Communities in an inter-firm network and their geographical perspectives

Ritsu Sakuramachi, Naoya Fujiwara, Shota Fujishima, Yuki Akiyama and Ryosuke Shibasaki

Abstract

We detect communities, or groups of firms having mutually close business relationship, in a huge inter-firm transaction network and study their geographical perspectives. We apply a random-walk based method proposed in network science for the community detection to a Japanese inter-firm network having around 0.7 million firms. We found that the community size follows a power-law distribution, and high-ranked communities in terms of size are nonlocalized nation-wide communities. Although the industrial proximity is generally important for the community formation, whether the geographical proximity matters depends on the type of business. Localized communities tend to be medium-sized. We also shed light on the hierarchical structure of community. We found that some of nation-wide communities are divided into several localized subcommunities.

R. Sakuramachi (Corresponding author) • N. Fujiwara • S. Fujishima • Y. Akiyama • R. Shibasaki
Center for Spatial Information Science, at the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, JAPAN
Email: rsakura@csis.u-tokyo.ac.jp

N. Fujiwara
Email: fujiwara@csis.u-tokyo.ac.jp

S. Fujishima
Email: sfujishima@csis.u-tokyo.ac.jp

Y. Akiyama
Email: aki@iis.u-tokyo.ac.jp

R. Shibasaki
Email: shiba@csis.u-tokyo.ac.jp

1 Introduction

Inter-firm transactions, which give us important information on national or regional business activities, have been studied intensively in order to comprehend economic phenomena. Such transaction patterns can be represented as an inter-firm network in which nodes and links respectively correspond to firms and transactions among them. Although inter-firm networks are huge in general, we are naturally interested in classifying firms into several "communities," which are groups of firms having mutually close business relationships, in order to effectively study economic phenomena[1]. The purpose of this paper is to detect communities in a real inter-firm network in which links are weighted by transaction volumes.

The inter-firm network has been well studied from the perspective of network theory. For example, Ohnishi et al. (2009) show that both in-degree and out-degree distributions in a Japanese inter-firm network follow power-law distributions.[2] On the other hand, geographical perspectives, such as the spatial distribution of firms within a community, have received little attention. We shed light on this issue by projecting detected communities onto a map by using GIS, and investigate relationships between business connections among firms and the geographical proximity. Because our network does not have any direct spatial factors a priori, it is not obvious whether detected communities are geographically localized. Our ultimate interest lies in studying in what cases the geographical proximity plays or does not play a primary role for the business connections among firms.

As for the community detection, various algorithms have been proposed in the network science.[3] One of the most widely-used methods is the modularity method (Newman and Girvan, 2004) that pins down the most "plausible" partition of links relative to the case where links are randomly assigned. Focusing on the manufacturing sector, Fujiwara and Aoyama (2010)[4] use the modularity method to detect communities in a Japanese inter-firm network. They detect more than 1,000 communities, and find that

[1] Bargigli and Gallegati (2011) consider an expected degree model based on random graph theory, and statistically reject the hypothesis that a real inter-firm network in Japan is consistent with the random network model. As they argue, their result implies that the community structure is important in the inter-firm network.

[2] Miura et al. (2012) provide a theoretical model that is consistent with this observation. In our context, in-degree is the number of firms that make payments to a given firm whereas out-degree is the number of firms to which a given firm makes payments.

[3] See Fortunato (2010) for an overview of community detection methods.

[4] Rosvall and Bergstrom (2011) improve their original method of Rosvall and Bergstrom (2008) to account for the hierarchical structure of communities. We use their improved method.

small communities generally consist of firms specializing in the same type of business and are geographically localized.

Contrary to their work, we use the random-walk based method of map equation developed by Rosvall and Bergstrom (2008) which detects the community structure that minimizes the code length necessary to describe the flow of random walkers on a network. This method is flexible enough to account for transaction volumes among firms and the hierarchical structure of communities, both of which are not addressed in Fujiwara and Aoyama (2010). In particular, the hierarchical structure would be quite relevant for inter-firm transaction networks because inter-firm relationship is generally not on an equal footing due to, for example, the contractor-subcontractor relationship.[5]

The data set used in this study, which consists of approximately 4.5 million transactions among approximately 0.7 million Japanese firms performed in 2013, was provided by Teikoku Databank, Ltd. Although the original data indicates whether transactions between firms exist or not and the only partial information on transaction volumes is available, Tamura et al. (2012) estimate pair-wise transaction volumes on the whole network. We also use their estimates for the community detection.

Applying the random-walk based method to the Japanese inter-firm network above, we detected about 24,000 communities. However, we found that the community size follows a power-law distribution. This implies that most of the communities are quite small, whereas only a few communities are significant. Indeed, the probability of random walker visiting one of the 30 largest communities is more than 0.5.

Our main claim is that the importance of the geographical proximity is mixed: in other words, the geographical proximity is not always a dominant factor for economic connections of firms. On the one hand, we found large communities in which firms having similar industrial profiles are spread all over Japan. The existence of such nation-wide communities, which are far from localized, imply that the industrial proximity can be sufficient for firms to form a community. On the other hand, by measuring the degree of geographical concentration of community members with Ripley's K-function (Ripley, 1981), we found medium-size communities that are geographically localized. Firms in those communities have similar industrial profiles as in large communities. Therefore, whether the geographical proximity matters depends on the type of industry.

We also found that some communities have the hierarchical structure. That is, some communities are further divided into several subcommunities.

[5] Whereas this is concerned with directed relationship *within* a community, Iino and Iyetomi (2012a) study directed relationship *between* communities which they detect with the modularity method in a Japanese inter-firm network.

As far as we know, this point has not been addressed within a unified framework in the literature studying a real inter-firm network.[6] We present a community that consists of firms spread all over Japan but is divided into several localized subcommunities. Thus, looking at the hierarchical structure reveals that the geographical proximity is actually important for this nation-wide community. However, we should not overestimate this point because many of nation-wide communities do not have the hierarchical structure.[7]

Finally, we point out that some communities are similar in terms of both industrial and geographical profiles. This implies that some firms engage in the same kind of business and locate close to each other, but belong to different communities. Thus, the industrial and geographical proximities are not the only factors for the community formation, and some other factors such as whether they belong to the same business combine would be relevant for this case.

The rest of the paper is organized as follows. Section 2 explains our data set and the methodology of community detection. Section 3 reports our results and discusses them. Section 4 concludes and discusses subjects of future research.

2 Methodology

2.1 Data

The data set used in this study, which consists of approximately 4.5million transactions among approximately 0.7 million Japanese firms performed in 2013, was provided by Teikoku Databank, Ltd. It includes the information of each firm-transaction such as addresses of both orderer and receiver, main contents of transaction, and its estimated value.

[6] Iino and Iyetomi (2012b) propose the recursive community detection in which the modularity maximizations are conducted again for each community that has been detected in the first round. Applying this method to a Japanese transaction data, they find that large communities are further divided into several subcommunities.

[7] The largest community that has the hierarchical structure is ranked 23rd.

2.2 Random walk model and map equation

In this paper, we detect communities of the inter-firm transaction network using the random walk based method. Here we regard the flow of money caused by the firm transaction as the random walk. Let us consider an inter-firm transaction network of N firms. The available data is the amount of transaction from a firm *j* to another firm *i* during one period, M_{ij}. We assume that the transaction network is closed hence the total amount of money is conserved. The transition matrix representing the transition probability from node *j* to node *i* is given as

$$T_{ij} = \alpha \frac{M_{ij}}{\sum_k M_{kj}} + (1-\alpha)\frac{1}{N}. \tag{1}$$

The diagonal elements are set as $M_{ii} = 0$ which implies that firms spend all money for the transaction in the next period in this model. The second term in the right hand side represents the random jump with the probability $1-\alpha$, where we took $\alpha = 0.85$. Then the time evolution of the money flow is given as

$$c_i(t+1) = \sum_j T_{ij} c_j(t). \tag{2}$$

If the network is strongly connected, the largest eigenvalue of the transition matrix is one corresponding to the steady state of this evolution equation. We define the *flow* c_i^*, the probability that a random walker stays in the *i*th node in the steady state, which satisfies

$$c_i^* = \sum_j T_{ij} c_j^*. \tag{3}$$

It is evident that c_i^* is the right eigenvector for the largest eigenvalue of T_{ij}.

A novel community detection algorithm, the map equation (Rosvall and Bergstrom, 2008) has been proposed based on the above random walk model. Random walkers are expected to stay inside one community for long time and sometimes jump to another community. The above intuition leads us to describe the paths of random walkers using the label of nodes taking into account the community structure. Length of the description can be discussed within the framework of the information theory. Namely, it is a problem to find the group of nodes which minimizes the description length

$$L(M) = qH(Q) + \sum_{i=1}^{m} p_i H(P_i), \tag{4}$$

where $H(Q)$ and $H(P_i)$ dentote the Shanon entropy of a random walker to change community and that to move inside community i, which are computed using the steady state probability c_i^*. The first and second terms of the above equation give the code length required to describe the jump between communities and the jump inside a community, respectively. In this work, we employ the extended version of the above algorithm to the multi-level community structure (Rosvall and Bergstrom, 2011).

2.3 Visualization

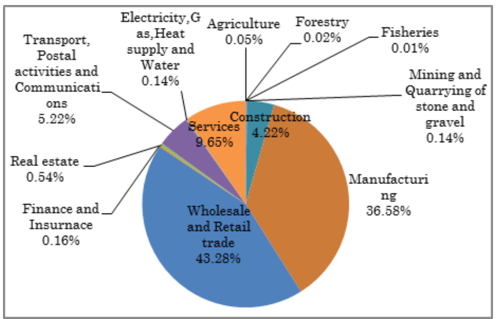
We visualized the detected communities by projecting distribution of firms in each community as points onto a map by using GIS. We color-sorted points of firms by their flow value in order to grasp which firms play an important role in their community. Furthermore, we indicate the shares of major three-digit industrial categories for each community in order to find the characteristic of transactions among it.

3 Result

Applying the map equation program to the Japanese inter-firm network described in Section 2.1, flows, which are the probabilities of random walker visiting a particular node at the steady state as we defined in the previous section, are assigned to each firm and approximately 24 thousand communities are found among approximately 0.7 million firms.

3.1 Flow

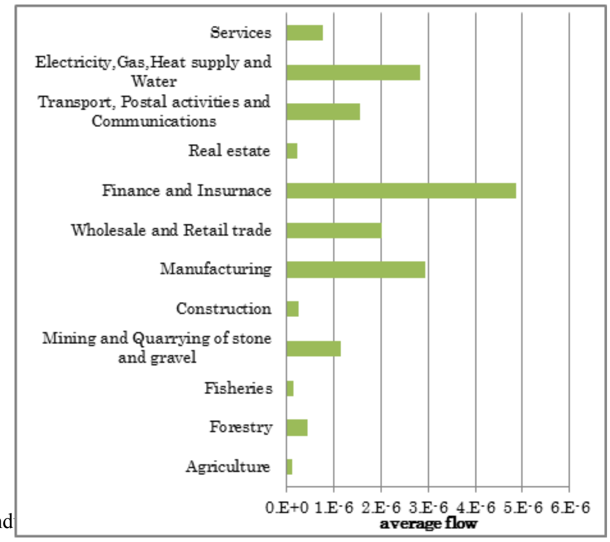
To obtain the big picture on our Japanese inter-firm transaction network, we classify firms according to their types of business into 12 categories



and compute aggregate flows for each industrial category.[8] Fig.1-(a) shows the ratios of the aggregate flow to the total flow for each business category. It illustrates that the flows of "wholesale and retail trade" and "manufacturing" account for most of the total flow. This fact indicates that these types of business play important roles in the inter-firm transaction network in Japan, which agrees with the finding of Iino and Iyetomi (2012b) who detect communities in a Japanese inter-firm network with the modularity method. However, the number of firms greatly depends on business categories as shown in Fig.1-(b). Taking this into account, the average flows are shown in Fig.1-(c) for each business category. From a viewpoint of the average flow, the largest one is attained in "finance and insurance," followed by "manufacturing" and "electricity, gas, heat supply and water." Although the flow of "finance and insurance" has the share of only 0.16%, the number of firms in the category is small enough to make its average flow the largest. It indicates that each single firm of "finance and insurance" plays an important role in the network of inter-firm transaction. The "manufacturing" and "wholesale and retail trade" are major industries in Japan in terms of both the total flow and the average flow.

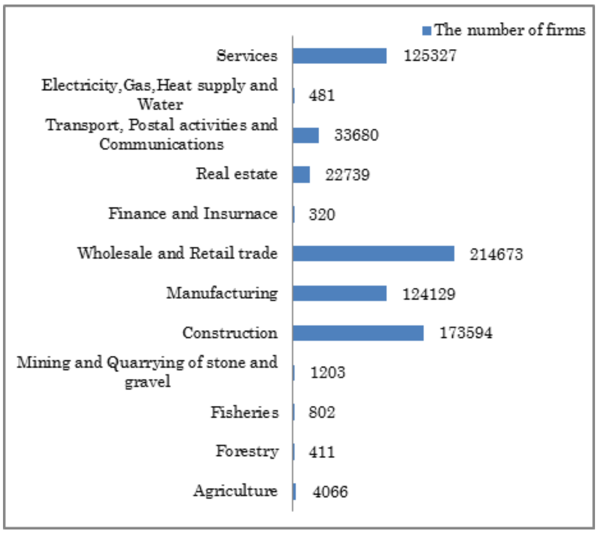
Fig. 1 Classification of firms according to the 12 types of business categories**.**

(a) Aggregate flow of 12 business categories



[8] Although we follow the ind
nese Standard Industrial Classification.

(b) the number of firms of 12 business categories



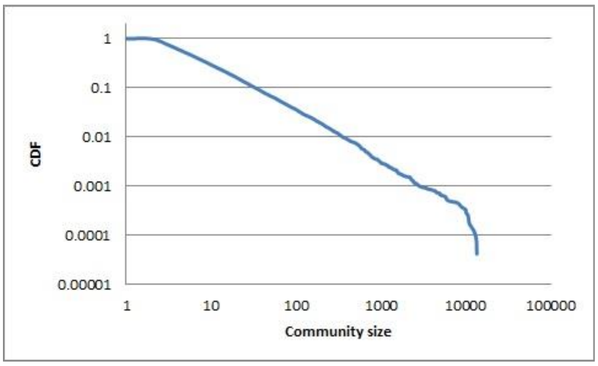
(c) average flow of 12 business categories, are depicted

3.2 Community structure

The community size, which is defined as the number of firms in a community, varies considerably over communities. While some communities consist of more than 10 thousand firms, some communities consist of only a few firms. Fig. 2 shows the cumulative distribution of community size. It is evident from this graph as a line that the community size follows a power-law distribution. This implies that economic activities in our inter-firm network can be mostly explained by a small number of communities. Indeed, looking at the flow that is an important measure of economic activities in our context, we found that more than 50% of the total flow is governed by 30 communities.

Besides, some communities are further divided into subcommunities. Specifically, approximately 1,460 communities have their subcommunities and, moreover, 51 of those communities have two-level hierarchical structures.

Fig. 2 Cumulative distribution of community size.



3.3 Communities and the Geographical Proximity

We study whether the geographical proximity matters for the community formation. At first, we project the locations of firms on a map for each of the three largest communities in terms of the aggregate flow to visualize the spatial distributions of their firms (see Fig. 3).

The Fig. 3 shows that firms in these communities are broadly distributed all over Japan. One might expect that these communities can be divided into several localized subcommunities. However, as we will see, our algorithm does not detect such a hierarchical structure for any of them. Quite naturally, firms with high flows tend to locate in large cities such as Tokyo and Osaka. The three communities have the common feature that they are composed of firms having similar industrial profiles. For example, the three largest three-digit industrial categories in Community 1, which accounts for about 70% of all categories there, are all related

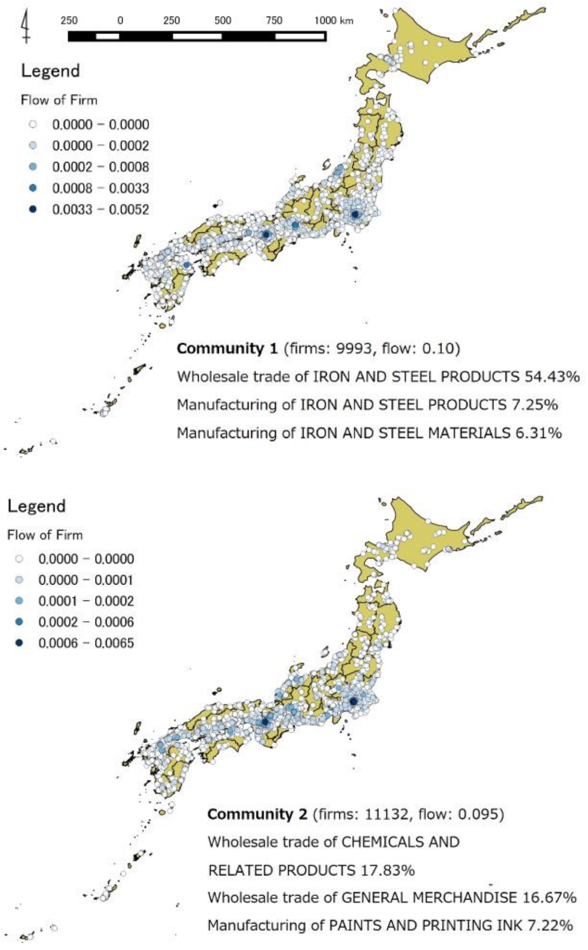


Fig. 3 The spatial distributions of firms in principal communities. We pick up the three largest communities in terms of the aggregate flow. The locations of firms having high flows are represented by dark blues. The shares of major three-digit industrial categories are indicated for each community.

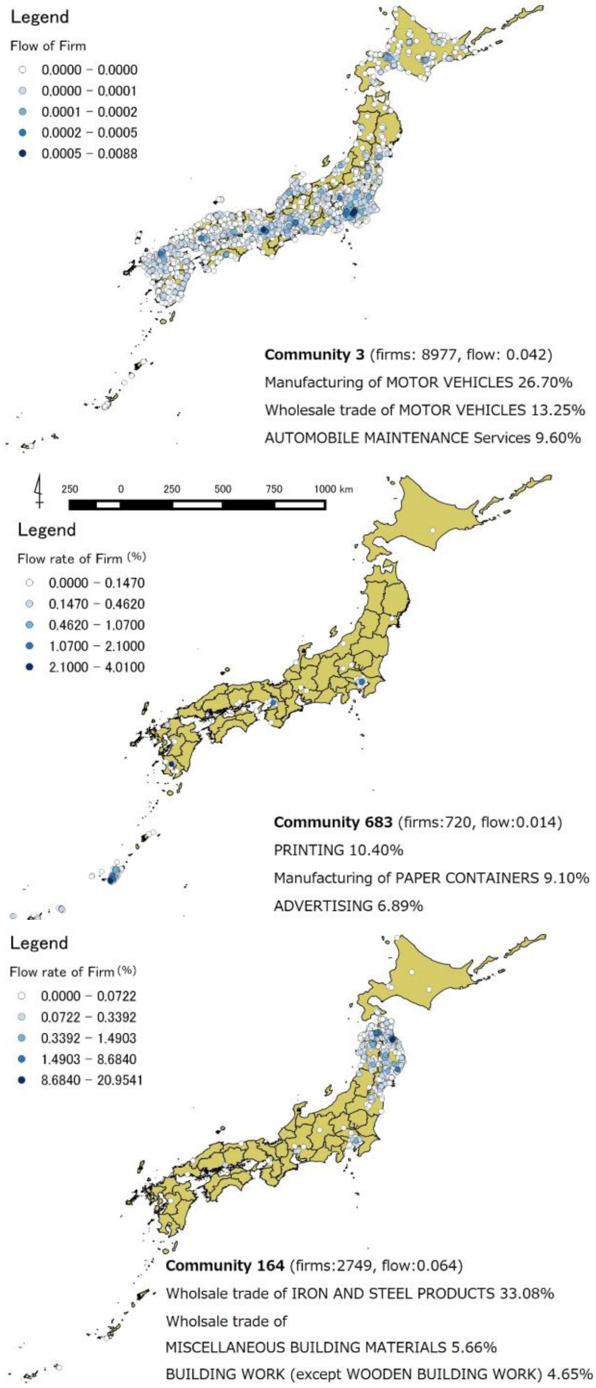
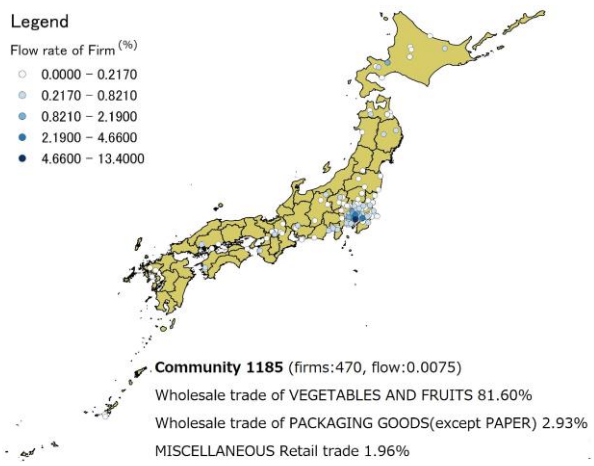


Fig. 4 The spatial distributions of firms in especially centralized communities. We pick up the three most centralized communities according to Ripley's K-function. The locations of firms having high flows are represented by dark blues. The shares of major three-digit industrial categories are indicated for each community.



to iron and steel. Therefore, for the primary nation-wide communities, what is important is the industrial proximity rather than the geographical proximity.Next, we look at localized communities. To this end, we measure the degree of spatial centralization of firms for each community by Ripley's K-function (Ripley, 1981) in order to extract especially centralized communities. We pick up three communities that can be regarded as localized in Fig. 4 and discuss several observations.

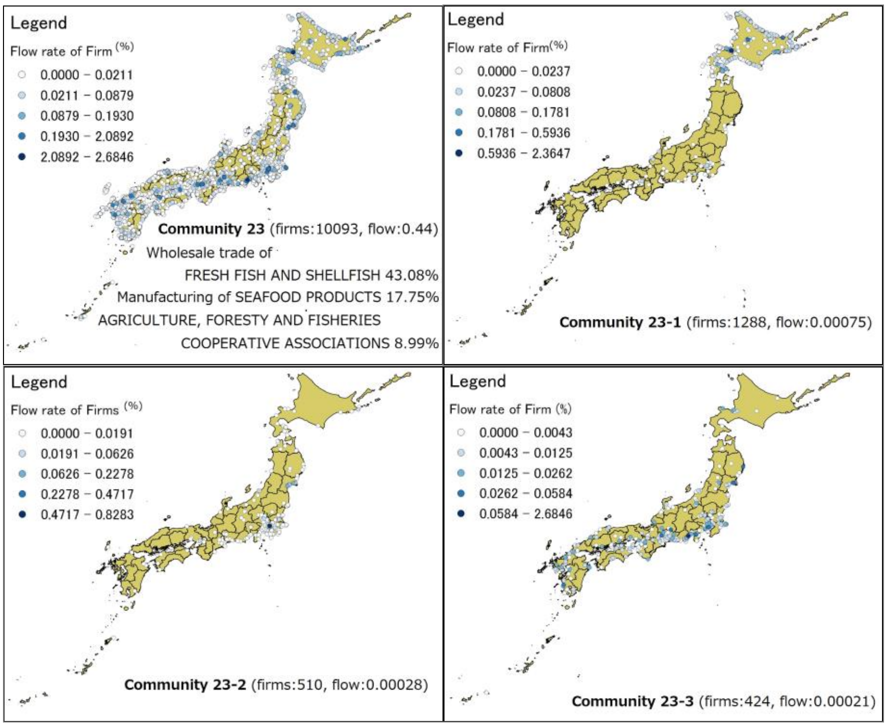
First, we found that many of the most localized communities in terms of K-function value are placed in Okinawa prefecture, which locates in the southern tip of Japan, as the place of agglomeration. See Community 683 for such an example. It can be expected that Okinawa has many localized communities because of its geographical isolation from the Mainland. Second, firms in Community 164 are agglomerated in Tohoku, the North area except Hokkaido island. It turned out that this community is mainly composed of affiliated companies of a top-ranked firm in terms of flow which sells construction materials in Tohoku. Third, firms in Community 1185 are agglomerated in the Tokyo metropolitan area and the share of "Wholesale trade of vegetables and fruits" is extremely high. The geographical proximity apparently matters for these communities, while they do not have very high aggregate flows.

Based on the above observations, it can be reasonably concluded that, while the industrial proximity is generally important for the community formation, the combination of type and place of business tends to be relevant for whether the geographical proximity matters.

3.4 A Community with Hierarchical Structure

Some communities are further divided into subcommunities and have the hierarchical structure. We present Community 23 in Fig. 5 as an example. This community is the largest of all communities having the hierarchical structure. The industrial profile of this community is characterized by fishery, and many firms are located on the coast. Around 300 subcommunities exist in the lower level of this community. The five largest subcommunities in terms of the aggregate flow are mapped in Fig. 5. Unlike the parent community, the subcommunities have the tendency that one city or region attracts a dominant number of firms. For example, most of firms are located in Hokkaido, the northernmost island, in the first subcommunity, and most of firms are located in Tokyo in the second subcommunity.

Thus, although the parent community is a nation-wide community with firms having similar industrial profiles, the geographical factor actually seems to matter in the lower class of the community, unlike the primary nation-wide communities without hierarchical structure in Section 3.2.



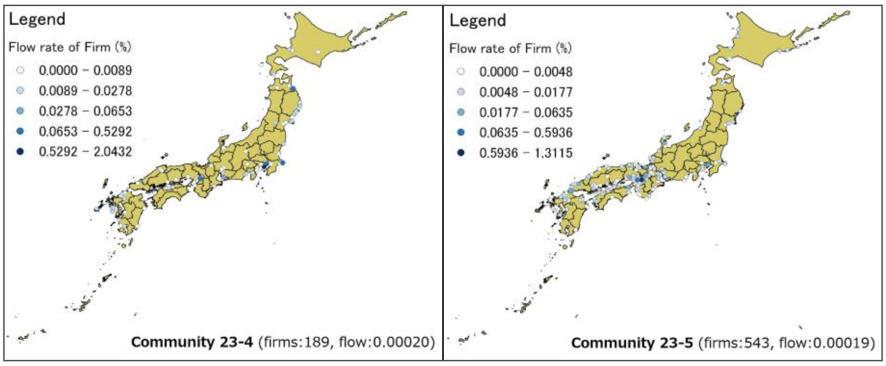


Fig. 5 The spatial distributions of firms in the community 23, the largest one having the hierarchical structure. We also pick up the five largest subcommunities in terms of the aggregate flow (community 23-1 to 23-5). The locations of firms having high flows are represented by dark blues. The shares of major three-digit industrial categories are indicated for each community.

3.5 Communities that are similar in terms of both industrial and geographical profiles

Although we have looked at the industrial and geographical proximities as factors for the community formation, we could expect that these are not the only ones. To see this point, we present two distinct communities such that both industrial profiles and spatial distributions of community members are similar to each other. The two communities in Fig. 6 serve as such an example. Comparing the two communities, four of the five largest industrial categories overlap and the spatial distributions of firms are quite similar. Thus, there would be firms that have similar industrial profiles and locate close to each other but belong to different communities. This implies that factor(s) other than the industrial and geographical proximities is relevant here. One possible factor would be whether they belong to the same business combine, or the zaibatsu.

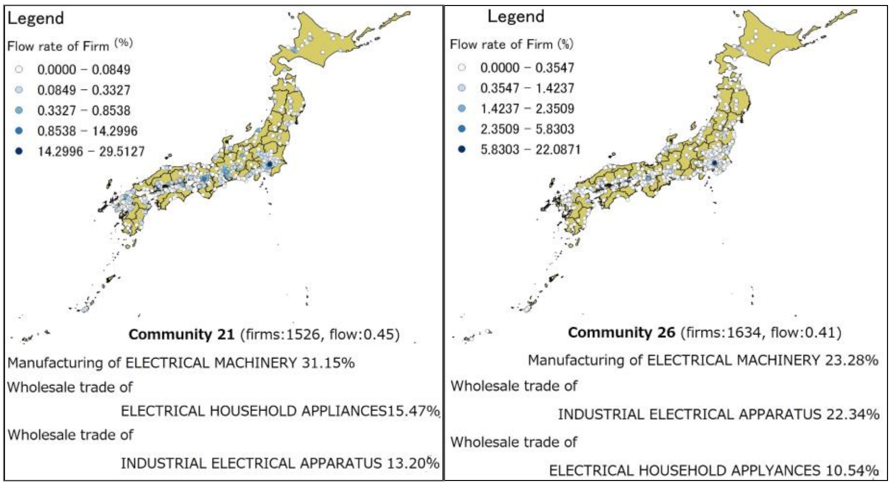


Fig. 6 The spatial distributions of firms in the communities 21 and 26, communities such that both industrial profiles and spatial distributions of community members are similar to each other. The locations of firms having high flows are represented by dark blues. The shares of major three-digit industrial categories are indicated for each community.

4 Conclusion and Future Research

We applied the random-walk based method for the community detection developed by Rosvall and Bergstrom (2008, 11) to a Japanese inter-firm network with approximately 0.7 million firms. Our primary interest lies in whether the geographical proximity matters for the community formation. Investigating the detected communities led us to the observation that, although the industrial proximity is generally important for the community formation, whether the geographical proximity matters depends on the type of industry. That is, the geographical proximity is not always indispensable for the business connections. We also found that some communities have the hierarchical structure such that a community is further divided into subcommunities.

We close our paper by discussing several subjects for future research. First, it is desirable to study properties of detected communities more formally. In particular, we need a reliable measure to evaluate the “closeness” among community members in terms of the types of business they do. Moreover, although we used the K-function to evaluate the geographical locality of communities, this argument depends on the particular distance we choose. Addressing these problems would enable us to rigorously evaluate which factor is dominant for a community, industrial proximity or geographical proximity.

Second, because our study is about real regional economies, it is desirable to derive some policy implications. For example, it is often pointed out that, while big cities such as Tokyo are attaining the dominant share of population, small cities are shrinking in Japan. By investigating how vulnerable localized communities in small cities are to various kinds of risks such as macroeconomic shocks and disasters, we might be able to discuss what the local and central governments can do to enhance the economic performances of small cities.[9]

Finally, although we focused on economic activities of firms by using inter-firm transaction data, we could also take economic activities of consumers into account by using person flow data and commuter trip data. With these data, for example, we might be able to provide a reasonable definition for metropolitan area or functional region, which would have an important implication for urban planning.[10]

References

Bargigli, L. and M. Gallegati. (2011), "Random digraphs with given expected degree sequences: A model for economic networks," *Journal of Economic Behavior & Organization*, **78**, 396-411.

Battiston, S., D. D. Gatti., M. Gallegati., B. Greenwald., and J. E. Stiglitz. (2007), "Credit chains and bankruptcy propagation in production networks," *Journal of Economic Dynamics and Control*, **31**, 2061-2084.

Farmer, C. J. Q. and A. S. Fotheringham. (2011), "Network-based functional regions," *Environment and Planning A*, **43**, 2723-2741.

Fortunato, S. (2010), "Community detection in graphs," *Physics Reports*, **486**, 75-174.

Fujiwara, Y. and H. Aoyama. (2010), "Large-scale structure of a nationwide production network," *The European Physical Journal B*, **77**, 565-580.

[9] Studying how such risks propagate over an inter-firm transaction network is an active research subject. See, for example, Henriet et al. (2012) and Battiston et al. (2007) for theoretical work and Todo et al. (forthcoming) and Mizuno et al. (2014) for empirical work.

[10] See Farmer and Fotheringham (2011) for research along this line.

Henriet, F., S. Hallegatte., and L. Tabourier. (2012), "Firm-network characteristics and economic robustness to natural disasters," *Journal of Economic Dynamics and Control*, **36**, 150-167.

Iino, T. and H. Iyetomi. (2012a), "Directional bias between communities of a production network in Japan," In: Watada, J. (eds.), *Intelligent Decision Technologies Vol. 2*, Springer, 273-280.

Iino, T. and H. Iyetomi. (2012b), "Subcommunities and their mutual relationships in a transaction network," *Progress of Theoretical Physics Supplement*, **194**, 144-157.

Miura, W., H. Takayasu., and M. Takayasu. (2012), "Effect of coagulation of nodes in an evolving complex network," *Physical Review Letters*, **108**, 168701 (4 pages).

Mizuno, T., W. Souma., and T. Watanabe. (2014), "The structure and evolution of buyer-supplier networks," *Plos One*, **9**, e100712.

Newman, M. E. J. and M. Girvan. (2004), "Finding and evaluating community structure in networks," *Physical Review E*, **69**, 026113.

Ohnishi, T., H. Takayasu., and M. Takayasu. (2009), "Hubs and authorities on Japanese inter-firm network: Characterization of nodes in very large directed networks," *Progress of Theoretical Physics Supplement*, **179**, 157-166.

Ripley, B. D. (1981), *Spatial Statistics*, John Wiley.

Rosvall, M. and C. T. Bergstrom. (2008), "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, **105**, 1118-1123.

Rosvall, M. and C. T. Bergstrom. (2011), "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems," *PLoS ONE*, **6**, e18209.

Tamura, K., W. Miura, M. Takayasu., H. Takayasu., S. Kitajima., and H. Goto. (2012), "Estimation of flux between interacting nodes on huge inter-firm networks," *International Journal of Modern Physics: Conference Series*, **16**, 93-104.

Todo, Y., K. Nakajima., and P. Matous., "How do supply chain networks affect the resilience of firms to natural disasters? Evidence from the great east Japan earthquake," *Journal of Regional Science*, forthcoming.