Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 3, No. 1, 2025

Similarity-Based Network in the Industrial Community of Joyo City

Keita Takeuchi [1,*], Masashi Iwasaki [1] and Masato Shinjo [2]

[1] Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, 606-0823, Japan; s824732002@kpu.ac.jp (K.T.); imasa@kpu.ac.jp (M.I.)

[2] Faculty of Data Science, Osaka Seikei University, 533-0007, Japan; shinjo@osaka-seikei.ac.jp (M.S.)

***** Correspondence: s824732002@kpu.ac.jp

Article History

Received 3 February 2025
Revised 2 March 2025
Accepted 6 March 2025
Available Online 5 May 2025

Keywords:
Industrial community
Similarity analysis
Network expression
Joyo City

Abstract

Data utilization is becoming increasingly widespread in a variety of fields around the world, and has become especially important in the industrial world. Data utilization techniques and approaches can contribute to the development of not only individual companies but also certain groups of companies. In this paper, we consider the industrial structure of Joyo City, Japan, by analyzing data collected through interviews with company presidents and managers. The main purpose of this paper is to grasp it in terms of similarity across industrial categories. We first express the features of each company as a vector with entries determined from the interview data. We then compute vector similarities in order to draw a graphical network, in which nodes corresponding to similar companies are linked by an edge. From the resulting network, we derive the most similar companies in the same and different industrial categories for each company. Moreover, we then classify Joyo City's companies into new groups across the standard categories.

1. Introduction

Recent developments in information technology (IT) have made data analysis more accessible across various fields [1, 2], and then have globalized various industrial fields, giving rise to more complicated competition among companies beyond international borders [3]. Thus, to prevent competing company from declining, it is necessary to leverage the possibilities of data analysis [4, 5]. Many companies at the top of the world's market capitalization ranking have strong data utilization capabilities and functions.

Of course, data utilization is expected to enrich not only individual companies but also their related companies. For example, Airbus, a well-known aeronautics and space company, shares various data about its aircraft through the Skywise platform, and this allows airlines to utilize that data to enhance their own services [6]. Recently, in Japan, municipal governments have applied data utilization to support their local industrial community, with data-based decision-making enabling investment in local industries, initiatives to attract new companies, and promotion to tourists [7, 8]. A better understanding of the local industrial structure should be useful in creating new business matchings that benefit local companies as well as the municipality overall, including through online services that better convey and promote the entirety of the municipality's industry.

The main purpose of this paper is to present data analysis techniques for capturing a municipality's industrial structure, with a case study of the industrial community of the municipality of Joyo City (hereafter, Joyo). Joyo is located in the southern part of Kyoto Prefecture, Japan. Its area is 32.71 km^2, which is 1,439th of the 1,741 municipalities in Japan. As of January 2024, the population of Joyo is 72,814, which is ranked 377th. Joyo's predecessor was formed about half a century ago by the merger of four different villages: Kutsukawa, Terada, Tonosho, and Aodani. The principal industries of those four villages are different, and the current industrial community in Joyo inherits the diversity. Moreover, Joyo develops diversely as a significant transportation hub along with the development of other nearby prefectures such as Osaka, Hyogo, Nara, Shiga, and Mie. The industrial community of Joyo, which benefits from the transportation convenience, comprises various kinds of manufacturing, including Japan's leading companies for metalworking, fabric dyeing, lamé yarn manufacturing, information display system manufacturing, and semiconductor-related manufacturing. With respect to agriculture, various kinds of crops grow well in Joyo's warm climate, with its specialties being Japanese tea, figs, plums, and sweet potatoes. The abundant groundwater makes it possible to cultivate aquatic plants and to produce beverages and Japanese saké. Moreover, Joyo has several long-established Japanese confectionery stores. According to the 2021 data for Joyo's industry provided by Regional Economy Society Analyzing System (RESAS) [9], wholesale-retail trade has the highest percentage in terms of the number of companies, accounting 19.8% of the total, with 339 companies, while manufacturing has the highest percentages in terms of the number of employees and valued-added, accounting 28.7% and 47.1% of the total, with 4,493 employees and 27,396 million yen, respectively.

In recent years, Joyo has been actively undertaking significant infrastructure developments, including highways, industrial lands, and a large shopping mall, to attract various businesses and strengthen the local economy. However, without a strategic approach to interaction across industrial categories, there is a risk that these investments may not be fully utilized. Joyo's government requested that we design a website that captures the principal structure of Joyo's industry, spurring this study. Joyo's government also expected that the resulting website would promote interaction among companies in different industrial categories as well as those in the same ones. We were then unable to find an industry summary website with hyperlinks that definitely span across industrial categories. For example, on the website “Nippon no shacho. TV [10]”, hyperlinks are organized based on viewer browsing history, and as a result, there are some hyperlinks across industrial categories. Of course, not every company has hyperlinks with companies from different categories, and hyperlinks do not necessarily reflect company features. Therefore, based on company features, we propose a new approach to reveal relationships among companies not only in the same industrial categories but also in different ones.

Our first step was to represent each company in Joyo as a network based on similarities among companies, regardless of category, and then to reveal similarities among companies based on their features, which play a key role in constructing webpage hyperlinks. While it may seem intuitive and easy to consider companies within the same industrial category as similar, it is less straightforward to identify similarities across different categories. In this paper, we formulate a method for identifying the most similar companies in both the same and different categories for each company. Based on company feature similarity, we also propose a classification that is independent of the standard industrial categories.

2. Materials and Methods

In this section, we first describe our data collection approach, which is interview-based. Next, we propose to draw a graphical network in which nodes correspond to companies and nodes of similar companies are connected with lines. Within that explanation, we describe how to extract company features from the interview data and convert the results into vectors, with the help of matrix singular value decomposition (SVD). Finally, we describe how to investigate industrial features of Joyo across categories by analyzing the similar company network. Employing graph analytics techniques, we identify the most similar companies within and across industrial categories. We further develop a company clustering across the industrial categories to understand Joyo's industrial community.

2.1. Data Collection in Joyo's Industrial Community

We tried to collect numerical data to understand the status of Joyo's prominent companies. However, there were limitations to publicly available numerical data, so we could not completely quantify status with this approach. For example, from a company's website, we could relatively easily find the number of employees, but not necessarily the number of visitors per day. Thus, in this section, we describe the data collection approach based on interviews with company presidents and managers.

We conducted interviews in 2018 and 2019, with fifty companies in Joyo which were informed as key companies by Joyo’s government, as shown in Table 1, which are classified into six categories as reflected in the company IDs: agriculture (AC), manufacturing (MF), retail trade (RT), food service (FS), professional service (PS), and leisure service (LS).

The interviews included both common topics and industry-targeted or unique topics. However, in this

Table 1. IDs and overviews of companies we interviewed.

Company ID	Company Overview
AC01	A tourist farm where visitors can experience harvesting sweet potatoes, which is a specialty of Joyo.
AC02	A nursery that grows and sells aquatic plants.
AC03	A chicken farm that produces purely domestic chicken.
AC04	A farm that grows figs, which is a specialty of Joyo.
AC05	A farm that grows many different kinds of plants, including ornamental plants and vegetables.
AC06	A farmer's market that sells agricultural products, such as vegetables, fruits and their processed foods.
MF01	A lamé yarn manufacturer.
MF02	A manufacturer of testing systems to assist automotive manufacturing.
MF03	A manufacturer of a wide range of equipment, such as mainly semiconductor manufacturing and inspection equipment.
MF04	An ironworker that specializes in metalworking such as metal cutting and precision sheet metal assembly using precision welding.
MF05	A print fabric manufacturer.
MF06	An ironworker that specializes in metalworking such as punching, bending, and welding.
MF07	An ironworker that specializes in metalworking of long parts and large plates.
MF08	A seafood processor that specializes in tuna processing.
MF09	A manufacturer that specializes in press work.
MF10	A manufacturer that specializes in large sheet metal fabrication.
MF11	A manufacturer of soft drinks, alcoholic beverages, and health foods.
MF12	A seafood processor that specializes in konjac, soy milk, and Yuba.
MF13	A traditional fabric dyer.
MF14	A manufacturer of nutritional and dietary supplements.
MF15	A manufacturer of lighting equipment and information display systems.
MF16	A manufacturer of tofu and fried tofu, and other soybean foods.
RT01	A store of seaweed and sprinkles.
RT02	A Japanese sake brewery.
RT03	A Japanese tea store.
RT04	A bakery of pies sandwiches, breads, and gelatos.
RT05	A women's apparel store that specializes in clothing made from natural materials such as hemp and cotton.
RT06	A sporting goods store that mainly sells and maintains baseball gloves.
RT07	A bakery of breads and original sandwiches.
RT08	A bakery of various kinds of breads.
RT09	An artisanal and artistic Japanese confectionery store.
RT10	An artisanal Japanese confectionery store, where signature confectioneries are mochi including beans and strawberry daifuku.
RT11	An artisanal Japanese confectionery store with eat-in corner.
RT12	A confectionery store whose feature is fusion of Japanese and Western styles.
RT13	A confectionery store, where a signature confectionery is strawberry daifuku.
RT14	A confectioner store, where signature confectioneries are mochi including beans and baked mochi.
RT15	A store that mainly sells processed foods made from the high-quality plum “Joshuhaku”, which is a specialty of Joyo.
FS01	A restaurant of a course meal that consists mainly of breads.
FS02	A restaurant that serves fresh vegetables and seafoods with local sake.
FS03	A restaurant that serves traditional Japanese “Kaiseki” cuisine.
PS01	A building and renovation contractor.
PS02	An orthopedic clinic.
PS03	An individual barbershop.
PS04	A tax accountant office.
PS05	A tourist information center.
LS01	A manager of futsal field.
LS02	A manager of cultural complex facility equipped with a concert hall, a planetarium, and multipurpose halls.
LS03	A manager of a leisure land equipped with playground equipment, restaurants, BBQ facilities and accommodation.
LS04	A manager of a big public park.
LS05	A manager of an outdoor facility.

paper, we focus only on the common topics because we are focusing on similarities among companies across categories, and industry-specific questions would not easily contribute to that aim. The common topics are as follows:1. Capital or gross sales per year, 2. Number of employees, 3. Company site area, 4. Working hours per day, 5. Number of weekly holidays, 6. Timing of busy seasons, 7. Timing of busy hours, 8. Sufficiency of manpower, 9. Degree of popularity, 10. Accessibility to train stations, 11. Accessibility to major roads, 12. Number of years in business, 13. Degree of customary, 14. Degree of relevance to region, 15. Sufficiency of nearby accommodation facilities, 16. Number of visitors per day, 17. Average visitor dwell time, 18. Average visitor spending amount, 19. Ratio of domestic visitors from within Joyo to from outside Joyo, 20. Degree of visitor particular, 21. Age range of most visitors, 22. Visitor gender ratio, 23. Ratio of foreign visitors, 24. Capital expenditures and advertising expenses, 25. Degree of new product development, 26. Degree of innovative attempts, 27. Degree of internet service usage, 28. Degree of collaboration with other companies, 29. Degree of collaboration with Joyo City Hall.

At each company, multiple interviewers in our project team scored each topic using grades that were translated to a 1 to 5 integer scale. The interview topics can be divided into those scored with absolute standards and those scored with relative standards. In the former topics, for example, in the interview topic 1. Capital or gross sales per year, if the capital was around ten million yen, we decided to score it as 3. Of course, these criteria should be adjusted based on the municipalities analyzed. We scored for the latter topics by discussing with company presidents and managers how much they feel their company deviates from the standard. For example, in topic 16. Number of visitors per day, interviewers not only asked for the specific number of visitors but also whether company presidents and managers felt there were many or few, and they scored based on that response. For each topic, we then set the final value to the average score for that topic for that company. Where a topic could not be directly scored for a particular company, we substituted the average score across companies to enable data analysis.

2.2. Graphical Network of Similarities Among Companies

We present a graphical network in which a node corresponds to a company, while an edge signifies a similarity between the nodes it connects. The features of each company have been represented as a 29-dimensional vector whose entries are the average ratings in the 29 common interview topics. To represent all fifty companies at once, we also have constructed a 50-by-29 matrix by arranging the feature vectors of the fifty companies.

2.2.1. Feature Extraction of Company Data Using Singular Value Decomposition

Singular value decomposition (SVD) of a matrix is a useful technique for extracting features. Thus, we applied SVD to the company feature matrix to examine the features of Joyo's industrial structure. We then removed minor features from the feature matrix based on the SVD to derive a company's feature vectors focused on major features.

Now, we briefly explain the SVD of a rectangular matrix $A \in \mathbb{R}^{m\times n}$. The ℓ eigenvalues of $A^\top A$ are all nonnegative where $\ell = \min(m,n)$ and $\top$ denotes the transpose. Then, the square roots, conventionally denoted by $\sigma_1, \sigma_2, \cdots, \sigma_\ell$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_\ell$, are called singular values of A. Employing two suitable orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_\ell)$, we can decompose A into their products as follows:

$$A = \begin{cases} U\begin{pmatrix}\Sigma \\ O\end{pmatrix}V^\top, & m > n, \\ U\Sigma V^\top, & m = n, \\ U(\Sigma, O)V^\top, & m < n. \end{cases} \tag{1}$$

This matrix decomposition is the SVD of A. The matrices U and V are respectively called the left and right orthogonal matrices. The ith column vectors, denoted by $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$, are the left and right singular vectors corresponding to σ_i. Using the singular values and left and right singular vectors, we can rewrite Equation (1) as:

$$A = \sum_{i=1}^{\ell} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top. \tag{2}$$

This implies that the magnitude of σ_i is the dominant ratio of $\boldsymbol{u}_i\boldsymbol{v}_i^\top$ in constructing A by a linear combination of $\boldsymbol{u}_1\boldsymbol{v}_1^\top, \boldsymbol{u}_2\boldsymbol{v}_2^\top, \cdots, \boldsymbol{u}_\ell\boldsymbol{v}_\ell^\top$. For i where $i < \ell$, the rank i approximation matrix of A is determined as $A_i \coloneqq \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^\top + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^\top + \cdots + \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^\top$. A common criterion for determining the value of i is the cumulative contribution ratio, which is the ratio of $\sigma_1 + \sigma_2 + \cdots + \sigma_i$ to the sum of all the singular values:

$$r_i \coloneqq \frac{\sigma_1 + \sigma_2 + \cdots + \sigma_i}{\sigma_1 + \sigma_2 + \cdots + \sigma_\ell}. \tag{3}$$

The cumulative contribution ratio is also recognized as a criterion of how well A_i reconstructs A. For example, in a low-rank approximation of A such that the cumulative contribution is 80 %, we regard A_i as a matrix that preserves 80% of the features of A. We note that the 20% of features removed are relatively minor. Thus, the low-rank approximation based on the SVD enables us to extract major features by removing minor features.

In this study, we considered the SVD of the 50-by-29 matrix C whose rows are the feature vectors of fifty companies $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{50}$, and then derived the low-rank approximation matrix $\bar{C}$ using eighteen pairs of singular values and their corresponding singular vectors such that the cumulative contribution ratio is less than and close to 90 % since $r_1 \approx 0.4037, r_2 \approx 0.4624, r_3 \approx 0.5063, r_4 \approx 0.5466, r_5 \approx 0.5830, r_6 \approx 0.6154, r_7 \approx 0.6460, r_8 \approx 0.6754, r_9 \approx 0.7028, r_{10} \approx 0.7301, r_{11} \approx 0.7551, r_{12} \approx 0.7779, r_{13} \approx 0.7990, r_{14} \approx 0.8191, r_{15} \approx 0.8391, r_{16} \approx 0.8579, r_{17} \approx 0.8749, r_{18} \approx 0.8908, r_{19} \approx 0.9057, \cdots, r_{29} = 1$. The reason why we set the threshold value of cumulative contribution ratio to 90% in generating the low-rank approximation matrix $\bar{C}$ is that this threshold value is widely adopted in several papers such as Koyama et al. [11], Fukumoto et al. [12], and Fukui et al. [13]. In the case where we want to further eliminate the influence of singular pairs with low contribution ratio, we should set the threshold value to less than 90%. We hereinafter regarded the row vectors $\bar{\boldsymbol{c}}_1, \bar{\boldsymbol{c}}_2, \cdots, \bar{\boldsymbol{c}}_{50}$ in $\bar{C}$ as feature extraction vectors of the fifty companies with minor features removed with respect to the entirety of Joyo's industry.

2.2.2. Constructing the Similar Company Network

We describe the construction of a graphical network that expresses similarities among companies based on the feature extraction vectors $\bar{\boldsymbol{c}}_1, \bar{\boldsymbol{c}}_2, \cdots, \bar{\boldsymbol{c}}_{50}$. We applied Spearman's correlation method [14] to quantify the similarity of each pair of companies. Companies with high similarity are likely to have the same good or bad scores on the scored items. For two feature extraction vectors $\bar{\boldsymbol{c}}_i := (\bar{c}_i(1), \bar{c}_i(2), \cdots, \bar{c}_i(29))$ and $\bar{\boldsymbol{c}}_j := (\bar{c}_j(1), \bar{c}_j(2), \cdots, \bar{c}_j(29))$, we evaluated the degree of correlation between them from the 29 pairs $(\bar{c}_i(1), \bar{c}_j(1)), (\bar{c}_i(2), \bar{c}_j(2)), \cdots, (\bar{c}_i(29), \bar{c}_j(29))$. However, we did not use the 29 pair values directly as in the well-known Pearson's correlation method [15]. Rather, for each k, we converted $\bar{c}_i(k)$ to the rank, denoted by $\hat{c}_i(k)$, in descending order of values among $\bar{c}_i(1), \bar{c}_i(2), \cdots, \bar{c}_i(29)$ where $\hat{c}_i(k_1) = \hat{c}_i(k_2) = \cdots = \hat{c}_i(k_t) = [s + (s+1) + \cdots + (s+t)]/t$ if the values of $\bar{c}_i(k_1), \bar{c}_i(k_2), \cdots, \bar{c}_i(k_t)$ are tied for sth place, and similarly for $\bar{c}_j(k)$. Note here that the term "rank" is used in a different sense than in the previous subsection. Thus, we obtained:

$$r(\bar{\boldsymbol{c}}_i, \bar{\boldsymbol{c}}_j) := 1 - \frac{1}{4060} \sum_{k=1}^{29} \left(\hat{c}_i(k) - \hat{c}_j(k)\right)^2. \qquad (4)$$

According to Imai et al. [16], we can consider $\bar{\boldsymbol{c}}_i$ as having a weak correlation with $\bar{\boldsymbol{c}}_j$ if $r(\bar{\boldsymbol{c}}_i, \bar{\boldsymbol{c}}_j) = 0.35$. Thus, we judged that company i shares some similarity with company j if $r(\bar{\boldsymbol{c}}_i, \bar{\boldsymbol{c}}_j) > 0.35$, and added an edge between those two company nodes, as shown in Figure 1. Note that the similarity company network in Figure 1 does not include isolated nodes that do not link to any other nodes. If we want to draw a company network presenting strong similarity, we should add link nodes i and j that satisfy $r(\bar{\boldsymbol{c}}_i, \bar{\boldsymbol{c}}_j) > 0.7$. In this network, 32 nodes became isolated. Recall here that the important mission is to construct hyperlinks among companies across industrial categories on the website. Since isolated nodes are clearly undesirable, we thus set the threshold value to 0.35 to determine if nodes i and j are linked.

2.3. Methods for Analyzing the Similar Company Network

We investigate industrial features of Joyo across categories by analyzing the similar company network. We first consider the most similar companies within and across industrial categories, which is beyond the information given in Figure 1. We next found groups of similar companies across categories using the Clauset-Newman-Moore greedy modularity maximization [17].

2.3.1. Identifying the Most Similar Companies in the Same and Different Categories

For some companies in Figure 1, we can identify similar companies in the same category, but cannot identify similar companies in a different category, and vice versa. We treated the similar company network as a weighted graph by interpreting company similarities as edge weights. Moreover, we replaced every edge weight with its reciprocal, and regarded the reciprocal as the length of an edge. Note that the edge length decreases as the company similarity increases. Thus, by finding one or more nodes connected to each node by multiple edges where the sum of edge lengths is shorter than the other sums, we identified one or more pairs of similar companies across categories. To identify such pairs, we applied the Dijkstra's algorithm [18].

2.3.2. New Community Detection from the Similar Company Network

The Clauset-Newman-Moore greedy modularity maximization is a useful method for finding communities of nodes in node-edge networks. We considered the adjacency matrix A corresponding to a node-edge network, where the (i,j)th entry $a_{i,j}$ is 1 if the nodes i and j are linked to each other by an edge and 0 otherwise. We respectively denote the numbers of nodes linked to the nodes i and j as k_i and k_j. Moreover, for the new communities C_i and C_j to which the nodes i and j respectively belong, we introduced a function $\delta(C_i, C_j)$ that returns 1 if $C_i = C_j$ and 0 otherwise. In Clauset-Newman-Moore greedy modularity maximization, we focused on the value given using $a_{i,j}, k_i, k_j, \delta(C_i, C_j)$, and the number of edges m as:

$$Q := \frac{1}{2m} \sum_{i,j} \left(a_{i,j} - \frac{k_i k_j}{2m}\right) \delta(C_i, C_j). \qquad (5)$$

Note that Q grows larger as the number of edges within each community increases and those connecting to nodes in other communities decreases. Thus, we needed to seek a community division that maximizes the value of Q. We completed this process by the greedy algorithm [17] in the Clauset-Newman-Moore greedy modularity maximization. Note here that the number of resulting

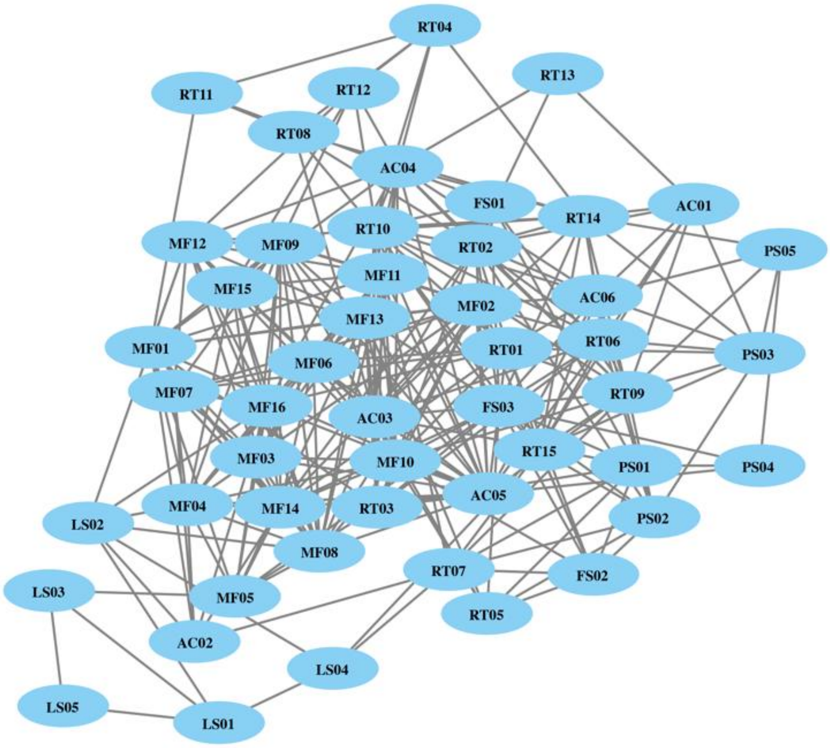

Figure 1. A graphical network of similar companies in Joyo based on Spearman's rank correlation coefficients.

communities is automatically determined depending on a network being analyzed.

3. Results and Discussion

In this section, we present results obtained by analyzing the similar company network.

3.1. The Most Similar Companies in the Same and Different Categories

Employing Dijkstra's algorithm, we obtained Table 2, where the values in parentheses in the second and third columns are the sums of the edge lengths between the two nodes corresponding to two similar companies, expressed to two decimal places. Table 2 responds to the request from Joyo's government to identify, for each company, only one most similar company from the same category and only one from a different category for the convenience of unifying the website's layout. We emphasize here that all companies can be ranked according to their similarity to each other.

In the second column of Table 2, we can observe AC05 most often in the agriculture category, MF08 most often in the manufacturing category, RT06 and RT09 most often in the retail trade category, FS03 most often in the food service category, PS03 and PS05 most often in the professional service category, and LS01 most often in the leisure service category. Thus, these companies have been concluded to have principal features of their own categories.

We now turn to the third column of Table 2. For the companies in the categories of agriculture, food service and professional service, the most similar companies in other categories are disparate. In contrast, in the categories of manufacturing and retail trade, we observe specific companies repeatedly occurring as the most similar company. In particular, for nine companies in the manufacturing category, AC05 is the most similar non-manufacturing company. In the retail trade category, more than half of the most similar non-retail-trade companies are agriculture companies. In the category of leisure services, two companies are most similar to AC02 and RT15, and the other three companies are similar to MF05.

3.2. Similar Companies Across Categories and Their Communities

Using the Clauset-Newman-Moore greedy maximization, we have identified three new company communities, as color-coded in Figure 2. We here emphasize that $Q \geq 0.3$ is generally considered to indicate that a significant community structure is found in a network, and $Q \approx 0.3227$ in this case. Moreover, by computing the average of each interview topic rating in each company community and focusing on the top five and bottom five average scores in each company community, we have

Table 2. The most similar companies in the same and different categories.

Company ID	Company ID in the same category	Company ID in the different category
AC01	AC06(1.95)	PS03(1.47)
AC02	AC05(3.82)	MF04(1.88)
AC03	AC04(2.22)	MF13(1.51)
AC04	AC05(1.91)	RT12(1.72)
AC05	AC06(1.82)	MF02(1.39)
AC06	AC05(1.82)	RT09(1.38)
MF01	MF04(1.61)	RT03(1.62)
MF02	MF10(1.36)	AC05(1.39)
MF03	MF08(1.14)	AC05(2.36)
MF04	MF16(1.39)	AC02(1.88)
MF05	MF14(1.78)	LS03(1.85)
MF06	MF07(1.25)	AC05(2.10)
MF07	MF06(1.25)	RT01(2.60)
MF08	MF03(1.14)	AC05(1.91)
MF09	MF08(1.31)	AC05(1.97)
MF10	MF02(1.36)	AC05(1.39)
MF11	MF08(1.30)	AC05(1.66)
MF12	MF09(1.31)	AC03(2.42)
MF13	MF08(1.59)	AC03(1.51)
MF14	MF04(1.56)	AC05(1.58)
MF15	MF09(1.60)	RT12(1.63)
MF16	MF04(1.39)	AC05(1.83)
RT01	RT06(2.22)	MF02(1.73)
RT02	RT06(1.46)	AC06(1.88)
RT03	RT06(1.99)	MF01(1.62)
RT04	RT08(1.63)	AC04(2.50)
RT05	RT15(3.40)	FS02(1.94)
RT06	RT09(1.32)	AC06(1.58)
RT07	RT10(2.14)	PS02(2.14)
RT08	RT04(1.63)	AC03(2.76)
RT09	RT14(1.23)	AC06(1.38)
RT10	RT09(1.46)	AC06(1.88)
RT11	RT08(1.95)	FS01(2.39)
RT12	RT04(1.93)	MF15(1.63)
RT13	RT02(2.29)	AC01(2.33)
RT14	RT09(1.23)	AC04(1.77)
RT15	RT14(1.61)	PS02(1.42)
FS01	FS03(2.45)	PS01(1.97)
FS02	FS03(2.35)	PS02(1.63)
FS03	FS02(2.35)	MF10(1.63)
PS01	PS04(2.28)	FS02(1.88)
PS02	PS03(2.15)	RT15(1.42)
PS03	PS05(1.89)	AC01(1.47)
PS04	PS05(2.24)	AC05(2.26)
PS05	PS03(1.89)	RT14(2.19)
LS01	LS03(1.66)	MF05(3.51)
LS02	LS04(1.87)	AC02(2.08)
LS03	LS01(1.66)	MF05(1.85)
LS04	LS02(1.87)	RT15(2.39)
LS05	LS01(1.83)	MF05(4.67)

grasped the features of each company community, as shown in Table 3, where the contents within the parentheses in the third column are the focused topics and averages expressed to two decimal places. The company communities colored in green, blue, and red in Figure 2 are respectively Comm01, Comm02, and Comm03, in Table 3. From Table 3, we immediately have seen that Comm01 and Comm02 primarily contain retail trade companies and manufacturing companies, respectively, while the companies in Comm03 are all in the leisure service category.

4. Conclusions

In this paper, by analyzing data collected through interviews with company presidents and managers, we have captured the industrial structure of Joyo City. We

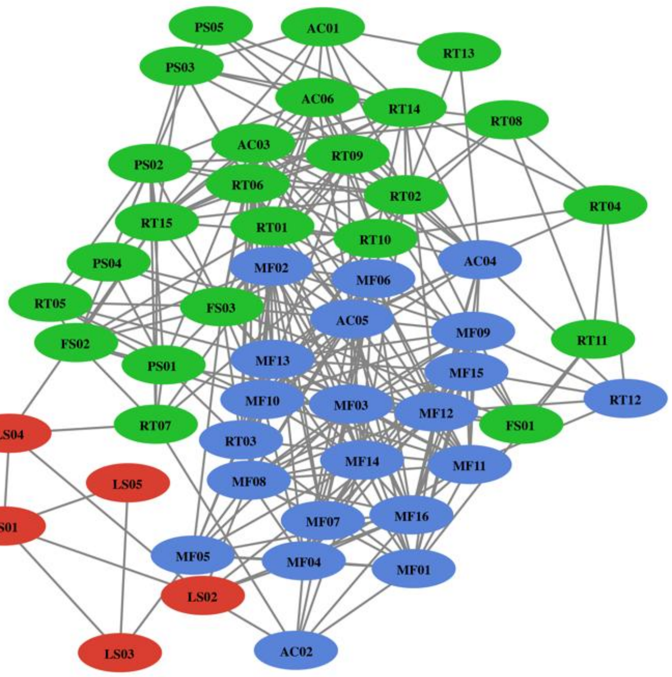

Figure 2. Three communities of companies colored in the similarity company network.

Table 3. Three communities and their major features.

Community ID	**Category rates of companies**	**Five major features**
Comm01	Agriculture: 12.5% Retail trade: 54.2% Food service: 12.5% Professional service: 20.8%	・High relevance to region (Topic 14: 3.62) ・Narrow target customer (Topic 20: 3.52) ・The majority of visitors are older (Topic 21: 3.45) ・Small number of employees (Topic 2: 2.01) ・Small company's site area (Topic 3: 2.19)
Comm02	Agriculture: 14.3% Manufacturing: 76.2% Retail trade: 9.5%	・Attractive for domestic visitors from the outside (Topic 19: 4.07) ・Large capital or gross sales (Topic 1: 3.72) ・Short distance from major roads (Topic 11: 3.59) ・High capital expenditures and advertising expenses (Topic 24: 3.55) ・Large company's site area (Topic 3: 3.51)
Comm03	Leisure services: 100.0%	・Large company's site area (Topic 3: 4.46) ・Long visitor dwell time (Topic 17: 4.06) ・Long working hours (Topic 4: 3.86) ・Few collaborations with other companies (Topic 28: 3.86) ・Casual ambience (Topic 13: 1.82)

emphasized the features of each company by using low-rank approximation based on matrix singular value decomposition. We then presented a graphical network that reveals one or more similar companies for each company with the help of Spearman's correlation method. Next, by applying Dijkstra's algorithm to the similar company network, we have clarified the most similar companies in the same and different industrial categories for each company. Moreover, based on the Clauset-Newman-Moore greedy modularity maximization, we have found company communities in terms of feature similarity across categories.

Future work should consider how these results and techniques can be applied to formulate policies and establish grants that promote the development of Joyo's industry, such as building cross-category collaboration. We also would like to consider whether our industrial analysis can be generalized and applied to other cities. As municipalities grow larger, the number of companies tends to increase, so we may need to consider alternative data collection methods instead of interviews with company presidents and managers. In very large municipalities, the amount of data collected will be enormous, so we may also need computational strategies to smoothly process big data statistically.

Author Contributions: Conceptualization, K.T., M.I. and M.S.; methodology, K.T.; software, K.T.; validation, K.T.; formal analysis, K.T.; investigation, M.I. and M.S.; resources, K.T., M.I. and M.S.; data curation, M.S.; writing—original draft preparation, K.T.; writing—review and editing, M.I. and M.S.; visualization, K.T.; supervision, M.I.; project administration, M.I.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Grant-in-Aid for Early-Career Scientists No. 21K13844 from the Japan Society for the Promotion of Science.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study is available upon request to the corresponding author.

Acknowledgments: We sincerely thank the anonymous reviewers for their constructive feedback and valuable suggestions.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Berisha, B., Mëziu, E., and Shabani, I. (2022). Big Data Analytics in Cloud Computing: An Overview, *Journal of Cloud Computing*, Vol. 11, 24. doi:10.1186/s13677-022-00301-w.
2. Mishra, A., and Kumar, G. (2021). Big Data Analytics on AWS Cloud, *International Journal of Engineering Research & Technology (IJERT)*, Vol. 10, No. 04. doi:10.17577/IJERTV10IS040325.
3. Schwertner, K. (2017). Digital Transformation of Business, *Trakia Journal of Sciences*, Vol. 15, Suppl.1, 388–393. doi:10.15547/tjs.2017.s.01.065.
4. Santos, M. Y., e Sá, J.O., Costa, C., Galvão, J., Andrade, C., Martinho, B., Lima, F. V., and Costa, E. (2017). A Big Data Analytics Architecture for Industry 4.0, *Recent Advances in Information Systems and Technologies. WorldCIST 2017. Advances in Intelligent Systems and Computing*, Vol. 570, 175–184. doi:10.1007/978-3-319-56538-5_19.
5. Janík, S., Szabó, P., Mĺkva, M., and Mareček-Kolibiský, M. (2022). Effective Data Utilization in the Context of Industry 4.0 Technology Integration, *Applied Sciences*, Vol. 12, No. 20, 10517. doi:10.3390/app122010517.
6. Izzo, F. (2019). Management Transition to Big Data Analytics: Exploratory Study on Airline Industry, *International Business Research*, Vol. 12, No. 10, 48-56. doi:10.5539/ibr.v12n10p48.
7. Makita, Y., and Fujiwara, M. (2018). Joint Endeavors of Public and Private Sectors in Open Data Utilization Activities: The Advanced Prefecture, Fukui. The Data City, Sabae. (in Japanese), *Journal of Information Processing and Management*, Vol. 60, No. 11, 798–808. doi:10.1241/johokanri.60.798.
8. Hori, R., Uga, T., Urata, M., Endo, M., Yasuda, T., and Yamada, M. (2022). Promotion of Big Data Utilization for Tourism-Based Community Development: Analysis of Traffic Routes at Intersections, *Journal of Global Tourism Research*, Vol. 7, No. 1, 75–79. doi:10.37020/jgtr.7.1_75.
9. Machi·Hito·Shigoto Soseihombu. (2015). RESAS (in Japanese), from https://resas.go.jp/#/13/13101.
10. Kabushikigaisha Deinoshisutemu. (2011). Nippon no Shacho.TV (in Japanese), from https://j-president.net/.
11. Koyama, N., Yamada, T., and Yamada, T. (2024). A Proposal for an Ultra-fast Flood Analysis Method Using Dynamic Mode Decomposition (in Japanese), *Advances in River Engineering*, Vol. 30, 459–464. doi:10.11532/river.30.0_459.
12. Fukumoto, K., Maeda, A., and Nakaizumi, F. (2016). The Analysis of Unstable Modes of Sit-to-Stand Movement with Restraint of Human Trunk (in Japanese), *The Proceedings of JSME Annual Conference on Robotics and Mechatronics (Robomec)*, Vol. 2016, 1P1-12b5. doi:10.1299/jsmermd.2016.1P1-12b5.
13. Fukui, K., Tomio, S., and Tsujita, K. (2014). 1P1-D01 A study on Mechanical design of a double-arm type power assist device based on the motion analysis of the human meal behavior(Welfare Robotics and Mechatronics (2)) (in Japanese), *The Proceedings of JSME Annual Conference on Robotics and Mechatronics (Robomec)*, Vol. 2014, 1P1-D01. doi:10.1299/jsmermd.2014._1P1-D01_1.
14. Spearman, C. (1904). The Proof and Measurement of Association between Two Things, *The American Journal of Psychology*, Vol. 15, No. 1, 72-101. doi:10.2307/1412159.
15. Pearson, K. (1896). VII. Mathematical Contributions to the Theory of Evolution.—III. Regression, Heredity, and Panmixia, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, Vol. 187, 253–318. doi:10.1098/rsta.1896.0007.
16. Imai, Y., and Kubo, A. (2019). Relationships between Body Composition and Body Image in Young Females (in Japanese), *Rigakuryoho Kagaku*, Vol. 34, No. 5, 713–717. doi:10.1589/rika.34.713.
17. Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding Community Structure in Very Large Networks, *Physical Review E*, Vol. 70, No. 6, 066111. doi:10.1103/PhysRevE.70.066111.
18. Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik*, Vol. 1, No. 1, 269–271. doi:10.1007/BF01386390.