disappears when removing its elements [1]. Because of the high versatility of this theory, it has been applied to a wide range of real world problems, such as electrical conduction [2] and Internet traffic congestion [3].

Since the BA model was proposed [4], percolation theory has been applied to complex networks with an inhomogeneous structure in connection with the concept of small-world [5]. Studying the percolation process in such complex networks plays an important role from the viewpoint of the fragility of a given system. It is well known that scale-free networks lose connectivity at low density if nodes are removed randomly and at high density if nodes are removed in descending order of the degree [6]. Because these studies can be viewed as a kind of stress test, percolation theory is also important for application study.

In the next section, we explain a dataset composed of about 600,000 Japanese firms and describe its basic properties as a complex network. We present the basic results of our percolation simulation in Section 3. The statistical properties of the survival rate and the theoretical analysis are provided in Section 4. In Section 5, we discuss the network robustness of the prefecture in Japan. Finally, we conclude this study and mention our plans for future work in Section 6.

2 Business Relation Network

The dataset we used in this study was provided by TEIKOKU DATABANK, Ltd., a Japanese credit research company. It included information about the direction of money flow, sales and employees of each firm in operation in 2011. From the point of view of a network study, the dataset provided a complex network consisting of 612,133 nodes and 3,841,496 links. As we were interested in the percolation properties of this network, we ignored the direction of the links and severed so-called dangling bonds, i.e. the bonds that could be removed from the network by the removal of a single link. We then extracted the largest strongly connected component (LSCC) from the raw network [7], and ignored the direction of each of the links for simplicity. As a result of this process, our network was composed of 327,721 nodes and 2,960,370 links. This operation enabled us to reduce the amount of numerical calculation in the following analysis.

Next, we present the basic properties of this network. The link number, namely, degree k , is distributed across a wide range, and this distribution is approximated by a power law for a large degree.

$$F(\geq k) \propto k^{-\alpha} \quad (1)$$

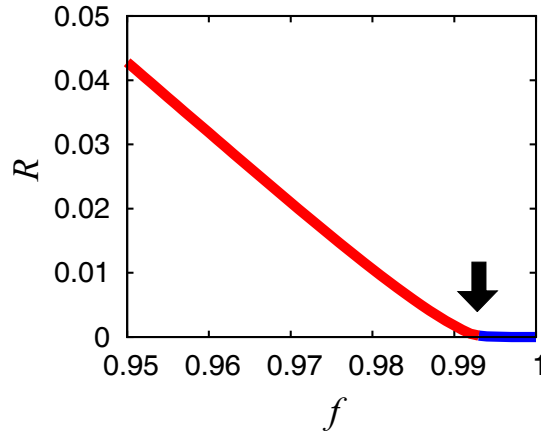


Fig. 1 Largest cluster size R normalised by all links in the range of f between 0.95 and 1.00. The arrow indicates the percolation transition point. The average was taken over 100 trials.

The cumulative exponent α is roughly estimated to be 1.5. Hence, the business relation network is a typical scale-free network [8]. In addition, it also has the small-world property [9].

In this network, we introduce k-shell decomposition, which is a general method intended to reveal the layer structure in a complex network [10]. Application of this method enabled this network to be decomposed into 25 layers, which are also called shells. A shell number is defined for each node, and the number of nodes with the shell number 7 is most numerous [11].

3 Percolation Simulation

A detailed observation of the changes in the network topology became possible when links were randomly removed from the network one by one especially around the percolation transition point. In this case, we did not apply node removal as this could be viewed as a kind of correlated link removal. We calculated the largest cluster size R as an order parameter, which was defined as the ratio of the number of links in the cluster to all the initial links. Here, the control parameter f is the ratio of the number of removed links to all links. As shown in Fig. 1, the order parameter R is sufficiently small for f larger than f_c , which is referred to as the percolation transition point. We estimated the value of f_c as 0.994. Its value is approximately 1, but not exactly 1, and this result is consistent with the findings of previous research in which percolation simulation was applied to a complex network [12]. The properties around this point are discussed in detail from the viewpoint of statistical physics including the finite-size effect [11].

Table 1 Spearman’s rank correlation coefficient between the survival rate and principal parameters (degree, shell number, sales, and number of employees).

	Spearman’s rank correlation coefficient
Degree	0.729
Shell Number	0.765
Sales	0.378
The number of employee	0.356

4 Survival Rate

4.1 Basic Properties of Survival Rate

In this section, we introduce the survival rate for each node and provide its basic properties. At the transition point ($f_c = 0.994$), the survival rate is defined as the ratio of the number of trials, in which the node belongs to the largest cluster, to the total number of trials. In this study, 100,000 trials were performed to estimate the value of the survival rate for each node. This parameter is widely distributed, and its large-scale behaviour approximates a power law [11]. It should be noted that this index is able to characterise the global connectivity of each node in the network as we explain in Section 4.3.

Next, we discuss the correlation between the survival rate and important parameters characterising firms, such as degree, shell number, sales and the number of employees. Spearman’s rank correlation coefficient was chosen for this purpose, because Pearson’s correlation coefficient is susceptible to outliers. As shown in Table. 1, there is a positive correlation between the survival rate and all the parameters, and this is especially strong for values characterising the network, such as the degree and shell number.

The correlation between degree k and the survival rate P_s was investigated in more detail. The variation of the survival rate P_s was clarified by plotting its distribution for degree k as shown in Fig. 2. We found that the survival rate P_s varies even in the same range of degree k . Therefore, the survival rate P_s is not completely determined by information relating to the local connectivity, such as degree k ; the degree k can explain this value roughly. This fact suggests that the survival rate P_s is determined by the critical cluster, which includes the information of the whole network topology. In this sense, this robust index includes information about the global connectivity, such as the shell number, as opposed to local connectivity such as the link number.

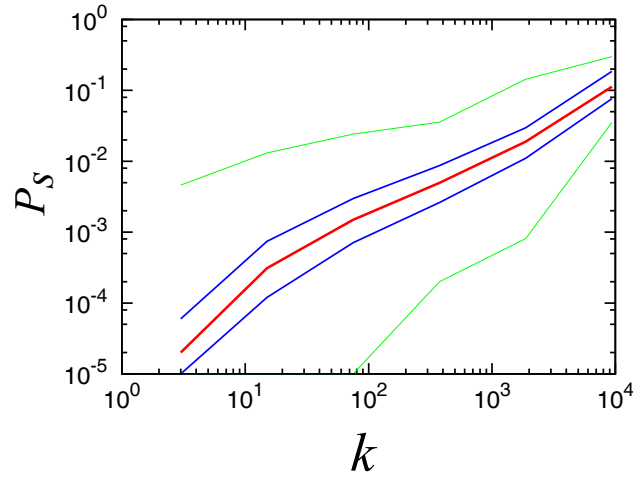


Fig. 2 Degree k vs the survival rate P_s in a log-log scale. Minimum, 1st quartile, median, 3rd quartile and maximum are plotted for log bin. In cases where the representative value was 0, we replaced the observation limit, 1.0×10^{-5} .

4.2 Practical Meaning of P_s

It is important to note that the nodes with the same survival rate P_s were confirmed to have widely distributed link numbers as shown in Fig. 2. We subsequently investigated the features of the nodes with a high survival rate for small link numbers and those with a low survival rate for large link numbers. First, we specified a range of degree k from 1 to 10 as a set of small link numbers within a certain range of the survival rate ($5.0 \times 10^{-4} \leq P_s \leq 5.0 \times 10^{-3}$), which includes approximately 20,000 nodes. When we investigated the industry these nodes represent, it was revealed that the nodes categorised as belonging to the construction industry captured 27% of the share, whereas the share was 21% of the network in its initial state. This result means that nodes belonging to the construction category have a higher survival rate than nodes in other categories.

Next, we focused on nodes with a large number of links within the same range of survival rate ($5.0 \times 10^{-4} \leq P_s \leq 5.0 \times 10^{-3}$). These nodes are characterised by a low survival rate and are fragile in spite of their many links. As an example, we paid attention to the node with large k and relatively small P_s , $(k, P_s) = (448, 3.3 \times 10^{-3})$. As shown in Fig. 3, most of its linking nodes are located on the same island, Hokkaido, and there are only 13 links (about 3%) connecting to firms outside this island. There are not many links connecting to nodes located outside of this island. This result suggests that this type of node bundles firms in a local region.



Fig. 3 Nodes with small survival rate (red circles) and its linking nodes (orange dots) plotted for Hokkaido Island in Japan.

4.3 Theoretical Estimation

A theoretical estimation of the survival rate was derived by using the degree and the rates of linking nodes as follows. In the case of a node that only has one link we have the following exact relation.

$$P_{s,i} = (1 - f_c)P_{s,j} \quad (2)$$

where, the subscript i represents the focusing node, and the subscript j represents the its linking node.

We next extended this formulation to the general case for nodes with multiple links. The probability of the focusing node being connected to the giant component, $P_{s,i}$ is approximated as follows.

$$P_{s,i} = 1 - \prod_{j=1}^k \{1 - (1 - f_c)P_{s,j}\} \quad (3)$$

On condition that the survival rate P_s is sufficiently small, we can approximate Eq. (3) by the following equation.

$$P_{s,i} \simeq (1 - f_c)Q_{s,j} \quad (4)$$

Here, $Q_{s,j}$ is defined as $\sum_{j=1}^k P_{s,j}$. This equation shows that the survival rate P_s is explained by the summation of the survival rates of linking nodes. In Fig. 4, we confirm that this relation is in good agreement. This examination revealed that the survival rate P_s depends on the link number k and the survival rates of the linking nodes, $P_{s,j}$. This result shows that the value $P_{s,i}$ is determined from the global

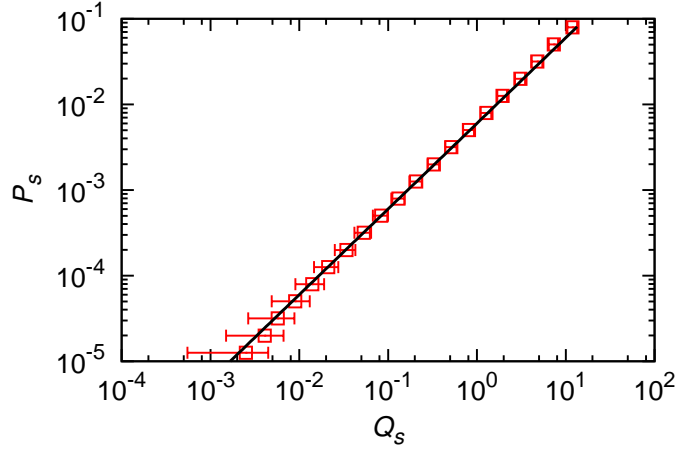


Fig. 4 Summation of the survival rate Q_s of nearest-neighbour nodes vs the survival rate P_s . The average is plotted in each log-scaled bin. Error bars are estimated by standard deviation in log-log scale. The line shows Eq. (4).

network topology, and that the mean field approach used in Eq. (3) works well in deriving Eq. (4).

5 Network Robustness of Prefectures in Japan

By using the survival rate, we define network robustness of each prefecture as follows. We picked up the top 10,000 ranking nodes (practically consisted of 10,005 nodes counting the same ranking) by the order of survival rates, and counted the number of nodes, and normalised this by the original number of nodes in the extracted network for each prefecture. As shown in Fig. 5, Tokyo, Osaka, and seven other prefectures were judged to belong to the most robust class.

As mentioned in section 2 we extracted the LSCC from the raw network when we made the percolation simulation. By this operation, the number of nodes was reduced from 612,133 nodes to 327,721 nodes. In order to check this effect, we re-calculated the network robustness normalized by the original number of nodes in the raw network for each prefecture as shown in Fig. 6. Comparing with Fig. 5, we confirm that changes by this modification are very small, and we find that the eliminated nodes do not affect the results.

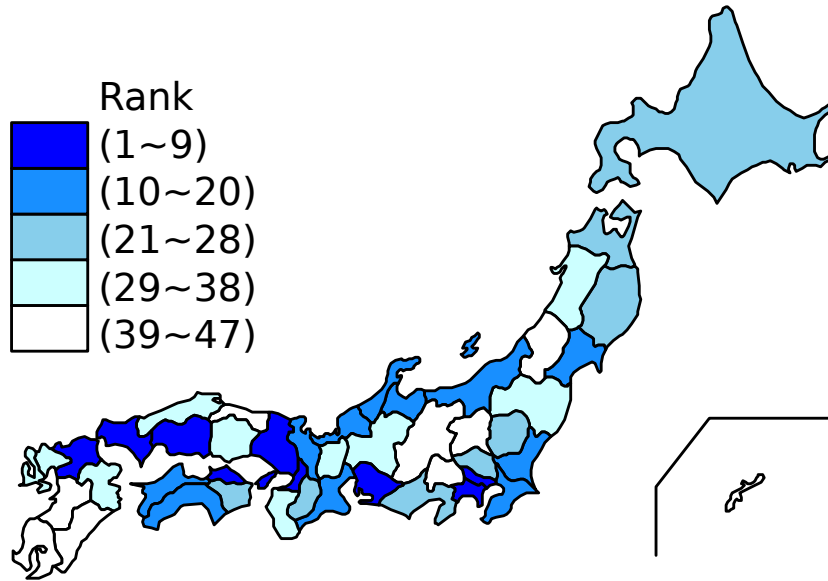


Fig. 5 Network robustness for each prefecture in Japan. Here, the network robustness of each prefecture is estimated by the number of top 10,005 robust nodes located in the prefecture divided by the number of nodes of the initially extracted LSCC network in the prefecture. Colours show the ranking of network robustness categorised into 5 classes, (1 ~ 9), (10 ~ 20), (21 ~ 28), (29 ~ 38), (39 ~ 47) from the deepest to the lightest.

6 Conclusion

This paper discussed the basic properties of survival rate of a business relation network in Japan based on percolation theory. First, we presented the statistical properties of the survival rate by characterising each node as an index measuring the global network connectivity. Values of survival rate are confirmed to be correlated to network connectivity indices such as degrees or shell numbers. However, as shown in Fig. 2, the values distribute widely for the nodes with the same degree number. It is proved in section 4.3 that the survival rate of a node is determined by the sum of survival rates of its neighbor nodes. As the survival rates of neighbors are determined by the next neighbors, and so forth, this value reflects information about wider area's network connectivity.

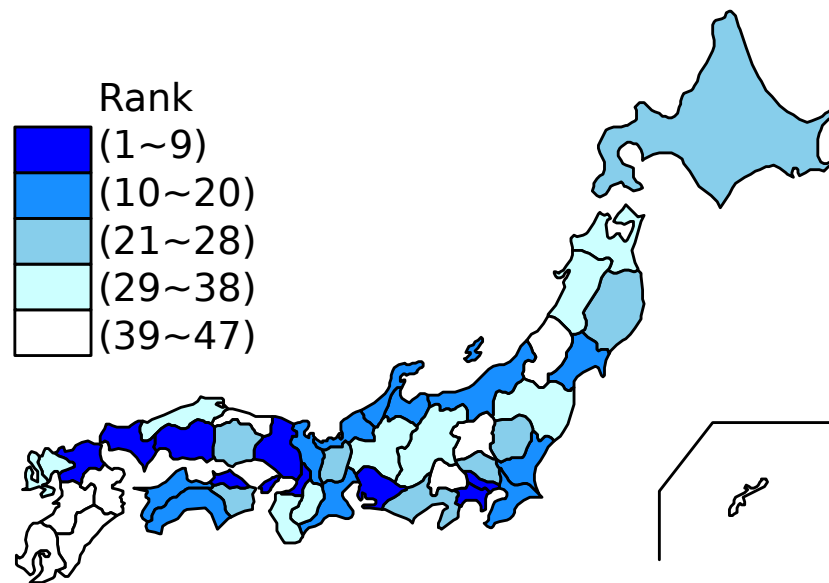


Fig. 6 Network robustness for each prefecture with modified normalization. Compared with Fig.5 normalization by the whole number of nodes in the original raw network is applied. Colours show the ranking of network robustness categorised into 5 classes, (1 ~ 9), (10 ~ 20), (21 ~ 28), (29 ~ 38), (39 ~ 47) from the deepest to the lightest.

We discussed regional differences from the viewpoint of network robustness. The proposed method enabled us to extract those prefectures that were determined to be robust from the viewpoint of complex network science.

This study involved an examination of the network robustness of a Japanese business relation network in 2011. In future, we plan to analyse the time series variation of network robustness paying attention to local economic activities.

Acknowledgements We would like to express our appreciation to Center for TDB Advanced Data Analysis and Modeling, Tokyo Institute of Technology for providing the datasets. This study is partially supported by the Grant-in-Aid for Scientific Research (C), Grant Number 24540395 and (B), Grant Number 26310207.

References

1. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1994).
2. B. J. Last and D. J. Thouless, *Phys. Rev. Lett.* **27**, 1719 (1971).
3. M. Takayasu, H. Takayasu, and T. Sato, *Physica A* **233**, 824 (1996).
4. A. L. Barabási and R. Albert, *Science* **286**, 509 (1999).
5. Duncan J. Watts and Steven H. Strogatz, *Nature* **393**, 440 (1998).
6. R. Albert, H. Jeong, and A. L. Barabási, *Nature* **406**, 378 (2000).
7. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **64**, 025101 (2001).
8. W. Miura, H. Takayasu, and M. Takayasu, *Phys. Rev. Lett.* **108**, 168701 (2012).
9. M. Takayasu, S. Sameshima, T. Ohnishi, Y. Ikeda, H. Takayasu and K. Watanabe, Annual Report of the Earth Simulator Center April 2007-March 2008 ,263-268 (2008) .
<http://www.jamstec.go.jp/esc/publication/annual/annual2007/pdf/2project/chapter3/263takayasu.pdf>
10. M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley and H. A. Makse, *Nature Phys* **29**, 1746 (2010).
11. Kawamoto H, Takayasu H, Jensen HJ, Takayasu M (2015). *PLoS ONE* 10(4): e0119979. doi:10.1371/journal.pone.0119979.
12. J. P. Onnela, J. Saramäki, J. Hyvonen, G. Szabo, M. A. de Menezes, K. Kaski, A. L. Barabási, and J. Kertész, *New J. Phys* **9**, 179 (2007).