



Algorithm for Defining Clusters based on Input–Output Tables: Case of Construction Cluster of Russia

Tatiana Kudryavtseva¹, Angi Skhvediani^{1*}, Valeriia Iakovleva¹, Alina Cherkas²

¹Graduate School of Industrial Economics, Peter the Great St.Petersburg Polytechnic University, St.Petersburg, Polytechnicheskaya, 29, 195251, Russia

²Laboratory of Industrial Data Streaming Processes, Peter the Great St.Petersburg Polytechnic University, St.Petersburg, Polytechnicheskaya, 29, 195251, Russia

Abstract. This research presents an algorithm for cluster identification based on input–output matrixes. Authors present an algorithm for downstream and upstream analysis of the symmetrical input–output matrix, which allows definition of the top input and output suppliers and consumers for each industry. As a result of the algorithm, related industries and clusters can be defined. The program, which implements the proposed algorithm, was written using Python. In this paper, the algorithm is applied to the analysis of «Construction» industry of Russia. We used the latest input–output matrix available for Russia for 2016, which contained information on 98 industries. We defined clusters and industries that are the top suppliers and consumers of the «Construction» industry. Among the top suppliers for the «Construction» industry are «Metal manufacturing», «Automotive cluster», and «Chemical products cluster», which account for 15.01%, 9.63%, and 5.95% of overall consumption, respectively. Top consumers of the «Construction» industry are «Public administration and defense; compulsory social security», «Real estate activities», and «Human health and social work activities», which account for 23.3%, 12.19%, and 6.26%, respectively, in the volume of output. The proposed algorithm can be used for analyzing input–output matrixes and cluster identification. Using the results of its application, the decision-makers can elaborate on policy for supporting the cluster-based development of the regions.

Keywords: Construction cluster; Input–output matrixes; Regional specialization

1. Introduction

Currently, in the context of forming Industry 4.0 and digitalizing industrial enterprises, innovatively active industrial clusters start playing a key role (Tashenova et al., 2020). Cluster development positively impacts the regional economy and employment levels (Moeis et al., 2020). Identification of industrial clusters, analysis of relationships between cluster presence, and economic performance are widely explored topics (Ketels and Protsiv, 2020). One of the main ideas behind the cluster is maximization of agglomeration effects, which can arise from localization, competition, and knowledge exchange. Researchers use two main approaches in order to identify clusters (Skhvediani and Sosnovskih, 2020). The first approach is based on analysis of localization quotients of the related industries. This approach allows to identify industries that have relatively high concentration at the regional level compared to the country average (Slaper et al., 2018; Kudryavtseva et al., 2020). The

*Corresponding author's email: shvediani_ae@spbstu.ru, Tel.: +7-812-534-7331
doi: [10.14716/ijtech.v12i7.5354](https://doi.org/10.14716/ijtech.v12i7.5354)

data on the values of regional localization coefficients make it possible to assess their competitive specialization and allow them to identify efficient regional clusters (Pavlov et al., 2015). The localization coefficient characterizes the concentration of enterprises belonging to a certain industry, but clusters can consist of enterprises of various industries connected by buyer-supplier relations. Another approach for cluster identification is based on the downstream and upstream analysis of symmetrical input–output matrix (Titze et al., 2011; Morrissey and Cummins, 2016). Input-Output Analysis shows that, in the overall economy, there are interrelationships and interdependencies between sectors. The tables present data on sales or shipments between companies in different industries, which allow to calculate what portion of its resources a company in one industry purchased from enterprises of other industries. The idea of cluster analysis lies in identifying strong patterns of cross-industry interaction. Groups of industries with strong connections are called value-added production chains or clusters. Despite the lack of practice in compiling cross-industry balance sheets at the regional level, such tables provide an overview of the possible relationship between enterprises and industries. Also, a combination of these methods can be used, as was done by Delgado et al. (2015).

Research on defining clusters based on input–output analysis is quite popular in scientific literature. For example, a worldwide, input-output network was built based on the global, multi-regional, input-output tables (Cerina et al., 2015). Also, input–output analysis was used to identify clusters of industries with high carbon emissions in Japan (Kanemoto et al., 2019), to identify clusters in German industries with intensive research and development (Kosfeld and Titze, 2017), and to find industrial clusters in the Beijing-Tianjin-Hebei region in China (Guo et al., 2019). Analysis of input-output tables was carried out in a study by Thai scientists to identify Thai rubber cluster (Tengsuwan et al., 2019). Indonesian researchers worked to identify the role of agricultural sectors in Jambi economy (Fitri et al., 2019). For Russian cases, this method was used to analyze 40 industries' input-output flows for 2007 (Markov and Markova, 2012), and the composition of the textile cluster of the Ivanovo region (Valitova et al., 2021). Therefore, there is quite a limited amount of research dedicated to cluster identification, based on input–output tables.

In this paper, we present a program algorithm, which allows to identify top suppliers and consumers. This algorithm was used in the «Construction» industry using the 2016 input-output table (Federal State Statistics Service, 2020).

2. Methods

For this research, we have used the latest available input-output matrix for Russian industries for 2016. The symmetrical input–output matrix contained data on 98 industries. We have utilized the maximum method, which consists of downstream and upstream algorithms, in order to analyze links between industries (Peeters et al., 2001; Luo and Yan, 2009).

A downstream cluster link is established when a supply is made by a specific supplier to its main customer and is a relative value that exceeds an arbitrarily set threshold. An upstream cluster link is established when a supply is made to a specific customer from its main supplier and represents a relative value that exceeds an arbitrarily set threshold. We statistically calculated threshold values by finding values of 99 and 95 percentiles of all ranked values of intermediate consumption of the table. Groups of industries connected by upstream and downstream links are clusters. In this study, we only consider the relationship of the construction industry with other clusters and sectors. In order to conduct this analysis, we have proposed an algorithm, which allowed us to automatically

identify top input and output types of activities for each industry. The code and step for program implementation are presented in Table 1.

The program is written using Python and designed to identify industrial clusters in the regional economy of the Russian Federation. The programs use the following analytical packages: Numpy, Pandas, Matplotlib, and Seaborn. The program can be used in the framework of the cluster analysis of the economy in the context of pairwise and multiple relationships of economic sectors. The program allows us to visualize the relationship between different types of activities, using as input data a symmetric table "input-output" for all sectors of the economy in accordance with the nomenclature adopted in the territory of the Russian Federation. The results of the program are the visualization of pairwise and multiple dependencies of the branches of the economy of the constituent entities of the Russian Federation in the form of tables and graphs.

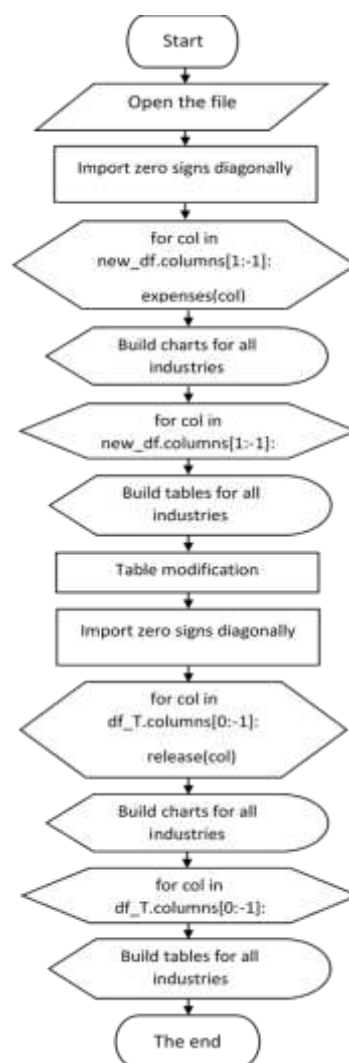


Figure 1 Algorithm for defining clusters based on input–output matrix

After implementing this algorithm, we obtained significant values above 5% and built a map of interconnected industries and defined clusters based on input and output flows. In this research, we will only discuss the construction cluster and types of activities and clusters related to it. Visualization of the relations between industries was performed in MS Visio. Further, on the basis of these "input-output" tables, the structure of consumption of domestic and imported products for the key suppliers in the construction industry was

analyzed.

3. Results and Discussion

3.1. Results of Input-output Table Analysis for «Construction» Industry

The results of the algorithm were used to visualize the relations between industries using MS Visio. Figure 2 presents results for the construction cluster. Classifier code of industry represents Russian National Classifier of Types of Economic Activity (OKVED) code of the industry, which is used in Russia as an analog of NACE – 2 codes in Europe.

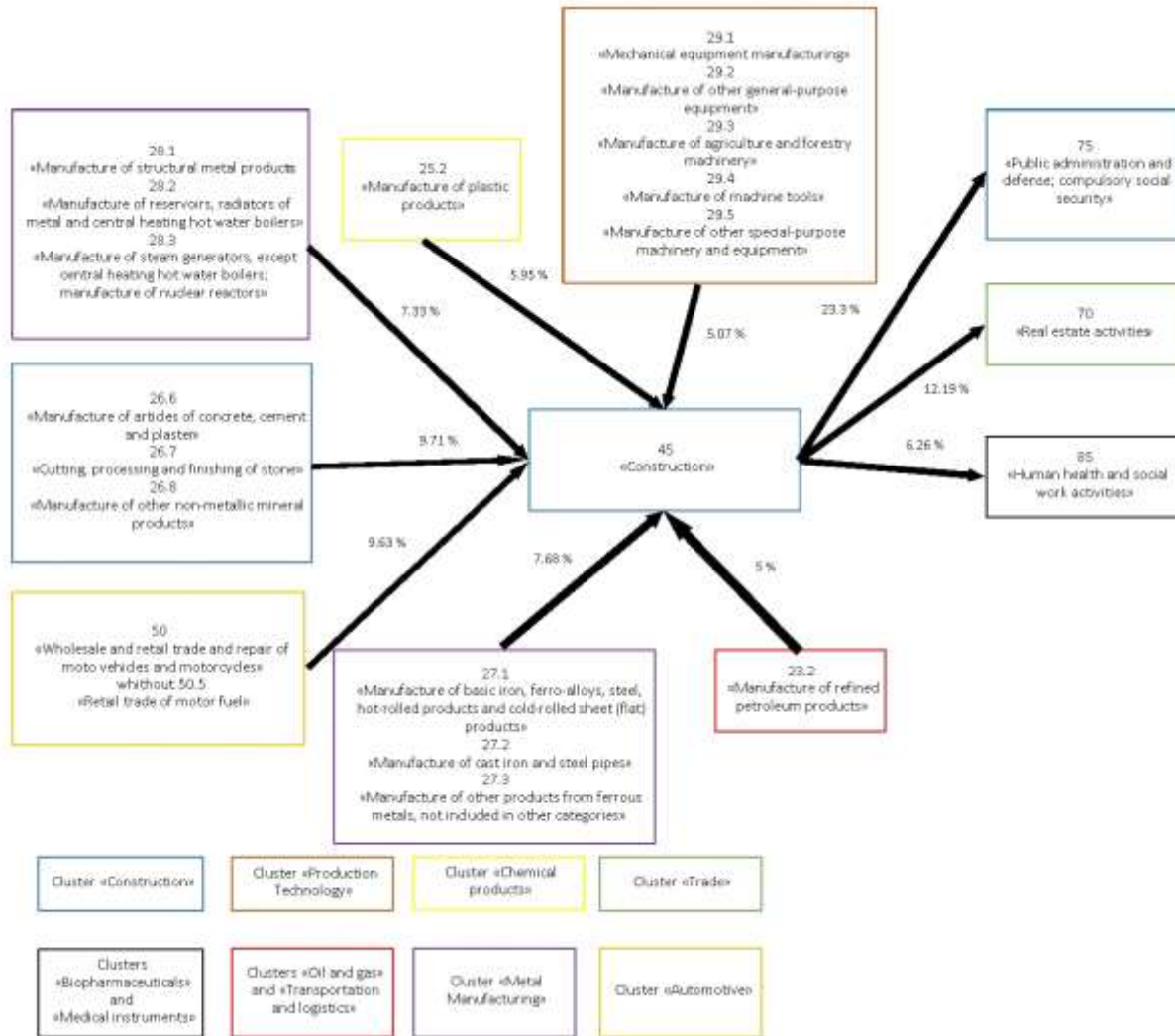


Figure 2 Visualization of input-output links of construction cluster

«Construction» (45) acquires various metal structures from the «Manufacture of structural metal products» (28.1), «Manufacture of reservoirs, radiators of metal, and central heating hot water boilers» (28.2), and «Manufacture of steam generators, except central heating hot water boilers; manufacture of nuclear reactors» (28.3). They account for 7.33% in the structure of costs of the «Construction» industry (45).

To create its own products, the «Construction» industry (45) purchases metal products from the following industries: «Manufacture of basic iron, ferro-alloys, steel, hot-rolled steel and cold-rolled sheet (flat) rolled products» (27.1), «Manufacture of cast iron and steel pipes» (27.2), and «Manufacture of other products from ferrous metals, not included in other categories» (27.3). They occupy 7.68% of the cost structure. The listed industries are

part of the Metal Manufacturing cluster.

The necessary production equipment is supplied by the following industries: «Mechanical equipment manufacturing» (29.1), «Manufacture of other general-purpose equipment» (29.2), «Manufacture of agriculture and forestry machinery» (29.3), «Manufacture of machine tools» (29.4), and «Manufacture of other special-purpose machinery and equipment» (29.5). They account for 5.07% in the structure of construction costs (45). These industries are part of the «Production Technology» cluster.

«Construction» purchases plastic materials and products supplied by the «Manufacture of plastic products» (25.2) industry, which is a part of the «Chemical products» cluster.

The «Wholesale and retail trade and repair of moto vehicles and motorcycles»—with the exception of the branch «Retail trade of motor fuel» (50.5), which is part of the cluster «Automotive»—supplies construction machinery. Its share in the cost structure of the industry «Construction» is equal to 9.63%.

The costs of the «Construction» industry (45) for the intermediate consumption of the products of the «Manufacture of refined petroleum products» industry (23.2) are 5%. In construction, petrochemical products are used as materials, energy carriers for special equipment and machinery, and for its maintenance. The Oil Products Industry (23.2) is also included in the Oil and Gas and Transport and Logistics clusters.

The industries «Manufacture of articles of concrete, cement and plaster» (26.6); «Cutting, processing and finishing of stone» (26.7); and «Manufacture of other non-metallic mineral products» (26.8) supply the basic production materials for the industry «Construction» (45). They account for the largest share of «Construction» industry costs, namely 9.71%.

The «Public administration and defense; compulsory social security» (75) industry consumes output from «Construction»; namely, the construction of infrastructure facilities to improve the welfare of society. It is the main consumer of the industry under study. Its share is 23.3% in the volume of output of the «Construction» industry (45). The above industries are part of the Construction cluster.

The «Real estate activities» industry (70) consumes 12.19% of the output of the «Construction» industry; namely, it purchases, sells, and leases property. The «Real estate activities» industry (70) is part of the Trade cluster.

The «Human health and social work activities» sector (85) consumes 6.26% of the output of the Construction sector for the construction of polyclinics, hospitals, and other medical institutions. The industry is also part of the «Biopharmaceuticals» and «Medical Instruments» clusters.

3.2. Structure of Consumption of Domestic and Imported Products by the Construction Industry

The large-scale global economic crisis caused by Covid-19, a new coronavirus infection, and the subsequent restrictive measures resulted in a number of negative trends in the leading sectors of the economy, including construction.

The construction industry remains extremely dependent on the macroeconomic environment. The introduction of anti-Russian sanctions, the rise in prices of imported materials and components, and the decline of Russian enterprises led to an increase in prices for construction materials.

A significant increase in prices for the products of industries supplying construction materials will cause a rise in the price of construction works and related industries, and will affect the functioning of clusters. More than half of the developers had to deal with a more than 5% increase in prices for construction materials (Badusheva and Palagin, 2020). The increase in prices for materials will affect the cost of construction work performed

(Semenenko, 2020). The coronavirus pandemic has already had, and will continue to have, a significant impact on the construction market in the future.

Table 1 presents data on the structure of domestic and imported intermediate consumption by the construction industry, from which it follows that imported goods occupy a significant share in the consumption structure.

Table 1 Structure of consumption of domestic and imported products by the construction industry for its main suppliers

OKVED code	Supplier Industry Name	Consumption by «Construction» industry				
		Domestic consumption		Import		Share in total imports of the «Construction» industry
		Million rubles	%	Million rubles	%	
29.1	«Mechanical equipment manufacturing»					
29.2	«Manufacture of other general-purpose equipment»					
29.3	«Manufacture of agriculture and forestry machinery»	296,091	62.54	177,373	37.46	23.31
29.4	«Manufacture of machine tools»					
29.5	«Manufacture of other special-purpose machinery and equipment»					
25.2	«Manufacture of plastics products»	347,162	78.91	92,792	21.09	12.2
28.1	«Manufacture of structural metal products»					
28.2	«Manufacture of reservoirs, radiators of metal, and central heating hot water boilers»	427,801	88.92	53,324	11.08	7.01
28.3	«Manufacture of steam generators, except central heating hot water boilers; manufacture of nuclear reactors»					
27.1	«Manufacture of cast iron, ferro alloys, steel, hot-rolled products, and cold-rolled sheet (flat) products»					
27.2	«Manufacture of cast iron and steel pipes»	448,268	91.40	42,156	8.60	5.54
27.3	«Manufacture of other products from ferrous metals, not included in other categories»					
26.6	«Manufacture of articles of concrete, cement, and plaster»					
26.7	«Cutting, processing, and finishing of stone»	566,393	94.09	35,603	5.91	4.68
26.8	«Manufacture of other non-metallic mineral products»					
23.2	«Manufacture of refined petroleum products»	291,610	99.07	2,746	0.93	0.36

Products of «Mechanical equipment manufacturing», «Manufacture of other general-purpose equipment», «Manufacture of agriculture and forestry machinery», «Manufacture

of machine tools», and «Manufacture of other special-purpose machinery and equipment» get imported the most: more than a third of goods are imported from these industries.

4. Conclusions

This paper contributes to the topic of cluster identification based on the input–output tables. We developed the program in Python and presented the programming algorithm for downstream and upstream analysis of the symmetric input and output tables. We used this algorithm at example of Russian data on 98 industries for 2016. In this paper, we presented the results for the «Construction» industry. In particular, we defined top suppliers and consumers of this industry and identified clusters that are related to it. To the best of our knowledge, this is one of the first works that used data on Russian input-output linkages for the last years in order to define related industries and form clusters. Previous works about Russia have mainly used localization quotients for cluster identification and international data on linkages between industries. There are several limitations to our research. The first limitation is that we determined the connectivity at the country level; we did not consider the presence of these clusters in the regions. The second is that we did not use coefficient on the localization employment variable in order to check whether related industries located at the same region. The third is that we extrapolate data from 2016 to the present date. Future research should consider the development of the programming algorithm to receive the map of interconnected industries automatically. In addition, it should allow to identify changes in structure of interconnected industries in time. Input–output table for 2021 is expected to be published by a Russian statistical service in 2022–2023. In the future, it will identify the spatial localization of clusters at the regional level, considering intersectoral relations.

Acknowledgements

This research was funded by the Russian Science Foundation. Project No. 20-78-10123.

References

- Badusheva, V., Palagin, A., 2020. Development of the Construction Industry under the Influence of COVID-19. *Vestnik Akademii Znaniy*, Volume 39(4), pp. 81–85
- Cerina, F., Zhu, Z., Chessa, A., Riccaboni, M., 2015. World Input-Output Network. *PloS One*, Volume 10(7), pp. 1–21
- Delgado, M., Porter, M., Stern, S., 2015. Defining Clusters of Related Industries. *Journal of Economic Geography*, Volume 16(1), pp. 1–38
- Fitri, Y., Amir, A., Murdi, S., Syafaruddin., 2019. Using Input-Output Analysis Approach to Identify the Role of Agricultural Sectors in Jambi Economy. *Russian Journal of Agricultural and Socio-Economic Sciences*, Volume 1(8), pp. 552–562
- Federal State Statistics Service, 2020. Official Site Federal State Statistics Service. Available Online at <https://rosstat.gov.ru/accounts>, Accessed on October 21, 2021
- Guo, J., Lao, X., Shen, T., 2019. Location-Based Method to Identify Industrial Clusters in Beijing-Tianjin-Hebei Area in China. *Journal of Urban Planning and Development*, Volume 145(2), June 2019
- Kanemoto, K., Hanaka, T., Kagawa, S., Nansai, K., 2019. Industrial Clusters with Substantial Carbon-Reduction Potential. *Economic Systems Research*, Volume 31(2), pp. 248–266
- Ketels, C., Protsiv, S., 2020. Cluster Presence and Economic Performance: A New Look Based on European Data. *Regional Studies*, Volume 55(2), pp. 208–220
- Kosfeld, R., Titze, M., 2017. Benchmark Value-added Chains and Regional Clusters in R&D-

- Intensive Industries. *International Regional Science Review*, Volume 40(5), pp. 530–558
- Kudryavtseva, T., Skhvediani, A., Berawi, M.A., 2020. Modeling Cluster Development using Programming Methods: Case of Russian Arctic Regions. *Entrepreneurship and Sustainability Issues*, Volume 8(1), pp. 150–176
- Luo, S., Yan, J., 2009. Analysis of Regional Industrial Clusters' Competitiveness based on Identification. *In: International Conference on Electronic Commerce and Business Intelligence*, Beijing, China, pp. 471–474
- Markov, L.S., Markova, V.M., 2012. Revealing Reference Clusters: Methodical Questions and the Practical Application to the Domestic Industry. *Bulletin of Nsu Social Economic Sciences*, Volume 12(1), pp. 95–108
- Moeis, A.O., Desriani, F., Destyanto, A.R., Zagloel, T.Y., Hidayatno, A., Sutrisno, A., 2020. Sustainability Assessment of the Tanjung Priok Port Cluster. *International Journal of Technology*, Volume 11(2), pp. 353–363
- Morrissey, K., Cummins, V., 2016. Measuring Relatedness in a Multisectoral Cluster: An Input–Output Approach. *European Planning Studies*, Volume 24(4), pp. 629–644
- Pavlov, K., Rastvortseva, S., Cherepovskaya, N., 2015. A Methodological Approach to Identifying Potential Clusters in Regional Economy. *Regional Economics: Theory and Practice*, Volume 10(385), pp. 15–26
- Peeters, L., Tiri, M., Berwert, A., 2001. Identification of Techno-Economic Clusters using Input-Output Data: Application to Flanders and Switzerland. *Innovative Clusters: Drivers of national innovation Systems. In: OECD proceedings*, 251-272
- Semenenko, V.Y., 2020. Global Construction Market in the Context of the Covid-19 Pandemic: Trends and Factors of Influence. *ASR: Economics and Management*, Volume 3(32), pp. 313–316
- Skhvediani, A., Sosnovskikh, S., 2020. What Agglomeration Externalities Impact the Development of the Hi-tech Industry Sector? Evidence from the Russian Regions. *International Journal of Technology*, Volume 11(6), pp. 1091–1102
- Slaper, T.F., Harmon, K.M., Rubin, B.M., 2018. Industry Clusters and Regional Economic Performance: A Study Across U.S. Metropolitan Statistical Areas. *Economic Development Quarterly*, Volume 32(1), pp. 44–59
- Tashenova, L., Babkin, A., Mamrayeva, D., Babkin, I., 2020. Method for Evaluating the Digital Potential of a Backbone Innovative Active Industrial Cluster. *International Journal of Technology*, Volume 11(8), pp. 1499–1508
- Tengsuwan, P., Kidsom, A., Dheera-Aumpon, S., 2019. Economic Linkage in the Thai Rubber Industry and Cluster Identification: Input-Output Approach. *Asian Administration & Management Review*, Volume 2(2), pp. 147–159
- Titze, M., Brachert, M., Kubis, A., 2011. The Identification of Regional Industrial Clusters using Qualitative Input–Output Analysis (QIOA). *Regional Studies*, Volume 45(1), pp. 89–102
- Valitova, L., Sharko, E., Sheresheva, M., 2021. Identifying Industrial Clusters based on the Analysis of Business Ties: A Case of the Textile Industry. *Upravlenets – The Manager*, Volume 12(4), pp. 59–74