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

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

Authors
Misako Takayasu Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
Sinjiro Sameshima Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
Takaaki Ohnishi Graduate School of Law and Politics, The University of Tokyo
Yuichi Ikeda Hitachi Research Institute
Hideki Takayasu Sony Computer Science Laboratories Inc.
Kunihiko Watanabe The Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

Network structure of about 1 million companies in Japan is studied by analyzing an exhaustive data of trade in Japan. Basic quantities of network structure, such as link numbers, the degrees of hub and authority, and PageRanks are observed. It is found that the network belongs to a typical scale-free network. The averaged number of links of each company is 4.5 and the exponent of the power law distribution is 1.3. Choosing a company as a start point and choosing another company as a destination, the probability that this pair is connected is about 45%, the mean distance is 4.2, the most-likely distance is 5, and the frequency decreases exponentially up to the maximum distance 21. These findings are expected to play an important role in characterization of the economic network's stability and functions.

Keywords: company networks, scale free networks, economic growth, economic stability

1. Introduction

Recently, it is often argued that human economic activity is affecting the Earth's climate such the global warming in relation with the density increase of carbon-dioxide. It is expected as a tomorrow's technology to control the whole economic system to be compatible to the natural environment avoiding a catastrophic recession. From scientific viewpoint the first step of controlling a real complex system is to observe the system in detail such as quantitative characteristics of elements and interactions among the elements. In the case of real economic system the units or elements can be companies and the interactions can be business dealings. In this project we analyze an exhaustive business dealings data of Japanese companies provided by RIETI (Research Institute of Economy, Trade and Industry).

The data we analyze in this project covers 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. In each company's list there are only up to 30 companies as business partners, however, by accumulating all companies' data there are big companies having several thousands of direct business partners. As a result the business relation matrix becomes of size about 1 million by 1million, namely, a matrix with 1 trillion elements is needed. This size of matrix can hardly be treated other than the Earth Simulator.

2. Companies' basic properties

Basic quantities characterizing the size of a company are sales (in unit of 10 million yen), incomes (in unit of 10 million yen), the number of employee and the number of offices. In Fig. 1 cumulative distributions of these quantities are plotted in log-log scale. As known from this figure all of these plots are on straight lines showing that they are approximated by power law distributions. Such scale-free properties for companies are well-known for company size distributions [1].

Network structures of business transactions are analyzed by solving the adjacency matrix of size about 1 million by 1 million. The (i, j) component of the 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. Figure 2 shows the cumulative distribution of link numbers in log-log scale, where the link number is defined by the number of non-zero compo-

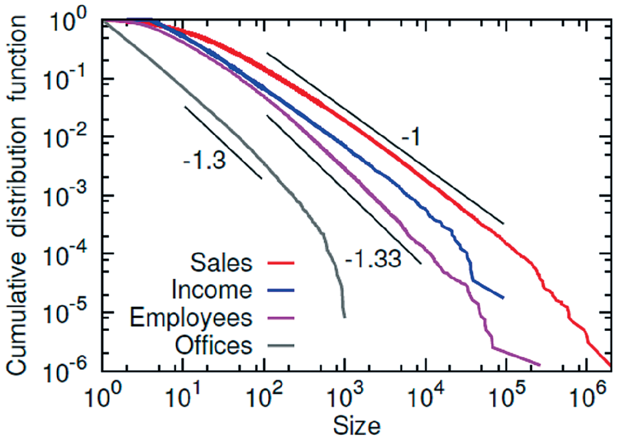

Fig. 1 Cumulative distributions of various sizes of companies. The amount of sales and incomes (in 10 million yen) and the numbers of employee and offices.

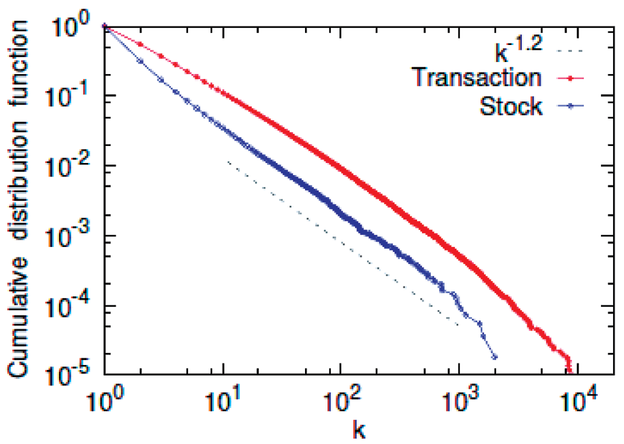

Fig. 2 Cumulative distribution of link numbers for business transactions and stock holders.

nents in the i-th row of the adjacent matrix. The distribution is well approximated by a power law with the exponent about 1.2 for the wide range from about 10 to 1000 [2, 3]. It is shown by comparing with randomly reconnected network keeping the link numbers that there is a tendency that companies with large link numbers are avoiding each other. Also there is a high tendency that the links are bi-directional.

Another adjacent matrix is defined also for stock holders relations, that is, if the i-th company holds the j-th company's stock, then the (i, j) component of the adjacent matrix of stock holder is 1 and otherwise 0. For this stock holder network the link number distribution also follows a power law as shown in Fig. 2. The link numbers of the stock holder network are smaller than those of the transaction network, however, the power law exponents of these two distributions are very close each other.

The distance of any pair of companies can be defined by the minimal number of links which connect these two companies. In Fig. 3 the cumulative distribution of distance of pairs is plotted in semi-log scale. The probability that there

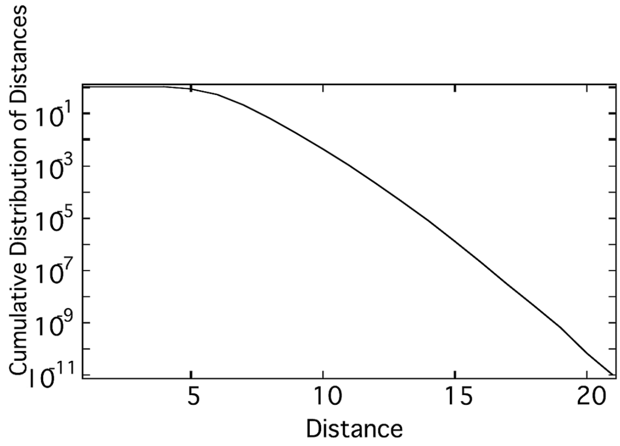

Fig. 3 Cumulative distribution of company pair distances.

exists at least one path from a randomly chosen company to a randomly chosen destination company is about 45%. The peak distance is 5 and the distribution of distance decays nearly exponentially from distance larger than 8 up to the maximum distance, 21.

Next, we define the degrees of authority and hubs which are playing important roles in web-page ranking [4]. Let M be the adjacent 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 as typically shown in Fig. 4. Namely, the degrees of hub are larger if the companies have direct links to those

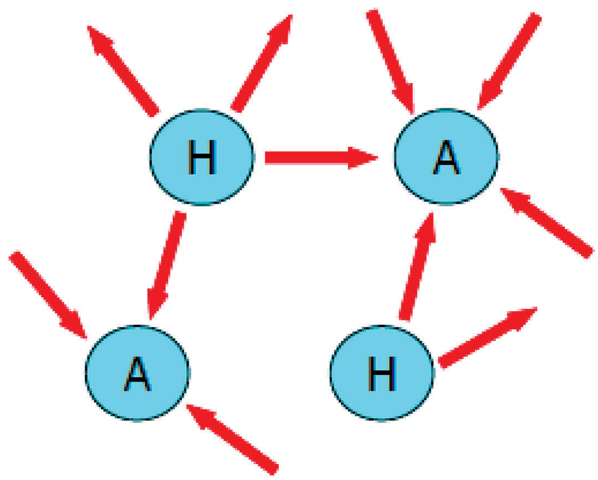

Fig. 4 Schematic relation of hub companies (H) and authority companies (A).

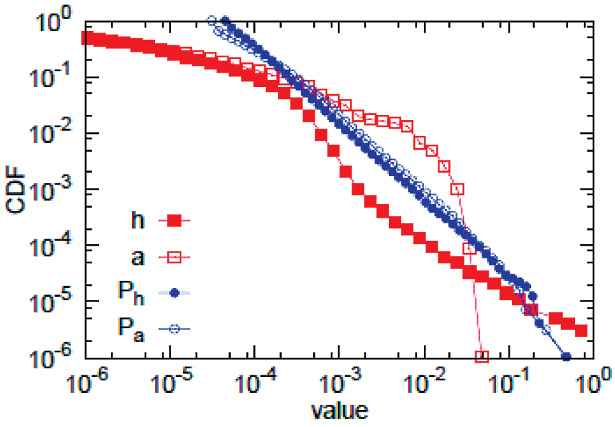

Fig. 5 Cumulative distributions of the degree of hub (h), degree of authority (a), Page ranks for money flow (Ph) and for materials and services flow (Pa).

companies having large degrees of authority, and the degrees of authorities are larger if the companies are linked by high authority companies.

In Fig. 5 the cumulative distributions of degrees of hub and authority are plotted in log-log scale. The distribution of degree of authority is characterized by a power law in the range of [10^{-6}, 10^{-3}] with the exponent about 0.3 and an exponential decay for larger values. On the other hand 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 [10^{-2}, 10^{0}] 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. It is confirmed that there is a tendency that big names of construction companies appear in the companies which have top rank degrees of authority. It is found that the degree of authority and sales are positively correlated. For the degree of hubs big names of retail trading companies tend to appear, but the correlation to the sales and the sizes are less clear than the degree of authority.

As another quantity of characterization of network structures we introduce PageRank, which is used in characterization of web page ranking such as Google. This quantity is closely related to the physical problem of diffusion on the network [5]. Consider 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. In the case of real company network the steady state probability distribution of this random walker, called the PageRank, is not uniform but the density distribution follows a power law distribution as shown in Fig. 5. Here, we consider two cases, one is the case that the random walker moves in the direction of money flow, and the other is the case that it moves in the opposite direction of money, namely, the direction of material or service. In both cases the PageRank distributions follow power laws with the exponent about 1.4 as shown in Fig. 5. The PageRanks can also characterize companies quantitatively from another angle.

3. Graphical representation of company networks

Graphical representation is very important for intuitive understanding of the complex network structure. However, this is not an easy job as the number of companies is too big. In Fig. 6 the network structure of banks is plotted. Even for this small category of companies the network structure is so complicated that little information can be retained from the figure intuitively.

Figure 7 is an example of network structure for companies in Kagoshima-prefecture. Here, small size companies measured by sales are neglected, companies with large values of authority are represented by squares, large hub companies are represented by ellipses, and colors specify job categories. It is known from this figure that two retail sales companies are playing the roles of hubs in this region, and several construction companies are making networks superposing on the whole networks. Other companies can be viewed as satellites, however, if we zoom up a satellite company it generally links many sub-satellite companies. This type of detail structure of company network depends on the region, size and job categories.

Figure 8 is an aggregated network structure representing interaction between job categories in Japan. Here, the whole companies are divided into 17 job categories and the interaction between categories are defined in the following way: To

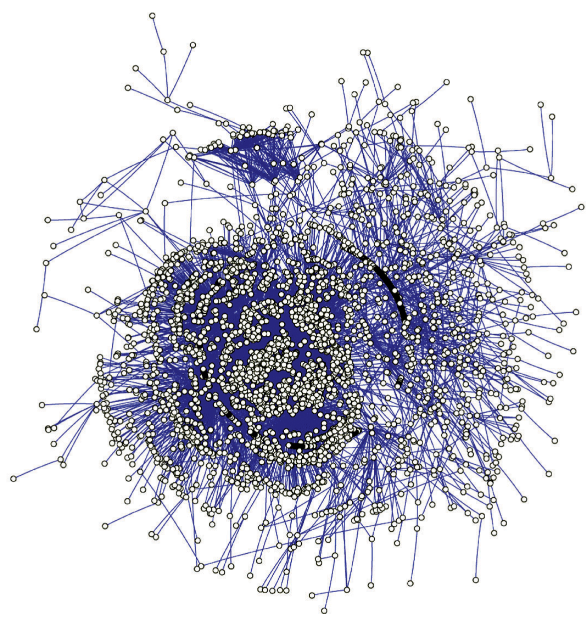

Fig. 6 An example of bank network in Japan shown by non-directional links.

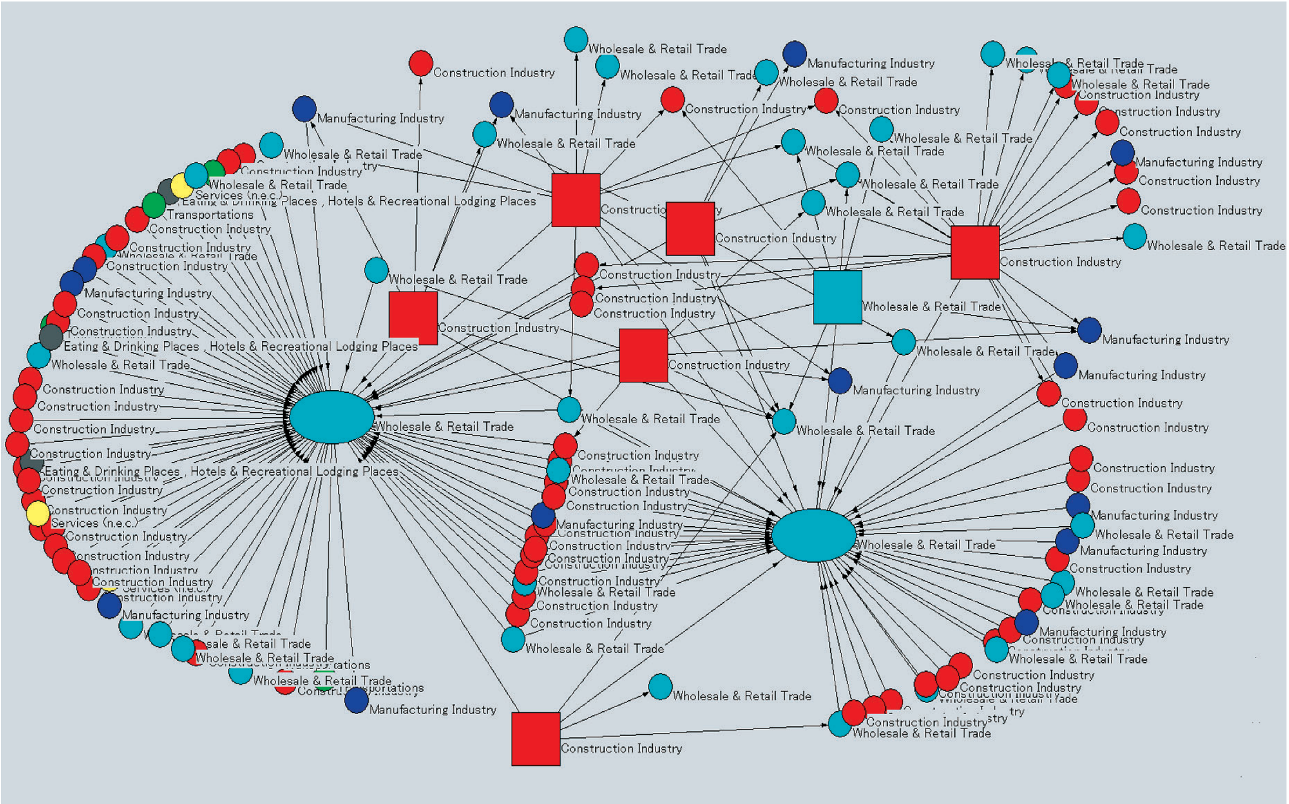

Fig. 7 Network structure of larger companies in Kagoshima-prefecture. Companies with high degree of authority are shown by ellipses, high hub companies by squares. Colors show job categories: Red; construction industry, Light blue; Wholesale and retail trades, Blue; Manufacturing industry, Gray; Restaurants and hotels, Green; Transportation, Yellow; Services.

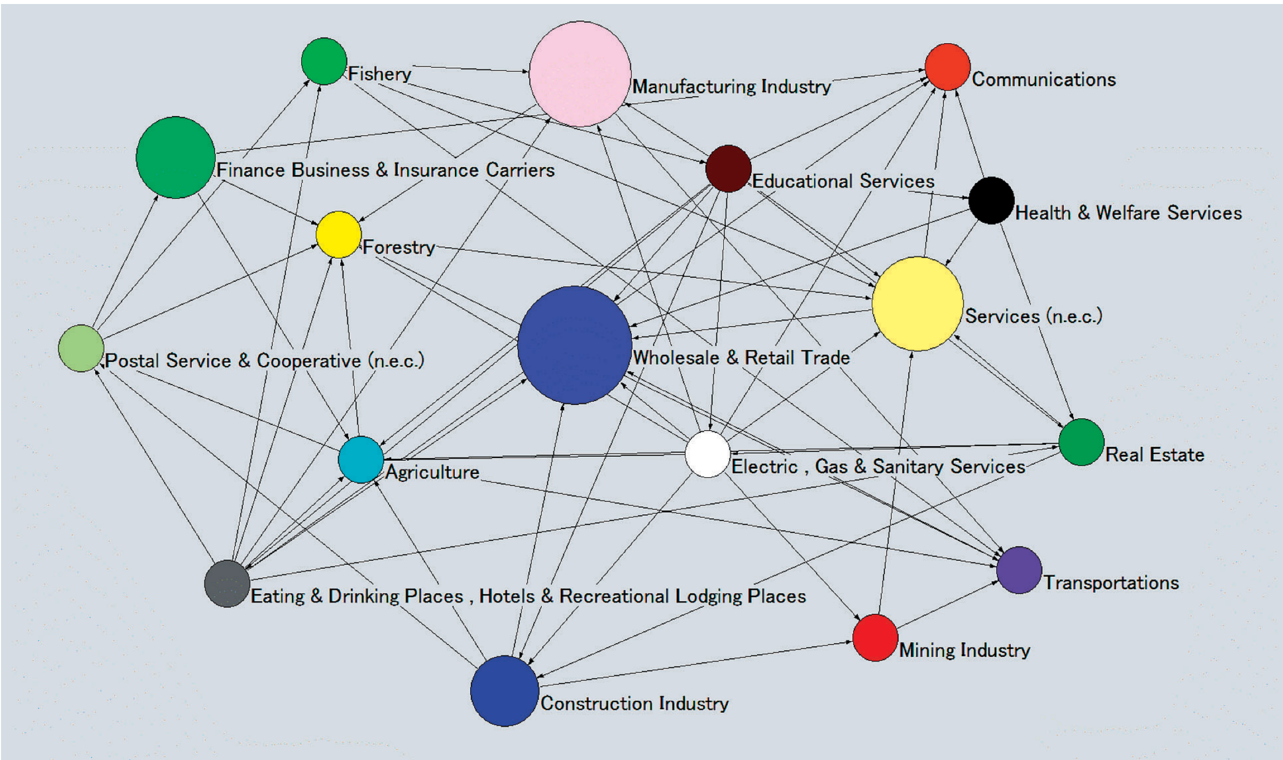

Fig. 8 Network structure of job categories in the whole Japan.

define the interaction from category A to category B, we count up the number of individual links each connecting a company in the category A to a company in category B. This number is normalized by the product of the numbers of companies in these two categories. Here, the top 60 links between categories are drawn in this job category networks. We can find aggregated interaction among companies from this figure.

4. Summary

In this project we analyzed the whole network properties of Japanese companies by solving the adjacent matrix of size about 1 million by 1 million. We observed basic quantities of networks, such as the link number distributions both for transaction network and for stock-holder network, the degrees of hub and authority and PageRanks. In any case the network shows fractal or scale-free properties. The distance of any pair of companies are calculated from the adjacent matrix and found that the peak distance is 5, the maximum distance is 21, and the distance distribution decays nearly exponentially. Visualization of company network is a challenging task and we introduced two new types of network representation. One is using the degrees of hub and authority which give intuitively consistent information to the network figure. The other is the category interaction network which aggregates many companies in the same category and counts the number of individual links between categories. By this method the global industrial interactions can be graphically represented.

References

[1] K. Okuyama, M. Takayasu, and H. Takayasu, Zipf's law in income distribution of companies, Physica A, 269(1999), 125–131.
[2] Takaaki Ohnishi, Hideki Takayasu and Misako Takayasu, Hubs and authorities on Japanese inter-firm network: Characterization of nodes in very large directed networks, preprint.
[3] Yoshi Fujiwara, Hideaki Aoyama, Wataru Souma, Large-scale structure of a nation-wide production network, in preparation.
[4] Jon M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM, 46 (1999), 604–632.
[5] Sergey Brin and Larry Page, The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, 30 (1998), 107–117.

高頻度経済データの経済物理解析

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　　東京工業大学　大学院総合理工学研究
鮫島伸二朗　　東京工業大学　大学院総合理工学研究
大西　立顕　　東京大学　大学院法学政治学研究科
池田　裕一　　(株)日立総合計画研究所　経営システムグループ
高安　秀樹　　(株)ソニーコンピュータサイエンス研究所
渡邉　國彦　　海洋研究開発機構　地球シミュレータセンター

日本の企業およそ100万社の取引関係のデータベース(独立行政法人経済産業研究所提供)を解析し、企業の取引ネットワーク構造を定量的に分析した。このデータは、ネットワークとしてはお金の流れる方向を表す有向グラフのネットワークを表している。任意の企業を選んだときに、そこから連結した企業数の分布はベキ分布にしたがっており、企業ネットワークは、様々な分野で普遍的に観測され注目を集めているスケールフリーネットワークとして分類される。インターネットのサイトのランキングを決める際に利用されているページランクという量は、ネットワーク上をランダムに移動する仮想的な粒子の存在確率として定義されるが、この量も企業ネットワークではベキ分布にしたがっている。また、インターネットのウェブのネットワーク解析で導入されているハブ度やオーソリティ度を計測し、企業を評価する上で有効な新しい指標となることがわかった。任意に選んだ企業から別の任意に選んだ企業、約1兆通りの組み合わせのうち、連結しているものはおよそ45%であり、連結している場合の平均距離は4.2リンク、最多距離は5リンクでリンク数とともに確率は減少し、連結している中で最も離れている場合は21リンクであることもわかった。

キーワード：企業ネットワーク, スケールフリーネットワーク, 経済成長, 経済的安定性