Massive Economics Data Analysis by Econophysics Methods - The case of companies' network structure II

Project Representative

Misako Takayasu Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

Authors

Misako Takayasu [*1], Sinjiro Sameshima [*1], Hayafumi Watanabe [*1], Takaaki Ohnishi [*2], Hiroshi Iyetomi [*3], Takashi Iino [*3], Yuki Kobayashi [*3], Kouyu Kamehama [*3], Yuichi Ikeda [*4], Hideki Takayasu [*5] and Kunihiko Watanabe [*6]

*1 Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

*2 Graduate School of Law and Politics, The University of Tokyo

*3 Department of Physics, Niigata University

*4 Hitachi Research Institute

*5 Sony Computer Science Laboratories Inc.

*6 The Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

Network structure of about 1 million Japanese companies is analyzed using the database provided by RIETI. We calculated various quantities using the adjacency matrix representing business interaction among companies. Degree of links, degrees of hub and authority, PageRanks and strongly-connected-components are calculated. It is shown that some of these methods are useful for evaluation of company's value taking into account the effect of network structure.

As a visualization method of the network structure we adopted molecular dynamics by representing each company by a particle having interaction with other particles. At an equilibrium state the configuration gives an intuitively reasonable clustered company network structure.

Keywords: company networks, evaluation of companies, network visualization

1. Introduction

Human economic activity is now considered as one of the main cause of the Earth's global warming, and it is an important first step to clarify detail structure of economy from the real data. For this purpose we have stated analysis of a huge business dealings data of Japanese companies provided by RIETI (Research Institute of Economy, Trade and Industry). As mentioned in the last year's annual report [1] the data consists of about 1 million companies, that is, practically all active companies in Japan. For each company, basic information such as job category, number of employee, annual sales, annual income, names of business partners are listed. The business relation matrix, or an adjacency matrix in the terminology of network analysis, becomes of size about 1 million by 1 million, and we need to analyze a matrix with about 1 trillion elements using the computational power of the Earth Simulator.

2. Company's network properties

Complicated network structures are characterized by various quantities. The most elementary quantity is the degrees of links, that is, simply the number of links which are connecting from a node. In the terminology of economy, this is the number of business partner companies. Based on the direction of money flow there are in-coming-links and outgoing-links for each company. For a given adjacency matrix M, these numbers are given by the number of non-zero elements for each column and row, where the (i,j) component of this adjacency matrix is 1 if the j-th company buys something from the i-th company, or equivalently if there is a money flow from the i-th to the j-th, otherwise the components are 0. The distributions of number of links for both directions are well approximated by power laws with the exponent about 1.3 for link numbers larger than 25 for both directions [2]. By comparing with randomly reconnected artificial networks keeping the link numbers, it is found that there is a tendency that companies with large link numbers are avoiding direct links each other.

The degrees of authority and hubs are defined by the following formulation [3]. Let M be the adjacency matrix, $\vec{a}$

and $\vec{h}$ be the vectors which satisfy the following relation:

$$\vec{h} = M\vec{a} \text{ and } \vec{a} = {}^tM\vec{h} \tag{1}$$

The i-th component of $\vec{a}$ is called the degree of authority of the i-th company, and the components of $\vec{h}$ are called the degrees of hub. As known from Eq.(1) $\vec{a}$ and $\vec{h}$ can be obtained as the maximum eigen vectors of the matrices ${}^tMM$ and $M\,{}^tM$, respectively. The companies having large values of degree of hubs and authorities satisfy a kind of conjugate relation: The degrees of hub are larger if the companies have direct links to those companies having large degrees of authority, and the degrees of authorities are larger if the companies are linked by high hub companies.

As already reported [1] the distribution of degree of authority is characterized by a power law with the exponent about 0.3, and the distribution of degree of hub shows a similar power law in the same small value range, but there is another power law behavior in the large value range with the exponent about 0.5. As known from this result the degrees of hub and authority can characterize companies quantitatively in a different way than the simple link numbers.

Another quantity of network characterization is called PageRank, which is commonly used in characterization of web page ranking such as Google [4]. This quantity is defined by considering an imaginary random walker on the company network, which moves to one of neighbor companies having direct links with the same transition probability when there is at least one outgoing link. In the case there is no outgoing link the random walker jumps to a randomly chosen company from all companies with uniform probability. We can define two types of PageRanks by considering ordinary link direction and the opposite link direction. The ordinary link direction is the direction of money flow and the opposite direction corresponds to material flow which is a conjugate of money flow. The value of steady state probability of this random walker is called the PageRank. It is confirmed that the distribution of PageRank for each direction type, i.e. the money flow or material flow, follows a power law with an exponent very close to that of link number distribution [1].

For these characterizing quantities of the company network, we checked mutual correlations and correlations with company's financial quantities such as sale, income, profit, number of employee and growth rate [2]. In order to avoid error due to statistical outliers we applied Kendall's correlation significance testing by comparing with randomly reconnected artificial networks. From Table 1 it is known that all quantities except growth rates have significant positive correlation, namely, for large companies almost all quantities are large. As for growth rate of profit only PageRanks have non-negligible correlation.

In the study of directed network structure a set of nodes are said to belong to the same strongly-connected-compo-

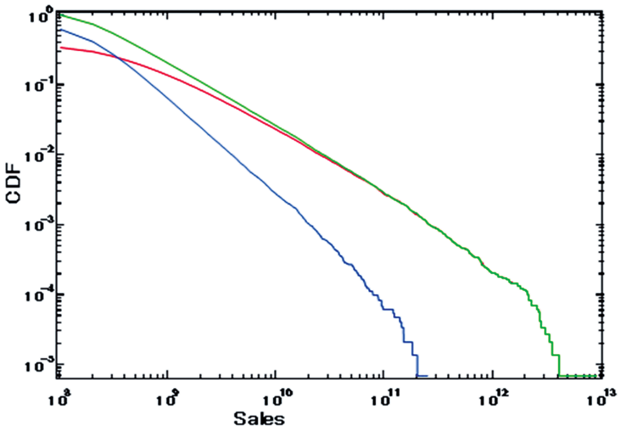

Fig. 1 Cumulative distribution function of sales of companies in manufacturing category. All companies (green), the largest strongly-connected-components (red) and the rest of companies.

Table 1 Correlation among characterization quantities. "indegree" and "outdegree" denote number of links for incoming and outgoing links. "p_a" and "p_h" show PageRanks for money-flow and material flow, respectively. "a" and "h" are the degrees of authority and hub. The last three "r"s are growth rates for sale, income and profit. The signs * show statistical significant level at p-value < 10^{-8}.

	indegree	*outdegree*	p_a	p_h	a	h
outdegree	0.228 *	–	–	–	–	–
p_a	0.753 *	0.138 *	–	–	–	–
p_h	0.143 *	0.709 *	0.080 *	–	–	–
a	0.665 *	0.103 *	0.528 *	0.052 *	–	–
h	0.157 *	0.622 *	0.090 *	0.469 *	0.093 *	–
sale	0.251 *	0.358 *	0.208 *	0.342 *	0.180 *	0.197 *
income	0.144 *	0.239 *	0.131 *	0.255 *	0.080 *	0.160 *
profit$^+$	0.120 *	0.146 *	0.125 *	0.155 *	0.082 *	0.095 *
employee	0.248 *	0.314 *	0.206 *	0.288 *	0.201 *	–0.070 *
r_{sale}	0.062 *	0.033 *	0.066 *	0.028 *	0.057 *	0.028 *
r_{income}	0.042 *	0.012	0.056	0.012	0.051 *	0.029 *
r_{profit^+}	0.004	0.003	0.026 *	0.008 *	–0.001	0.001

nent when any node in the set is connected to any other node in the set. Strongly-connected-components can be calculated from the adjacency matrix. For manufacturing category the number of companies is 143,628 and the number of largest strongly-connected-components is 50,367, about 35%. As known from Fig. 1 the companies which belong to the largest strongly-connected-elements tend to have larger sales than other companies.

3. Visualization of company networks

For visualization of company networks we apply MD simulation method assuming that companies are represented by particles having interaction. The interaction potential between nodes in the MD simulation is given by

$$\Phi(r_{ij}) = \frac{Q_i Q_j}{r_{ij}} - \frac{1}{2} k_{ij} r_{ij}^2, \tag{2}$$

where Q_i is the Coulomb charge for node i and k_{ij} is the spring constant between nodes i and j. Here we neglect the direction of links (flow of goods or money), and assume that Q_i and k_{ij} take identical values for all of nodes and all of pairs, respectively.

MD simulation for a system consisted of about one million charged particles is not a simple task because of the long-range nature of the Coulomb force; the computational task is of order of million times million without any tricks. To speed up the computation, we implemented the hierarchical tree algorithm [5] into our MD code. The whole space is divided by cells at hierarchical levels. The interactions of a given particle with other remote particles are replaced by the interactions with the cells to which those particles belong. As the distance between particles is larger, cells at higher level are used. On the other hand, the force on the particle is calculated directly for its neighbors residing in the same cell or the adjacent cells at the lowest level. This trick reduces the computational task to $O(N \log N)$ where N is the number of nodes.

We show a steady state structure of the transaction network in Japan obtained by this MD simulation in Fig. 2. We confirmed that the final structure is essentially independent of initial configurations or cooling processes we adopted. In this figure the companies are categorized by the size defined by the number of employee. It is found that large companies are located near the mass center of the particles.

The company configuration is decomposed by the sectors of industry in Fig. 3. As seen from this set of figures we find that each sector has its own characteristic configuration, for example, wholesale industry includes clustered structure while companies distribute rather uniformly in real estate industry.

As an application of visualization of company network we plot the companies which are involved in the scandal over sale of contaminated rice. Here, the trigger company firstly sold contaminated rice in Fig. 4 to a food processing company in the autumn of 2007 and then many companies having business relation are affected step by step in the company network. It is found that the shock is directed to the center of the network, indicating amplified impact on the economic system; actually a number of downstream firms and consumers were affected.

4. Summary

In this project we analyzed the whole network properties of Japanese companies by solving the adjacency matrix of size about 1 million by 1 million. We observed basic quantities of networks, the link numbers in both directions, the degrees of hub and authority, and the values of PageRanks. We calculated mutual correlation among these numbers and also we checked correlation with basic quantities of companies such as sale, income, number of employee, and growth rate. It is found that all of these quantities are positively correlated. An exception is the growth rate of a company, which is nearly independent for any of these quantities except the PageRank. This implies that larger companies and smaller companies are having almost similar chance in their growth, and the network structure is actually effecting slightly.

As the second stage of this research we introduced visualization of company network using MD simulation which is

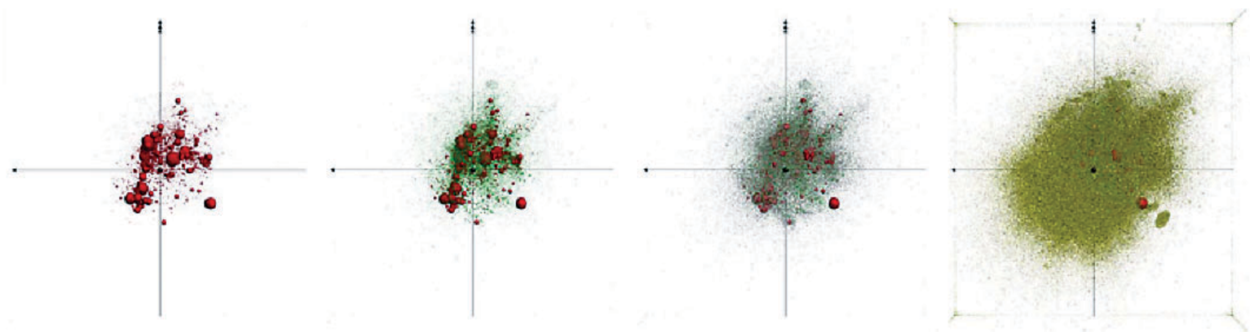

Fig. 2 The configuration of Japanese companies obtained by MD simulation. The radius of each sphere is proportional to the link numbers. (a) Large companies with the number of employee larger than 500. (b) Middle size companies with the number of employee larger than 100 are added. (c) Small size companies with the number of employee larger than 25 are added. (d) All companies.

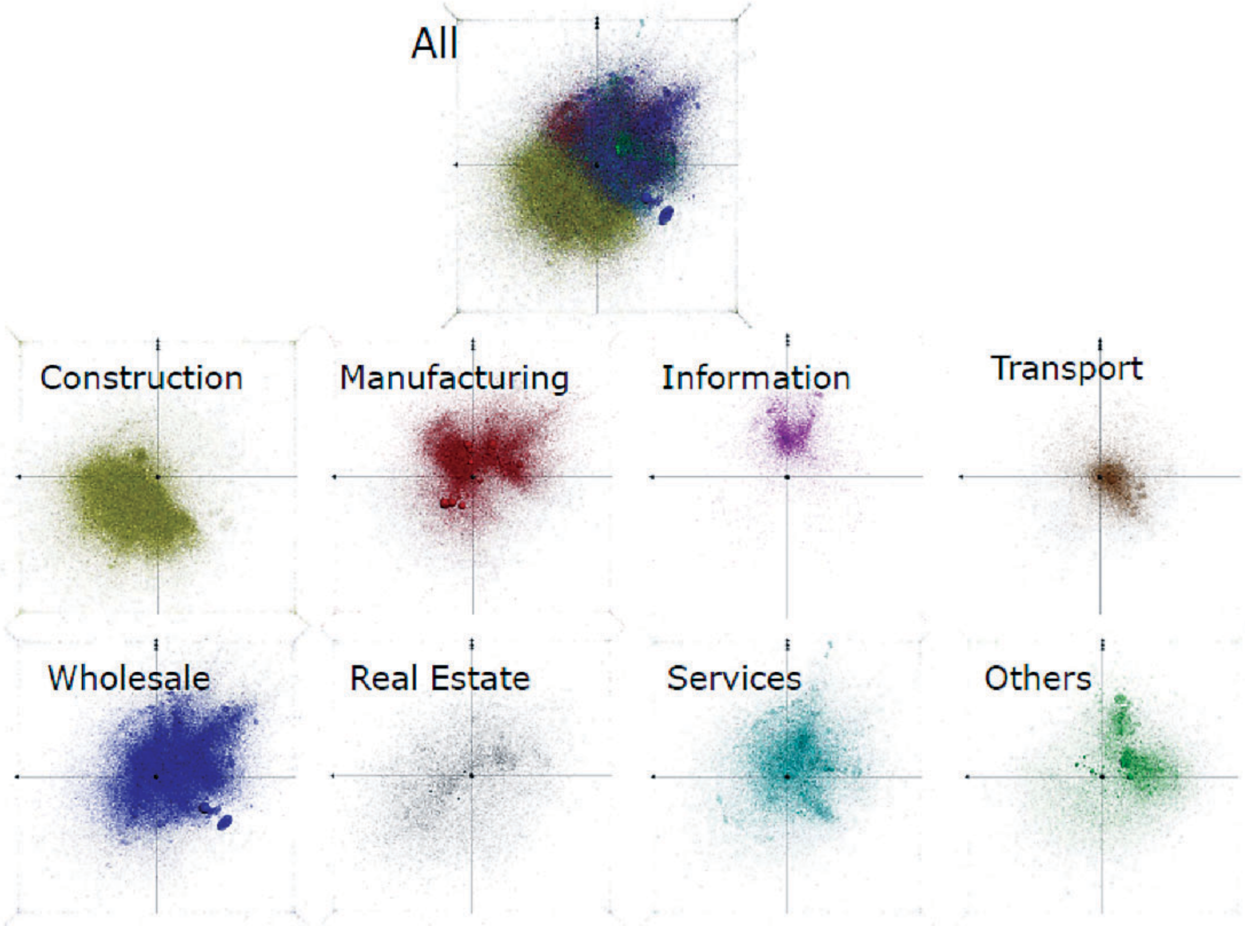

Fig. 3 An optimized network structure of whole economy in Japan (e); only nodes are drawn. And it is decomposed into various sectors: (a) construction, (b) manufacturing, (c) services, (d) real estate, (f) transport, (g) wholesales/retail trade, (h) information/communications, (i) others.

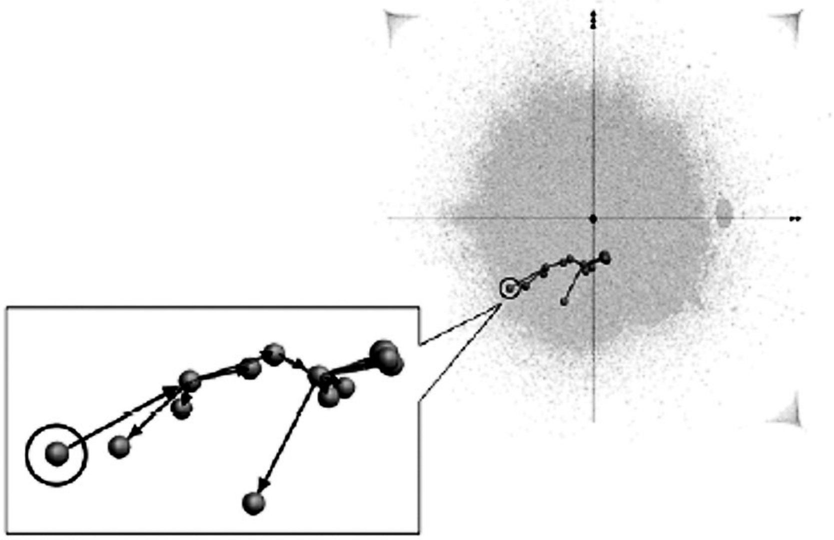

Fig. 4 Propagation of a shock due to the scandal over sale of contaminated rice by a food distribution company through the production network. The circle depicts the trigger company.

suitable for the Earth simulator. The configuration of all Japanese companies is oained, and the structure of each industry is depicted. As an application we checked the case of rice contamination scandal occurred in autumn of 2007. We confirmed cascading shock propagation on the company network toward the center of the whole network. This implies the possibility that the network configuration can be used for prediction of shock propagation in the company network when a big accident occurred on a company.

References

[1] Misako Takayasu et al, Massive Economics Data Analysis by Econophysics Methods - The case of companies' network structure, Annual Report of the Earth Simulator Center April 2007–March 2008, pp.263.

[2] Takaaki Ohnishi, Hideki Takayasu and Misako Takayasu, Hubs and authorities on Japanese inter-firm network: Characterization of nodes in very large directed networks, Progress of Theoretical Physics Supplement, to appear.

[3] Jon M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM, 46 (1999), 604–632.

[4] Sergey Brin and Larry Page, The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, 30 (1998), 107–117.

[5] S. Pfalzner, *Many-Body Tree Methods in Physics* (Cambridge University Press, New York, 1996).

高頻度経済データの経済物理解析II

プロジェクト責任者
高安美佐子　　東京工業大学　大学院総合理工学研究科

著者
高安美佐子[*1]，鮫島伸二郎[*1]，渡邊　隼史[*1]，大西　立顕[*2]，家富　　洋[*3]，飯野　隆史[*3]，小林　祐貴[*3]，亀浜　紘祐[*3]，池田　裕一[*4]，高安　秀樹[*5]，渡邉　國彦[*6]

＊1　東京工業大学大学院総合理工学研究科
＊2　東京大学大学院法学政治学研究科
＊3　新潟大学理学部物理学科
＊4　(株)日立総合計画研究所　経営システムグループ
＊5　(株)ソニーコンピュータサイエンス研究所
＊6　海洋研究開発機構　地球シミュレータセンター

昨年度に引き続き、独立行政法人経済産業研究所提供による日本の企業およそ100万社の取引関係のデータベースを解析し、企業の取引ネットワーク構造の解明を進めた。個々の企業をノードで表し、取引関係のデータから方向性のあるリンクでつながったネットワークの連結性を表現する隣接行列を求め、様々なネットワークを特徴づける量の計算をした。およそ100万 × 100万の成分を持つ行列の解析をするとき、一度に配列を切ることができるコンピュータによって計算の効率を上げることができる。

ネットワークを特徴づける量としては、リンク数分布、ハブ度、オーソリティ度、ページランク、強連結成分抽出を行った。その結果、これらのネットワーク構造を解析する手法が現実の日本の企業ネットワークの解析に有用であることが示された。特に、ページランクに関しては、企業の成長率と有為な相関を持つことが見出された。従来、企業の評価は単独に行われがちであったが、ネットワーク解析を活用することによって、新たな企業価値を評価することができるようになる可能性が切り開かれた。

さらに、複雑なネットワーク構造を可視化する方法として、分子動力学法を適用した。企業を3次元空間中の粒子とみなし、粒子間に仮想的に引力と斥力を仮定したうえで安定な状態を見出すことによって、効率よく類似性の高い企業が集まるような空間配置を数値解析によって得ることができた。このような可視化は、例えば、ひとつの企業に発生したトラブルがネットワーク上をどのように伝播していくのかを予測するような具体的な問題に取り組むための基盤技術となる。

キーワード：企業ネットワーク, 企業評価, ネットワーク可視化