

Finding Sustainable Clusters in Supply Chains Dynamics via Graph Partitioning

Raffaele D' Ambrosio, Stefano Di Giovacchino, and Carmela Scalone(✉)

Department of Engineering and Computer Science and Mathematics, University of L'Aquila, Via Vetoio, Loc. Coppito, 67100 L'Aquila, Italy
{raffaele.dambrosio,stefano.digiovacchino,carmela.scalone}@univaq.it

Abstract. This paper regards the clustering problem of supply chains. From a mathematical point of view, it is well-known that they may be modeled by connected graphs. Through this work, following the algorithm in [9], we address our attention on finding the smallest possible perturbation of the adjacency matrix of the graph associated to such supply chain, such that makes it disconnected. On a numerical analysis perspective, this problem consists in studying suitable approximations of a gradient system via spectral theory. Differently from standard methods like NMF or spectral clustering, the presented approach enables structured and interpretable perturbations, allowing for targeted policy analysis and scenario-based cluster exploration. Finally, selected numerical experiments on the special case of Japanese automotive industries will be reported.

Keywords: graph partitioning · differential equations · supply chain

1 Introduction

Nowadays, industrial and environmental sustainability is capturing a wider and wider attention among the scientific community and it represents an emergent real-world open issue. In particular, one of the most relevant addressed point is certainly the reduction of CO_2 emissions, that is one of the most effective responsible of climate warming. From a mathematical point of view, contributions concerning this topic are becoming increasingly common in the literature. In particular, we are interested in the possibility of studying the CO_2 emissions of supply chains of industries, through the clustering algorithms of graph theory. To a large extent, the interaction between industries may be studied by exploiting graphs and matrix theory; one of the main targets of such analysis has been oriented to the problem of finding appropriate algorithms for clustering nodes of the graphs. In a supply chain cluster, you group together actors (industries, sectors, or suppliers) that are closely interconnected and that contribute significantly to environmental impacts (e.g., CO_2 emissions) through energy and resources embedded in products. Members of a cluster share similar features, such as high levels of embodied energy or emissions, because their activities are

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15650, pp. 366–374, 2025.
https://doi.org/10.1007/978-3-031-96962-1_26

closely related (e.g., steel production and automobile manufacturing). By identifying clusters, efforts to improve energy efficiency or reduce emissions can be focused on the supplier groups that have the greatest overall impact. Not a single isolated actor or product is analyzed, but the entire network of relationships in the supply chain, improving the effectiveness of environmental strategies. Knowing which clusters are most critical helps design policies that have a broader impact and focus investments on the most influential areas.

In the scientific literature, there are few interesting contribution on that direction [7,8,13]. They mainly apply numerical algorithms, such as Fiedler partitioning and nonnegative matrix factorizations, for graphs clustering. They are based on associating a weighted graph with the supply chain using Leontief's model and then partitioning it using the above algorithms to identify clusters. The analysis is carried out considering the case study of the Japanese automotive industry. In this paper, we are interested in the same analysis of the aforementioned supply chain of clusters through the use of an algorithm that is perturbative in its own nature. Such a method has been introduced in [9] and it relies on finding the smallest possible perturbation of the *adjacency matrix* such that the corresponding graphs becomes disconnected ($\lambda_2 = 0$, being λ_2 the second eigenvalue of the *Laplacian* matrix associated to the graph.). The main tool of such method is the numerical analysis of a suitable gradient system. In this paper, first, using recursively the aforementioned approach, we provide the clusters of the graph, based on similar CO_2 emissions and then we will detect those most impacting on that direction. The reason for proposing the use of this method lies in the fact that it is robust with respect to the phenomenon of ambiguity, see [10], which we will discuss briefly in the paper, and is very flexible for studying the graph with respect to structured perturbations. While traditional approaches such as NMF or Fiedler-based spectral clustering extract partitions without structural constraints, the method adopted here allows for the explicit modeling of emission-related constraints and the design of scenario-specific perturbations. This makes it possible to simulate, for instance, the effect of freezing emission exchanges of specific industries or grouping them for policy purposes, offering interpretability and practical relevance often missing in standard graph partitioning. This paper proceeds as follow. In Sect. 2, we recall the existent methods for clustering graphs and we highlight their strength points and drawbacks. In Sect. 3, we illustrate the application of the algorithm in [9] to graph clustering problems. Finally, selected numerical experiments will be provided in Sect. 4.

2 Graph Partitioning and Supply Chain

We here introduce the notation and the mathematical framework for modelling supply chains. We denote as $n \in \mathbb{N}$ the number of all the commodities. The matrix $A \in \mathbb{R}^{n\times n}$ stands for the *direct requirement matrix*, i.e., its (i, j)-th element, (a_{ij}), $i, j = 1, \ldots, n$, indicates the input of commodity i directly required for producing one input of the commodity j. The matrix $B = (I - A)^{-1}$, where

I is the identity matrix of dimension n, is the *total requirement matrix*, i.e., its (i,j)-th element, $(b_{ij}), i,j = 1,\ldots,n$, indicates the input of commodity i directly and indirectly required for producing one input of the commodity j.
Moreover, to the generic commodity k, we associate the *final demand vector* $f_k \in \mathbb{R}^n$ with all zero elements except the one in position k and the output vector $x_k \in \mathbb{R}^n$, containing the commodities output induced by the final demand on the commodity k.
Also, for a given integer $m > 0$, we denote by $E \in \mathbb{R}^{m\times n}$, with elements $(e_{ij}), i = 1,\ldots,m,\ j = 1,\ldots,n$, the *direct factor input intensity matrix*; e_{ij} contains the quantity of factor inputs i, such as energy, that need for production of one unit of input of the commodity j. The vector $\tilde{E} \in \mathbb{R}^n$ contains all the consumptions of factor inputs needed for one unit of input of the commodity j, for any $j = 1,\ldots,n$.

If we denote by q the vector whose element $(q_i), i = 1,\ldots,n$ contains the total input factor that has been required directly or indirectly to produce the commodity i associated to the final demand of the commodity k, then, it is well-known that q satisfies the following equation

$$q = \mathrm{diag}(\tilde{E})x_k = \mathrm{diag}(\tilde{E})(I - A)^{-1}f_k. \tag{1}$$

By exploiting Taylor series arguments, after direct manipulations, one gets

$$Q = \mathrm{diag}(\tilde{E})\mathrm{diag}(f_k) + \mathrm{diag}(\tilde{E})A\mathrm{diag}(B_k)f_k, \tag{2}$$

where B_k is the k-th column of the matrix B and the matrix $Q \in {}^{n\times n}$ can be regarded as a weighted direct graph, whose elements (Q_{ij}) stand for the total factor input of commodity i. purchased to produce one input of j, finally needed for the final demand of the commodity k.
From Q, we derive a graph $G = (V,E)$, with $V = \{1,2,\ldots,n\}$ the set of nodes, $E = \{(i,j) : Q_{ij} > 0\}$ the set of edges and being Q_{ij} the weight to the edge connecting the nodes i and j.

From Q, we can construct the adjacency matrix Q^* as follows

$$Q^*_{ij} = \begin{cases} Q_{ij} + Q_{ji}, & i \neq j, \\ 0, & i = j. \end{cases}$$

Following [17], in [7] the authors makes use of Fiedler partition to cluster the graph in a manner that can be recalled as follows. Given two set C and D, the problem is to minimize the number of neglected edges and maximize the factor input intensity within each group. We may formulate this problem as follows.

$$\min \mathrm{Ncut}(C,D) = \frac{\mathrm{cut}(C,D)}{\sum_{u\in C} d_u} + \frac{\mathrm{cut}(D,C)}{\sum_{v\in D} d_v}, \tag{3}$$

with $C \cup D = V$ and $C \cap D = \emptyset$, where

$$\mathrm{Cut}(C,D) = \|Q^*_{CD}\| = \sum_{u\in C}\sum_{v\in D} Q^*_{uv}, \quad \mathrm{Cut}(D,C) = \|Q^*_{DC}\| = \sum_{u\in D}\sum_{v\in C} Q^*_{uv}$$

and

$$d_u = \sum_{v\in V} Q^*_{uv}, \qquad d_v = \sum_{u\in V} Q^*_{vu}.$$

The problem in (3) can be reformulated as follows [17]

$$\min \frac{y^T(D-Q^*)y}{y^T D y},$$

under the constraint $y^T Di = 0$, with $y_j \in \{1, -b\}$, $j = \{1,\ldots,n\}$, i is the vector of all ones and b is defined as

$$b = \frac{\sum_{i\in Y} d_i}{\sum_{i\in V\setminus Y} d_i}, \quad Y = \{i \in V | y_i = 1\}.$$

Here, the matrix D is a diagonal matrix whose diagonal elements are weighted degrees d_u. Also, the above problem can be reformulated in the following manner

$$\min \lambda = \frac{y^T(D-Q^*)y}{y^T D y}, \tag{4}$$

subject to $y^T Di = 0$, for any $y_i \in \mathbb{R}$, $i = 1,\ldots,n$. It is well-known that Equation (4) is solved by λ_2, were λ_2 is the second eigenvalue of the Laplacian matrix. Such eigenvalue is also denoted Friedler eigenvalue.

3 A Gradient System Method for Clustering

We refer to [9] to describe the numerical approach used in the present paper, see also [5,6]. Consider a graph $(\mathcal{V},\mathcal{E})$, where $\mathcal{V} = \{1,\ldots,n\}$ is the vertex set and edge set $\mathcal{E} \subset \mathcal{V}\times\mathcal{V}$. We suppose that the graph is *undirected*: if $(i,j) \in \mathcal{E}$, then $(j,i) \in \mathcal{E}$. With the undirected graph we associate *weights* w_{ij} for $(i,j) \in \mathcal{E}$, such that $w_{ij} = w_{ji} \geq 0$ for all $(i,j) \in \mathcal{E}$. The graph is *connected* if for all $i,j \in \mathcal{V}$, there exists a path from i to j for arbitrary length. Setting $w_{ij} = 0$ for $(i,j) \notin \mathcal{E}$, we have the symmetric weight matrix

$$W = (w_{ij}) \in \mathbb{R}^{n\times n}.$$

The degrees $d_i = \sum_{j=1}^n w_{ij}$ are collected in the diagonal matrix

$$D = \mathrm{diag}(d_i) = \mathrm{diag}(W\mathbb{1}), \qquad \text{where } \mathbb{1} := (1,\ldots,1)^T \in \mathbb{R}^n.$$

The *Laplacian matrix* $L = \mathrm{Lap}(W)$ is defined by

$$L = D - W, \quad \text{i.e.,} \quad \mathrm{Lap}(W) = \mathrm{diag}(W\mathbb{1}) - W.$$

All eigenvalues of L are nonnegative, and $L\mathbb{1} = 0$, so that $\lambda_1 = 0$ is the smallest eigenvalue of L. From Fiedler Theorem [11], we know that the connectivity of the graph is characterized by the second-smallest eigenvalue of L. In particular, the graph is disconnected if and only if $\lambda_2 = 0$. Moreover, if $0 = \lambda_2 < \lambda_3$, then the entries of the corresponding eigenvector orthogonal to $\mathbb{1}$ assume only two different values, of different sign, which mark the membership to the two connected components.

For this reason, the second smallest eigenvalue λ_2 of L is called *algebraic connectivity* of W. If λ_2 is simple, then the corresponding eigenvector is known as the *Fiedler vector*.

A typical strategy for partitioning a graph is to compute the Fiedler vector and to partition the graph according to the signs of the values of its components. This procedure becomes unreliable when a small perturbation of the weights yields a coalescence of the eigenvalues λ_2 and λ_3. In [9], the authors propose an approach based on finding the minimal norm perturbation of the weigth matrix W able to disconnect the graph, i.e. such that the second smallest eigenvalue of the corresponding perturbed Laplacian is zero.

The structure of the algorithm can be summarized as follows:

1. Given $\varepsilon > 0$, we search for a symmetric matrix $E = (e_{ij}) \in \mathbb{R}^{n\times n}$ with the same sparsity pattern as W (i.e., $e_{ij} = 0$ if $w_{ij} = 0$), of unit Frobenius norm, with $W + \varepsilon E \geq 0$ (with componentwise inequality) such that the second smallest eigenvalue of $\mathrm{Lap}(W + \varepsilon E)$ is minimal. The obtained minimizer is denoted by $E(\varepsilon)$.
2. We look for the smallest value of ε such that the second smallest eigenvalue of $\mathrm{Lap}(W + \varepsilon E(\varepsilon))$ equals 0.

For a given ε, the computation $E(\varepsilon)$ is performed using a suitable constrained gradient system for the functional

$$F_\varepsilon(E) = \lambda_2\big(\mathrm{Lap}(W + \varepsilon E)\big),$$

under the constraints of unit Frobenius norm and $W+\varepsilon E \geq 0$ and the symmetry and the sparsity pattern of E.

The optimal ε, denoted $\varepsilon^\star$, is computed by a Newton-bisection method.

The algorithm computes a partition of the graph as provided by the Fiedler vector corresponding to the weight matrix $W + \varepsilon^\star E(\varepsilon^\star)$.

4 Numerical Experiments

In this section we prove that partitioning problems associated to supply chains can be interesting applications of the perturbative method described in the previous section.

In particular, we refer to the data of the Japanese benchmark input-output table for the produced prices in 2000, see [12][1]. From the energy input-output database, the total energy consumption of each sector is calculated by summing the energy inputs for each sector. The total direct energy intensity of sector j, $\tilde{e}_j$, is then determined by dividing the total energy consumption by the sector's production output. Next, a diagonal matrix (393×393) containing the total energy intensities is created, and the adjacency matrix of energy consumption, Q, is constructed. This is done by substituting the energy intensity matrix $\text{diag}(\tilde{E})$, the input coefficient matrix A from the benchmark input-output table, and the diagonalized matrix for automobile domestic final demand (i.e., household consumption expenditure on automobiles, fixed capital formation in automobiles, and the increase in the stock of automobiles) $\text{diag}(F_{auto})$ into the right-hand side of Eq. 2. We observe that the graph contains several isolated nodes and a large connected component comprising 305 nodes. Naturally, we focus on this component to study its partitioning. Another important observation is the presence of many edges with very small weights, i.e., values close to zero. We proceed to remove these very weak connections (less than 10^{-3}) and truncate the decimal part of the remaining weights to improve the conditioning of the eigenvalue computations. Such techniques are commonly used to eliminate information that does not significantly influence the graph's structure, as discussed in [9]. We consider a partitioning of the graph, hierarchically applying the algorithm described in Sect. 2. We consider as parameters of the Newton-bisection method $\varepsilon_0 = 10^{-3}$ and the upper and lower bounds for the parameter ε, $\varepsilon_{lb} = 10^{-4}$ and $\varepsilon_{ub} = 10$. The tolerance parameter for accuracy is chosen as $tol = 10^{-6}$ and we consider a maximum number of nodes for each cluster equal to 40. Clearly, this choice is not unique. At the end we find 15 clusters. In Fig. 1, we report the structure of the fourth cluster obtained. In the following table, we report the results of the hierarchical partitioning:

For the correspondence between nodes and industries in the case studied we refer to the works [7,8]. Following [7,8],we measure the impact of industries in terms of CO_2 emissions in tons, considering the within-group sum estimates

$$\sum_{i \in C_k} \sum_{j \in C_k} Q^*.$$

In particular, after computing such quantity for all the obtained clusters in Table 1, we find that the most impactful are clusters number 9, 5 and 4, with sum estimates 2866, 985.74 and 409.54, respectively. The importance of using this algorithm compared to others used in the literature lies in the possibility of offering an interpretation of the results with respect to different scenarios. In fact, it is possible to study partitioning with respect to structured perturbations, for example, that have a certain sparsity pattern. The advantage is that this can be done with a simple projection onto the pattern considered in the method. This means that we are assuming that we are not perturbing certain weights

[1] The energy input data are from the website http://www.cger.nies.go.jp/publications/report/d031/eng/datafile/embodied/2000/401.htm.

Table 1. Distribution of the nodes in the graph clusters.

Cluster	Nodes
1	13, 23, 44, 47, 49, 50, 51, 52, 164, 179, 181, 185, 196, 216,
1	226, 227, 235, 271, 280, 303, 310
2	216, 226, 227, 235, 271, 280, 303, 310
3	151, 202, 203, 224, 273, 301, 307
4	15, 82, 89, 90, 91, 146, 149, 152, 153, 184, 190, 191, 194
4	199, 207, 219,220, 228, 237, 240, 241, 244, 257, 258, 262
5	4, 5, 6, 7, 10, 16, 17, 18, 19, 20, 25, 31, 32, 34, 37, 39, 45,
5	56, 58, 60, 62, 102, 120, 174, 176, 186, 236, 249, 252, 266,
5	267, 278, 279, 283, 293, 304, 311
6	11, 14, 24, 30, 72, 76, 77, 92, 96, 109, 110, 113, 117, 130,
6	160, 192, 195, 201, 208, 210, 225, 234, 247, 248, 251, 253,
6	255, 256, 269, 290, 296, 299, 305
7	217, 270
8	48, 57, 74, 88, 121, 122, 139, 144, 145, 159, 161, 162, 169,
8	218, 242, 261, 288, 289, 292, 294, 306
9	21, 29, 63, 69, 75, 101, 115, 119, 131, 133, 134, 135, 136,
9	141, 175, 182, 214, 215, 245, 246, 264, 265, 268, 287
10	9, 33, 36, 40
11	26, 55, 61, 66, 78, 84, 87, 94, 97, 98, 103, 105, 111,
11	114, 1, 16, 118, 127, 137, 138, 140, 143, 148, 150,
11	154, 170, 173, 187, 221, 230, 233, 238, 243, 263, 274,
11	275, 281, 300
12	67, 79, 80, 81, 85, 93, 95, 100, 112, 167, 177, 178,
12	180, 213, 232, 254, 260, 276, 282, 284, 285, 286, 298,
12	308, 309
13	59, 64, 65, 70, 73, 83, 147, 155, 166, 198, 206, 211,
13	212, 250, 259, 291, 297
14	22, 28, 68, 71, 123, 128, 156, 158, 163, 188, 193, 197,
14	200, 205, 229, 295
15	1, 2, 3, 8, 12, 27, 41, 42, 43, 53, 54, 86, 99, 104, 107,
15	108, 124, 125, 126, 129, 132, 157, 165, 168, 171, 172,
15	209, 222, 223, 231, 239, 277

(emissions) of certain nodes (industries). This makes it possible to provide an analysis of the clusters obtained for different configurations. The possibility of including membership constraints may also be of application interest, when one wants to constrain a group of industries in the same pattern for other environ-

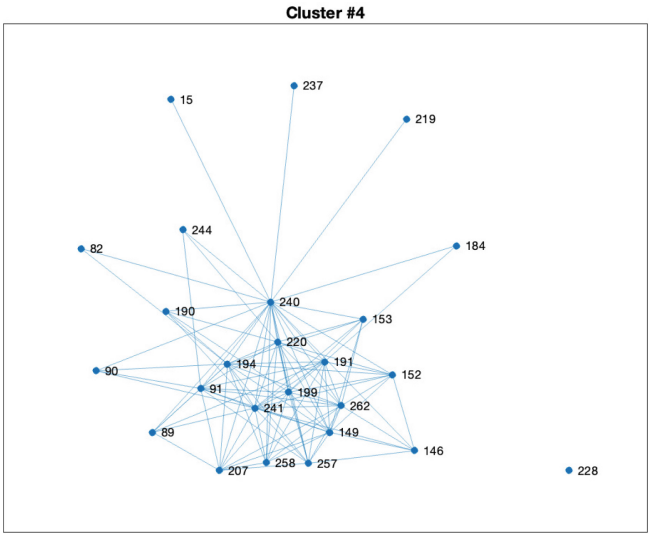


Fig. 1. Subgraph of the cluster 4.

mental impact reasons and then observe the repercussions in the hierarchical partitioning.

5 Conclusion

In this paper, we present a method suitable for partitioning a graph associated to a supply chain. This kind of analysis has provided a clustering useful to describe the impact of industrial groups in terms of CO_2 emissions. As also observed in [7], the direct effect of this analysis is meaningful in understanding how cluster-wide supply chain effort would contribute to saving energy and mitigating climate change. Future research perspectives could lead to the merging of clustering techniques in the supply chain domain with a machine learning context, see [1] and the extension of the theory presented in [3] to study the stability of stochastic networks associated with the presented model, as well as to establishing a formal paradigm of stochastic numerics (see [2,4,18]) for sustainability.

Acknowledgement. The work is supported by PRIN 2022 project "Stochastic numerical modelling for sustainable innovation" (20229P2HEA), CUP: E53C240022 80006. This work has been funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National National Centre for HPC, Big Data and Quantum Computing CN_00000013 - CUP: E13C22001000006).

References

1. Carissimo, N., D'Ambrosio, R., Guzzo, M., Labarile, S., Scalone, C.: Forecasting in shipments: comparison of machine learning regression algorithms on industrial applications for supply chain. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13957, pp. 462–470. Springer, Heidelberg (2023). https://doi.org/10.1007/978-3-031-36808-0_33
2. D'Ambrosio, R.: Numerical Approximation of Ordinary Differential Problems - From Deterministic to Stochastic Numerical Methods. Springer, Heidelberg (2023)
3. Di Giovacchino, S., Higham, D.J., Zygalakis, K.C.: A hierarchy of network models giving bistability under triadic closure. Multiscale Model. Simul. **20**(4), 1394–1410 (2022)
4. Higham, D.J., Kloeden, P.E.: An Introduction to the Numerical Simulation of Stochastic Differential Equations. SIAM (2021)
5. Guglielmi, N., Scalone, C.: An efficient method for non-negative low-rank completion. Adv. Comput. Math. **46**(2), 31 (2020)
6. Guglielmi, N., Scalone, C.: Computing the closest real normal matrix and normal completion. Adv. Comput. Math. **45**(5–6), 2867–2891 (2019)
7. Kagawa, S., Suh, S., Kondo, Y., Nansai, K.: Identifying environmentally important supply chain clusters in the automobile industry. Econ. Syst. Res. **25**(3) (2013)
8. Kagawa, S., Okamoto, S., Suh, S., Kondo, Y., Nansai, K.: Finding environmentally important industry clusters: multiway cut approach using nonnegative matrix factorization. Social Netw. **35**(3), 423–438 (2013)
9. Andreotti, E., Edelmann, D., Guglielmi, N., Lubich, C.: Constrained graph partitioning via matrix differential equations. SIAM J. Matrix Anal. Appl. **40**(1) (2019)
10. Andreotti, E., Edelmann, D., Guglielmi, N., Lubich, C.: Measuring the stability of spectral clustering. Linear Algebra Appl. **610**, 673–697 (2021)
11. Fiedler, M.: Algebraic connectivity of graphs. Czechoslov. Math. J. **23**(98), 298–305 (1973)
12. Matsushima, S., Kagawa, S., Nansai, K., Xue, J.: A comparison of deflation methods for carbon footprint calculations using Japanese data. Econ. Syst. Res. **36**(3) (2024)
13. Nansai, K., Kagawa, S., Suh, S., Inaba, R., Moriguchi, Y.: Simple indicator to identify the environmental soundness of growth of consumption and technology: eco-velocity of consumption. Environ. Sci. Technol. **41**, 1465–1472 (2007)
14. Nansai, K., Kagawa, S., Suh, S., Fujii, M., Inaba, R., Hashimoto, S.: Material and energy dependence of services and its implications for climate change. Environ. Sci. Technol. **43**, 4241–4246 (2009)
15. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**, 026113 (2004)
16. Newman, M.: Fast algorithm for detecting community structure in networks. Phys. Rev. E **69**, 066133 (2004)
17. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 888–905 (2000)
18. Milstein, G.N., Tretyakov, M.V.: Stochastic Numerics for Mathematical Physics. Scientific Computation, Springer-Verlag, Berlin (2004)