J Econ Interact Coord (2010) 5:171–180
DOI 10.1007/s11403-010-0066-6

REGULAR ARTICLE

Network motifs in an inter-firm network

Takaaki Ohnishi · Hideki Takayasu · Misako Takayasu

Received: 20 August 2009 / Accepted: 24 May 2010 / Published online: 6 June 2010

Abstract Firms build business relationships during economic activities. The goal of this paper is to clarify production mechanisms and economic functions by identifying characteristic patterns of inter-firm interactions. In this paper, we empirically analyze an inter-firm network consisting of about one million firms and four million directed links, in order to specify network motifs, which are small subgraphs that occur more frequently than expected in a randomly generated network. We found that V-shaped triads are network motifs, while feedforward and feedback loop are anti-motifs. By defining roles in the subgraph according to structural equivalence, we also detected the significance profile of roles characterizing the industry sector. The taxonomy of industries obtained from the profiles is economically meaningful. These empirical findings may serve to provide an easily interpretable view of the entire inter-firm network and to improve the efficiency and safety of economic systems.

T. Ohnishi (✉)
The Canon Institute for Global Studies, 11F, Shin-Marunouchi Bldg.,
1-5-1 Marunouchi, Chiyoda-Ku, 100-6511 Tokyo, Japan
e-mail: ohnishi.takaaki@canon-igs.org

T. Ohnishi
Graduate School of Economics, The University of Tokyo, 7-3-1 Hongo,
Bunkyo-Ku, 113-0033 Tokyo, Japan

H. Takayasu
Sony Computer Science Laboratories, 3-14-13 Higashigotanda,
Shinagawa-ku, 141-0022 Tokyo, Japan

M. Takayasu
Department of Computational Intelligence and Systems Science,
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology, 4259-G3-52 Nagatsuta-cho,
Midori-ku, 226-8502 Yokohama, Japan

1 Introduction

To understand a firm's economic activity, it is important to investigate not only the firm itself but also the structure of interactions on an inter-firm network. Nevertheless, it is only recently that the empirical study of the structure of inter-firm network has been conducted on a nation-wide scale (Saito et al. 2007; Fujiwara and Aoyama 2008; Ohnishi et al. 2009).

Traditional measures for analyzing complex networks, such as the clustering coefficient and the shortest path length have been widely employed to clarify the structural properties. However, they are not always useful for analyzing networks, because these measures ignore link directions and are intended for undirected networks. In particular, to understand the economic system better, it is necessary to take into account the direction of interaction, such as the flow of money, materials, and services. In fact, in the previous paper we showed that Google's PageRank algorithm, which measures the importance of each node in the directed network, is useful for characterizing the property of a firm (Ohnishi et al. 2009).

Complex networks have numerous patterns of connections. Subgraphs that occur significantly more often in the real network than in randomized networks are referred to as motifs, while those occurring less frequently are anti-motifs. Recently, network motifs have attracted attention as a tool for studying directed networks (Milo et al. 2002, 2004). In biological networks, a small set of network motifs appears to serve as basic building blocks. Such motifs are believed to perform specific functional roles.

This paper aims to gain some insight into the topological properties of inter-firm network by quantifying the basic network motif in large amount of economic data.

2 Data

We conducted an empirical study on Japanese inter-firm network of the year 2005, which is provided by Tokyo Shoko Research Ltd. (TSR). The network is defined as follows: firms are represented by nodes and directed links mean transactions. This data consists of transactions of materials and services. Financial transactions such as loan contracts and purchases of securities are not included. The directed link stands for flow of money, and the direction of link is from customer to supplier. The opposite direction shows flow of materials and services. The network consists of $961,318$ firms and $3,667,521$ directed links. Note that relationships with general consumer, government, or foreign firms are not included in this data. Accordingly, our study focuses on the interrelationships among Japanese firms. This data also contains basic information on each firm on job category; annual sales, incomes, and profits; number of employees; and so on.

3 Basic structure of an inter-firm network

In this study, we classify the directed links into three types: inlink, outlink, and mutual link. Inlink refers to the number of incoming links of a node, whereas outlink is the number of outgoing links. Mutual link is the number of double links with both

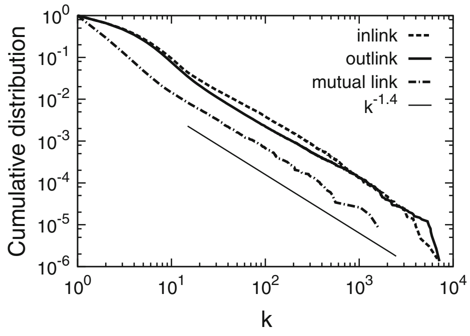


Fig. 1 Cumulative probability distributions for inlink, outlink, and mutual link. Power-law fits using the maximum likelihood method and the Kolmogorov-Smirnov statistic Clauset et al. (2007) yield $\alpha = 1.371 \pm 0.025$ ($x_{\min} = 87$) for inlink, $\alpha = 1.249 \pm 0.028$ ($x_{\min} = 85$) for outlink, and $\alpha = 1.350 \pm 0.076$ ($x_{\min} = 34$) for mutual link

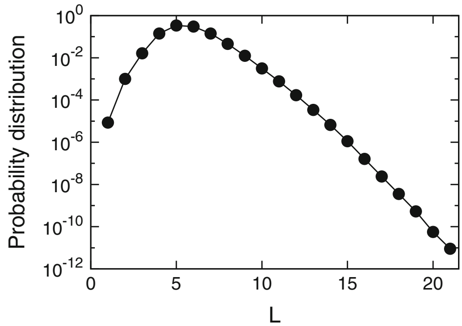


Fig. 2 Probability distribution of distance L decays nearly exponentially. The average distance is 5.62 and the maximum distance is only 21

incoming and outgoing links. Figure 1 shows the cumulative distribution of each type of link, that is, the probability that a value is larger than a threshold x, $P(> x)$. As reported in the previous findings (Saito et al. 2007; Fujiwara and Aoyama 2008; Ohnishi et al. 2009), every type of link follows the power law distribution

$$P(> x) \propto x^{-\alpha}$$

for $x > x_{\min}$. Best fits of power-law form using the maximum likelihood method and the Kolmogorov-Smirnov statistic (Clauset et al. 2007) result in $\alpha = 1.371 \pm 0.025$ ($x_{\min} = 87$) for inlink, $\alpha = 1.249 \pm 0.028$ ($x_{\min} = 85$) for outlink, and $\alpha = 1.350 \pm 0.076$ ($x_{\min} = 34$) for mutual link.

The distance between any two nodes is defined by the minimal number of links connecting a pair of nodes. As shown in Fig. 2, the distribution of distance decays exponentially. The average distance is 5.62 and the maximum distance is only 21, which indicates that the inter-firm network forms a small-world network, as seen in social or biological systems.

For undirected networks, the clustering around node i is quantified by the clustering coefficient C_i. The clustering coefficient is defined as the number of triangles containing node i, normalized by the maximum possible number of such triangles,

$$C_i = \frac{2t_i}{k_i(k_i - 1)},$$

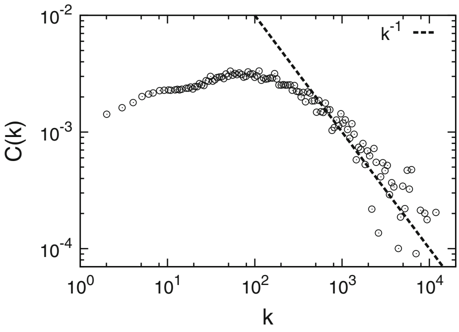


Fig. 3 Clustering coefficient $C(k)$ as a function of the node degree k. After a small increase, it decays as a power law, illustrating the hierarchical architecture

where t_i is the number of triangles around node i, and k_i is the degree of the node. To measure the clustering coefficient, we made the network undirected here. Figure 3 shows the clustering coefficient $C(k)$ as a function of the node degree k, defined as the average clustering coefficient of all nodes with the same degree. The obtained $C(k)$ function shows a small increase with an increasing k up to $k \sim 10^2$, followed by power-law decay with the exponent about 1. This indicates a hierarchical architecture in which higher the degree of a node, lower its clustering coefficient. The same scaling behavior, $C(k) \sim k^{-1}$, has been observed in many real networks (Ravasz and Barabási 2003).

In directed network, a variety of definitions of the clustering coefficient are possible, because there is more than one way of connecting three nodes: the cycle, feedforward loop, feedback loop, and so on. The motif analysis provides not only the detection of building-block patterns, but also the clustering measure in a directed network.

4 Motif analysis

In this study, we focus on three-node motifs. In order to connect three nodes with directed links, there are 13 possible three-node subgraphs as illustrated in Fig. 4. For the sake of detecting significant motifs, we calculate the number of appearances of these subgraphs in real and randomized networks (Itzhack et al. 2007). For each subgraph i, the statistical significance is evaluated by the Z score:

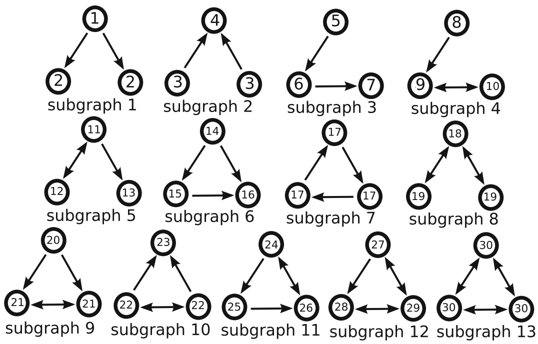


Fig. 4 All 13 types of three-node connected subgraphs with labels indicating the roles of the nodes. Each subgraph has from one to three roles. We classify every node into 30 types of roles

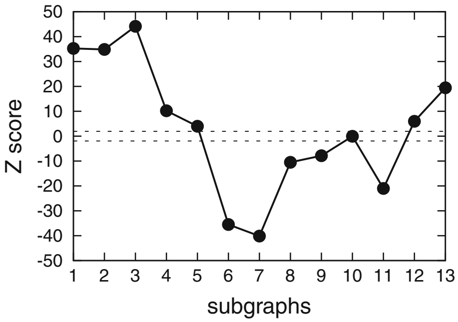


Fig. 5 Significance profile of subgraphs. The *horizontal broken lines* denote two-sided 5% critical values. V-shaped triads (subgraph 1, 2, and 3) are frequently occurring patterns

$$Z_i = \left(N_i - \left\langle N_i^{\text{rand}} \right\rangle\right) \Big/ \sigma_i \ ,$$

where N_i is the number of times the motif i appears in the network, $\left\langle N_i^{\text{rand}} \right\rangle$ and σ_i are the mean and the standard deviation of its appearances in the randomized networks respectively. The Z score is defined as the number of times the real network exceeds the randomized networks normalized by standard deviation.

For a stringent comparison, we use degree-preserving randomized networks that preserve not only the total number of nodes and links, but also the number of inlinks, outlinks, and mutual links for each node in the real network. Since these randomized networks have the same single-node characteristics as the real network, the number of appearances of two-node subgraphs is the same in both networks. This ensures that the detected significant deviation is independent of significant subpatterns. We generate $1,000$ different randomized networks by applying the Markov-chain Monte Carlo switching algorithm to a real network in which randomly chosen pairs of link are repeatedly switched (e.g., X1→Y1, X2→Y2 are replaced by X1→Y2, X2→Y1) until the network is well randomized (Milo et al. 2003).

Figure 5 shows the Z score of the 13 possible connected subgraphs (significance profile). V-shaped triads (subgraph 1, 2, and 3), which are common motifs in word-adjacency networks and bipartite model networks, are network motifs. This is related to the structure of inter-firm relationships in which a firm tends to transact with other firms in a different business category. Furthermore, clique (subgraph 13), in which all three firms are mutually connected, also plays an important role in inter-firm network. This framework is similar to the results obtained in social networks and World Wide Web hyperlink networks. On the other hand, feedforward loop (subgraph 6) and feedback loop (subgraph 7) are anti-motifs, showing that feedforward and feedback are not important. That is, the firm tends not to form circular relationship with other firms but forms V-shaped and clique structures.

5 Role analysis

In order to conduct a more detailed analysis, we classify nodes in a subgraph into roles by examining the structural equivalence of all possible pairs of nodes. For example, the feedback loop (subgraph 7) has only one role (role 17) because a cyclic permutation of the three nodes preserves its structure, whereas the feedforward loop (subgraph 6) has three roles: an input node (role 14), an internal node (role 15), and an output node (role 16). In this way, we can classify every node into 30 types of role as shown in Fig. 4. Note that each node can take more than one role.

Each firm is classified into one of the 96 industries according to the sector it belongs to as the primary industry. To determine the difference between industries, we focused on an industry and checked the role of every firm in that industry. Similar to the motif analysis, a significance profile of the role of each industry is obtained by estimating the Z score:

$$Z_i = \left(N_i - \left\langle N_i^{\mathrm{rand}} \right\rangle\right) \Big/ \sigma_i ,$$

where N_i is the number of times the role i appears in the firm belonging to the industry we focus on in the network, $\left\langle N_i^{\mathrm{rand}} \right\rangle$ and σ_i are the mean and the standard deviation of its appearances in the randomized networks respectively. In order to compare industries of different number of firms, we use normalized Z score:

$$\mathrm{SP}_i = Z_i \Big/ \left(\sum_i Z_i^2 \right)^{1/2} .$$

The similarities and differences between industries can be detected by the correlation coefficient between the significance profiles of two industries i and j:

$$\rho_{ij} = \frac{\left\langle \mathrm{SP}_i\ \mathrm{SP}_j \right\rangle - \left\langle \mathrm{SP}_i \right\rangle \left\langle \mathrm{SP}_j \right\rangle}{\sqrt{\left\langle \mathrm{SP}_i^2 - \left\langle \mathrm{SP}_i \right\rangle^2 \right\rangle \left\langle \mathrm{SP}_j^2 - \left\langle \mathrm{SP}_j \right\rangle^2 \right\rangle}} ,$$

where the angular brackets indicate an average of all 30 roles (Fig. 6). To classify 96 industries into groups, we performed a cluster analysis with Ward's method using Euclidean distance. The cluster analysis divides industries into groups so that industries from the same group are more similar to each other than those from different groups. In Fig. 7, we show the obtained dendrogram, which illustrates the industrial hierarchy

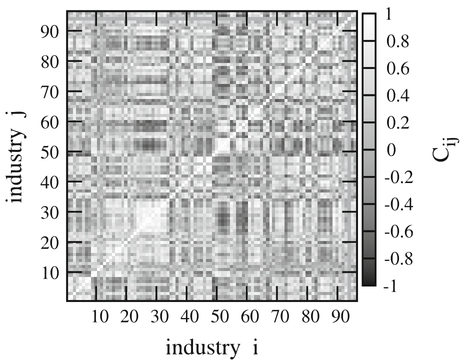
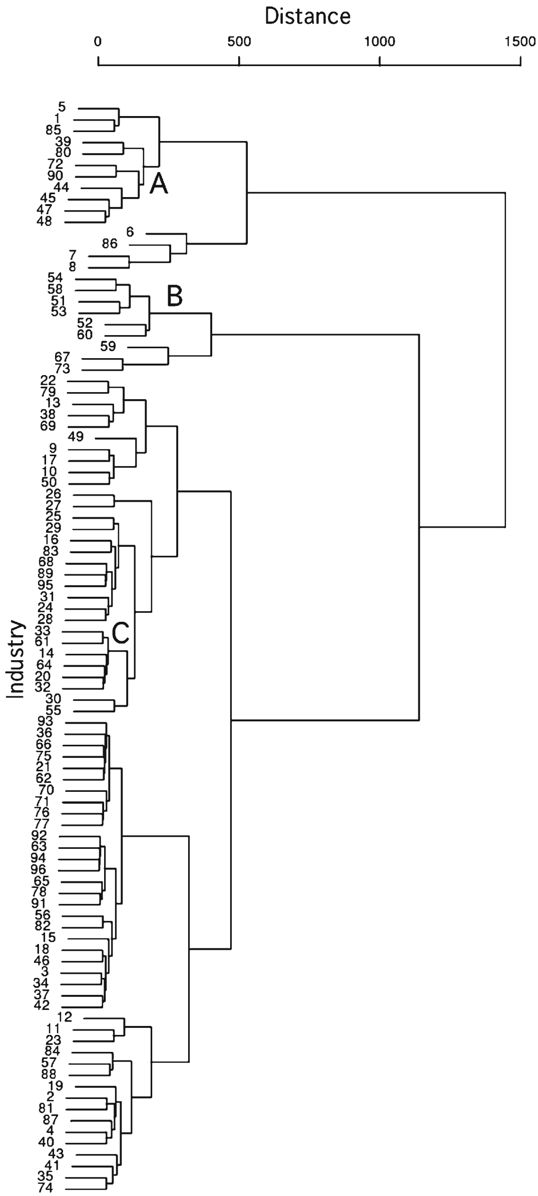


Fig. 6 The correlation coefficient matrix ρ_{ij} of the significance profiles of roles of the industries

Fig. 7 Dendrogram of the industries obtained with Ward's method using Euclidean distance



and defines groups at different levels. As a typical example, we select three groups mentioned below, whose member industries are listed in Table 1 and significance profiles are given in Fig. 8.

Group A includes information and communication, transport, and service industries. This group is characterized by role 23 in subgraph 10. A firm in the group tends to provide materials and services to two firms which have mutual relations with each

Table 1 Industries in groups A, B, and C

Group	Index	Industry
A	39	Information services
A	44	Road freight transport
A	45	Water transport
A	47	Warehousing
A	48	Services incidental to transport
A	72	Accommodations
A	80	Professional services
A	90	Miscellaneous business services
B	51	Wholesale trade (food and beverages)
B	52	Wholesale trade (building materials, minerals, and metals, etc.)
B	53	Wholesale trade (machinery and equipment)
B	54	Miscellaneous wholesale trade
B	58	Retail trade (motor vehicles and bicycles)
B	60	Miscellaneous retail trade
C	14	Manufacture of furniture and fixtures
C	20	Manufacture of rubber products
C	32	Miscellaneous manufacturing industries
C	33	Production, transmission, and distribution of electricity
C	61	Banking
C	64	Non-deposit money corporations engaged in the provision of finance, credit, and investment

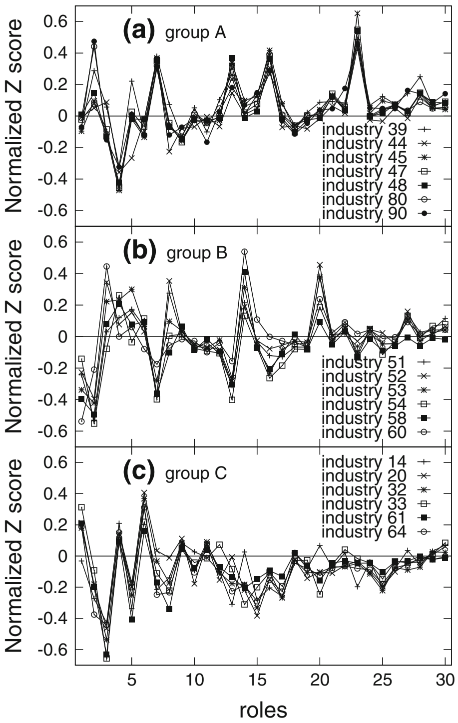
other. In other words, a firm in the group intermediates between mutually connected firms as a broker or an agency.

Group B consists of wholesale and retail trade industries. This group is characterized by role 14 in subgraph 6. A firm in the group tends to provide money to two firms, one of which provides money to the other. The relationships among car retailer (role 14), car manufacturer (role 15), and components manufacturer (role 16) is a typical one in this group. The car retailer buys cars and car components from the other two firms, and the car manufacturer buys car components from the components manufacturer, whereas the car components manufacturer does not buy anything from the other two firms. This represents a typical structure in production.

Group C contains manufacturing industries, characterized by role 6 in subgraph 3. A firm in the group tends to buy materials from a firm, process the materials, and sells the product to another firm. This represents a typical structure in processing.

As seen above, the observed grouping is economically explainable, indicating that the obtained dendrogram corresponds closely to the industrial sector. It is possible to extract the taxonomy of industries by using only the significance profile of a role. This result shows that the significance profile of a role possesses valuable economic information.

Fig. 8 The significance profiles of the roles of groups A (**a**), B (**b**), and C (**c**)



6 Conclusion

We have empirically analyzed the data of large-scale inter-firm networks. By using the motif analysis, we found that V-shaped triads appear most frequently (network motif), while feedforward and feedback loop appear least frequently (anti-motif). These motifs are the characteristics of an inter-firm network. Then, we performed a role analysis to investigate further relations between the firms. The clusters obtained using the significance profile of roles are economically meaningful, implying that we can extract important information from the local structure of the directed network. These findings provide a valuable insight into the relationship between the economic function and the network structure. More detailed investigations are being undertaken in order to find an easily interpretable view of the entire inter-firm network and to improve the efficiency and safety of economic system.

Acknowledgments The authors are grateful to the Research Institute of Economy, Trade, and Industry (RIETI) for allowing us to use the TSR data. This work was partially supported by a Grant-in-Aid for Young Scientists (B) No. 20760053 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (T.O.), the Zengin Foundation for Studies on Economics and Finance (T.O.), and an Initiative To Promote Supercomputer Usage Among Young Researchers from Information Technology Center, the University of Tokyo (T.O.). Numerical calculations were performed by Hitachi HA8000 at Information Technology Center, the University of Tokyo.

References

Clauset A, Shalizi C, Newman M (2007) Arxiv preprint arXiv:0706.1062

Fujiwara Y, Aoyama H (2008) Large-scale structure of a nation-wide production network. Arxiv preprint arXiv:0806.4280

Itzhack R, Mogilevski Y, Louzoun Y (2007) An optimal algorithm for counting network motifs. Physica A 381(15):482–490

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298(5594):824–827

Milo R, Kashtan N, Itzkovitz S, Newman MEJ, Alon U (2003) Uniform generation of random graphs with arbitrary degree sequences. Arxiv preprint cond-mat/0312028

Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U (2004) Superfamilies of evolved and designed networks. Science 303(5663):1538–1542

Ohnishi T, Takayasu H, Takayasu M (2009) Hubs and authorities on Japanese inter-firm network: characterization of nodes in very large directed networks. Prog Theor Phys Suppl 179:157–166

Ravasz E, Barabási AL (2003) Hierarchical organization in complex networks. Phys Rev E 67:026112

Saito YU, Watanabe T, Iwamura M (2007) Do larger firms have more interfirm relationships? Physica A 383(1):158–163