

Article

A Clustering Model for Three-Way Asymmetric Proximities: Unveiling Origins and Destinations

Laura Bocci ^{1,†}  and Donatella Vicari ^{2,*,†} 

¹ Department of Social and Economic Sciences, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy; laura.bocci@uniroma1.it

² Department of Statistical Sciences, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy

* Correspondence: donatella.vicari@uniroma1.it

† These authors contributed equally to this work.

Abstract: In many real-world situations, the available data consist of a set of several asymmetric pairwise proximity matrices that collect directed exchanges between pairs of objects measured or observed in a number of occasions (three-way data). To unveil patterns of exchange, a clustering model is proposed that accounts for the systematic differences across occasions. Specifically, the goal is to identify the groups of objects that are primarily origins or destinations of the directed exchanges, and, together, to measure the extent to which these clusters differ across occasions. The model is based on two clustering structures for the objects, which are linked one-to-one and common to all occasions. The first structure assumes a standard partition of the objects to fit the average amounts of the exchanges, while the second one fits the imbalances using an “incomplete” partition of the objects, allowing some to remain unassigned. In addition, to account for the heterogeneity of the occasions, the amounts and directions of exchange between clusters are modeled by occasion-specific weights. An Alternating Least-Squares algorithm is provided. Results from artificial data and a real application on international student mobility show the capability of the model to identify origin and/or destination clusters with common behavior across occasions.

Keywords: asymmetric dissimilarities; three-way data; clustering; origin/destination cluster; directed exchange; skew-symmetric matrix; Alternating Least-Squares



Citation: Bocci, L.; Vicari, D. A Clustering Model for Three-Way Asymmetric Proximities: Unveiling Origins and Destinations. *Symmetry* **2024**, *16*, 752. <https://doi.org/10.3390/sym16060752>

Academic Editor: Qing-Wen Wang

Received: 13 April 2024

Revised: 9 June 2024

Accepted: 11 June 2024

Published: 16 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Three-way proximity data are a structure of relationship data that is collected or measured between pairs of objects under multiple occasions, i.e., sources, settings, times, and experimental conditions. They extend the standard two-way proximity data, where the same objects are the entities in relation, to a three-dimensional framework. The additional dimension is given by the occasion on which the single pairwise proximity (similarity or dissimilarity) was collected. Three-way proximity data are typically arranged as a set of square data matrices (one for each occasion), each containing the proximities between all pairs of objects, and can be either symmetric or asymmetric.

In the latter case, where the relationship between the objects A and B is different from that between B and A , the asymmetry denotes a certain disequilibrium (imbalance) in the pairwise relationship. As a result, not only the presence of such relationships but also their directionality and magnitude, both of which are significant and cannot be ignored, define the true pattern of the data.

This phenomenon occurs in a variety of disciplines, including psychology, sociology, marketing research, social sciences, behavioral sciences, environmental sciences, and beyond, to name a few. Asymmetric proximity data can stem from diverse sources, such as judgments concerning the similarity between stimuli (confusion data), interactions among social actors (social networks), patterns of mobility, transitions between employment sectors, brand switching behaviors, or commercial transactions and trades between countries.

When dealing with asymmetric data, it may be of interest to unveil common behaviors of exchange in order to identify groups of objects that are primarily either origins or destinations of the exchanges, and in the presence of multiple occasions, such common patterns can also be analyzed and compared across them. For instance, the international student mobility between countries over several years yields a three-way asymmetric proximity array, where, in each annual matrix, the rows and columns correspond to the origins and destinations of the mobile students, respectively. The identification of possible pathways of student mobility over time can be an important source of information for policy makers.

Specially designed multiway models and methods are required to analyze such data. Within this framework, clustering three-way data is a complex and challenging task since asymmetric proximity matrices may subsume different classifications of the objects due to the heterogeneity across occasions, and the asymmetry may contain some information relevant to clustering efforts.

When analyzing asymmetric data, the asymmetry is often ignored by symmetrizing the proximities by averaging the two different values for each pair of objects. Nevertheless, asymmetry can be substantial and prominent, and deserves attention because it contains important information about the real patterns in the data, in terms of both the direction and the magnitude of the relationships.

A review of the literature on cluster analysis reveals a notable disparity: methods for asymmetric proximity data have received less attention compared with the large number of models and methods that have been proposed for symmetric proximity data, and even less for three-way data.

Clustering methods for asymmetric data have been developed for a single data matrix mainly following two approaches (see [1–3] for extended reviews of clustering methods for asymmetric data).

In the first approach, two different classification structures are estimated, one for the rows and one for the columns of the asymmetric matrix, since rows and columns are assumed to refer to two different sets of objects. Within this approach, the majority of the proposed clustering algorithms generally estimate hierarchical trees and identify non-overlapping clusters, i.e., each object belongs to only one cluster [4–7]. A non-hierarchical clustering algorithm, called GENNCLUS, that provides either overlapping or non-overlapping clusters for asymmetric or symmetric data has been proposed as a generalization of ADCLUS [8].

In the second approach, a single classification structure is estimated from both rows and columns of the asymmetric matrix because, in accordance with their true nature, they refer to the same set of objects. Within this approach, most of the methodologies extend the classical aggregative hierarchical methods [9–12], while few proposals concern non-hierarchical methods. Some of the latter are basically extensions of the k -means clustering [13–15]. In the same vein but with a different modeling, more recent proposals [16,17] concern non-hierarchical clustering methods that simultaneously fit both the symmetric and the skew-symmetric parts of the data resulting from the algebraic decomposition of the asymmetric matrix (see Section 1.1 for the definition). Specific clustering models for skew-symmetric data have been proposed by Vicari [18] and Vicari and Di Nuzzo [19]. In the first model, possible external covariates for the objects have been included, while the second proposal relies on the between-cluster effects modeled by Singular Value Decompositions that exploit the peculiar properties of the skew-symmetric matrices.

Regarding three-way asymmetric proximities, several models have been proposed within the framework of Multidimensional Scaling (MDS) (see [3] for a review of MDS methods for multiway asymmetric data), but to our knowledge, only one clustering model has been proposed in order to analyze three-way asymmetric proximity data. The proposal of Chaturvedi and Carroll [20] generalizes the INDCLUS model [21] to the asymmetric case by identifying two different sets of (overlapping) clusters of the objects (for the rows

and the columns of the data matrices, respectively) common to all occasions, while the three-way heterogeneity is accounted for by occasion-specific weights for the clusters.

We can observe that since Generalized INDCLUS [20] aims at finding two different sets of (possibly overlapping) clusters of objects, the identification of origins and destinations of exchange patterns is rather complicated from an interpretative point of view. This is because the clusters are only partially matched, which often leads to non-parsimonious solutions. Conversely, in many scenarios, when dealing with this type of data, the goal is to analyze the exchange between the same groups of objects that serve as either origins or destinations, because they contain objects with common exchange behavior toward the other groups across occasions, both in magnitude and in direction.

Accordingly, in the lack of proposals for clustering three-way asymmetric proximities, our research objectives are (1) to identify the same groups of objects (for the rows and columns of the data matrices) that are primarily origins or destinations of the exchanges and, together, (2) to measure the extent to which these clusters differ across occasions.

To fill this research gap, we start from the clustering model proposed by Vicari [17], which addresses the first issue and accounts for between-cluster effects when only a single dissimilarity matrix is available (two-way case). In the present paper, we generalize the model [17] to the three-way case with the aim of also accounting for heterogeneity across occasions.

The model is based on the decomposition of each asymmetric matrix into the sum of its symmetric and skew-symmetric components, which are modeled jointly. The asymmetric dissimilarities are assumed to subsume two clustering structures common to all occasions: the first defines a standard partitioning of all objects that fits the symmetric component of the exchanges; the second one, which fits the imbalances, defines an incomplete partition of objects, some of which are allowed to remain unallocated to any cluster. In both clustering structures, objects within the same cluster share the same behavior with respect to exchanges that are directed to objects in different clusters so that “origin” and “destination” clusters are identified. Objects possibly unassigned to any cluster of the incomplete partition represent “nearly” symmetric objects, characterized by small imbalances. As a novel contribution, this paper accounts for heterogeneity across occasions by estimating occasion-specific sets of weights that can capture both the average magnitude and the direction of exchange between clusters. This makes it possible to analyze the role of the common clusters as either origin or destination, which may differ across occasions.

This paper is organized as follows: After an illustrative example to introduce the problem (Section 1.1), the model is formalized in a general framework in Section 2 and an appropriate algorithm is proposed in Section 3. Applications to the artificial data of Section 1.1 and to real mobility data are presented in Sections 4 and 5, respectively, to illustrate the usefulness and effectiveness of the proposal, also in comparison with the Generalized INDCLUS [20]. Finally, Section 6 summarizes the findings and concludes with directions for future developments.

1.1. Illustrative Example: Artificial Data

Before discussing the model in detail, an illustrative example of the heterogeneous three-way asymmetric data we are dealing with is presented.

Let us consider a three-way array of asymmetric dissimilarity data pertaining to pairwise exchanges between $N = 9$ objects measured at $H = 3$ occasions. The three matrices X_h ($h = 1, 2, 3$) in Figure 1, containing the exchanges of each of the three occasions, have been artificially generated from model (4), which will be fully formalized in Section 2, and by assuming three clusters of objects, namely, $C_1 = \{a, b, c, d\}$, $C_2 = \{e, f\}$, and $C_3 = \{g, h, i\}$.

Exchanges are generally asymmetric: for example, the exchange from a to e is different from the exchange from e to a in each occasion. Furthermore, the exchanges from a to all other objects are greater than the exchanges from all other objects to a in occasions 1 and 3, while the opposite is true for occasion 2.

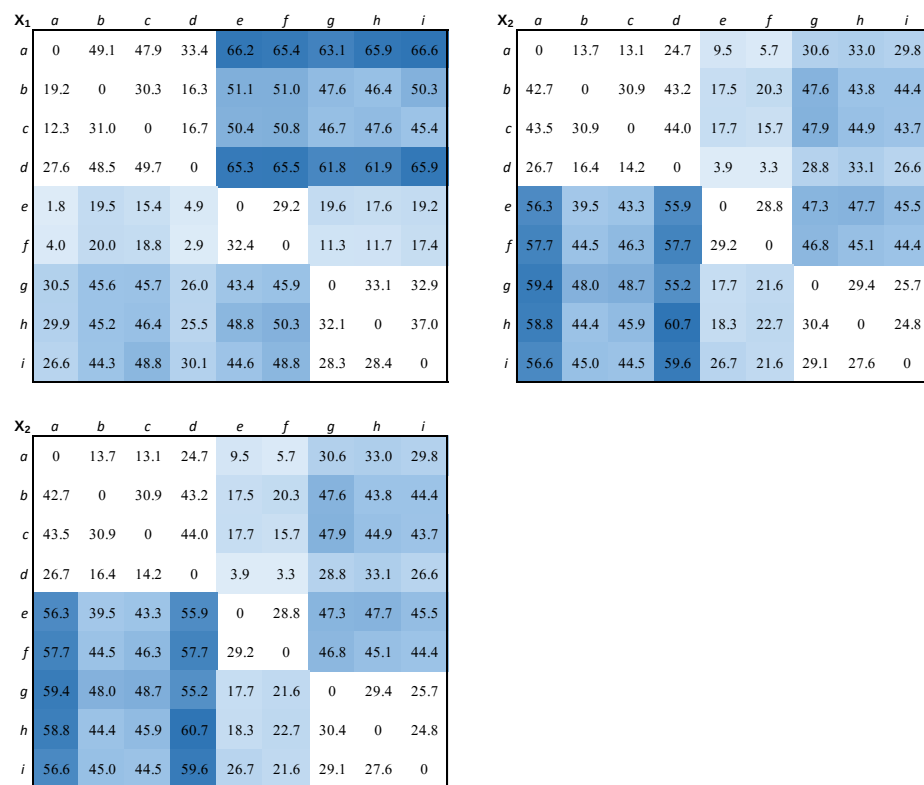


Figure 1. Artificial three-way dissimilarity data. Darker shades of blue represent higher magnitudes.

In addition, using the Gower decomposition [22], each matrix X_h ($h = 1, 2, 3$) can be decomposed into its symmetric and skew-symmetric components S_h and K_h , respectively, as shown in Figure 2.

Let us recall the Gower decomposition [22] of any square matrix X_h ($h = 1, \dots, H$), which can be uniquely decomposed into the sum of a symmetric matrix S_h and a skew-symmetric matrix K_h , both of size $(N \times N)$ and orthogonal to each other (i.e., $trace(S_h K_h) = 0$),

$$X_h = S_h + K_h = \frac{1}{2}(X_h + X_h') + \frac{1}{2}(X_h - X_h'), \quad (h = 1, \dots, H), \quad (1)$$

as elementary linear algebra can easily prove. The entry $s_{ijh} \in S_h$ represents the average amount of the exchange between objects i and j at occasion h ($h = 1, \dots, H$), while the entry $k_{ijh} \in K_h$ represents the imbalance between i and j , i.e., the amount by which k_{ijh} differs from its mean s_{ijh} ($i, j = 1, \dots, N$ and $h = 1, \dots, H$). Thus, each element of the skew-symmetric matrix K_h is, by definition, such that $k_{ijh} = -k_{jih}$, and conveys information about the direction of the exchange.

According to this decomposition, every exchange between a pair of objects in matrix X_h ($h = 1, 2, 3$), for example, $x_{ae1} = 66.2$ from the object a to the object e in the first occasion, can be decomposed into the sum of two entries: (1) the average amount of the exchange between a and e , $s_{ae1} = 34.0$ in S_1 , and (2) the imbalance between the exchanges from a to e and from e to a , $k_{ae1} = 32.2$ in K_1 , i.e., the amount by which the exchange between a and e differs from its mean s_{ae1} .

Based on the Gower decomposition, it is also possible to measure the percentage of asymmetry of each matrix X_h as the squared ratio between the Frobenius norm of its skew-symmetric component K_h and the Frobenius norm of X_h , i.e., $\frac{\|K_h\|^2}{\|X_h\|^2} 100$, which is a measure in $[0, 100]$.

The percentage of asymmetry of each of the matrices in Figure 1 is not negligible (15.5%, 17.8%, and 17.0%), which confirms that the structure of the data is characterized

by the direction and magnitude of the exchanges, which are both relevant and cannot be ignored.

The underlying pattern in the data is evident from Figure 1. In each heatmap, different shades of blue represent different magnitudes of exchange between pairs of objects in different clusters, with darker shades representing higher magnitudes. With a view of searching for origin/destination clusters, we are interested in modeling the exchange between clusters, while the exchange within clusters is not of interest here. Thus, the diagonal blocks are left white for better clarity. The heatmap of each data matrix shows three “blue” off-diagonal blocks, corresponding to the three clusters $C_1 = \{a, b, c, d\}$, $C_2 = \{e, f\}$, and $C_3 = \{g, h, i\}$, each with a common pattern across matrices.

Furthermore, the same clustering structure can be identified in both symmetric and skew-symmetric components S_h and K_h across all occasions (Figure 2).

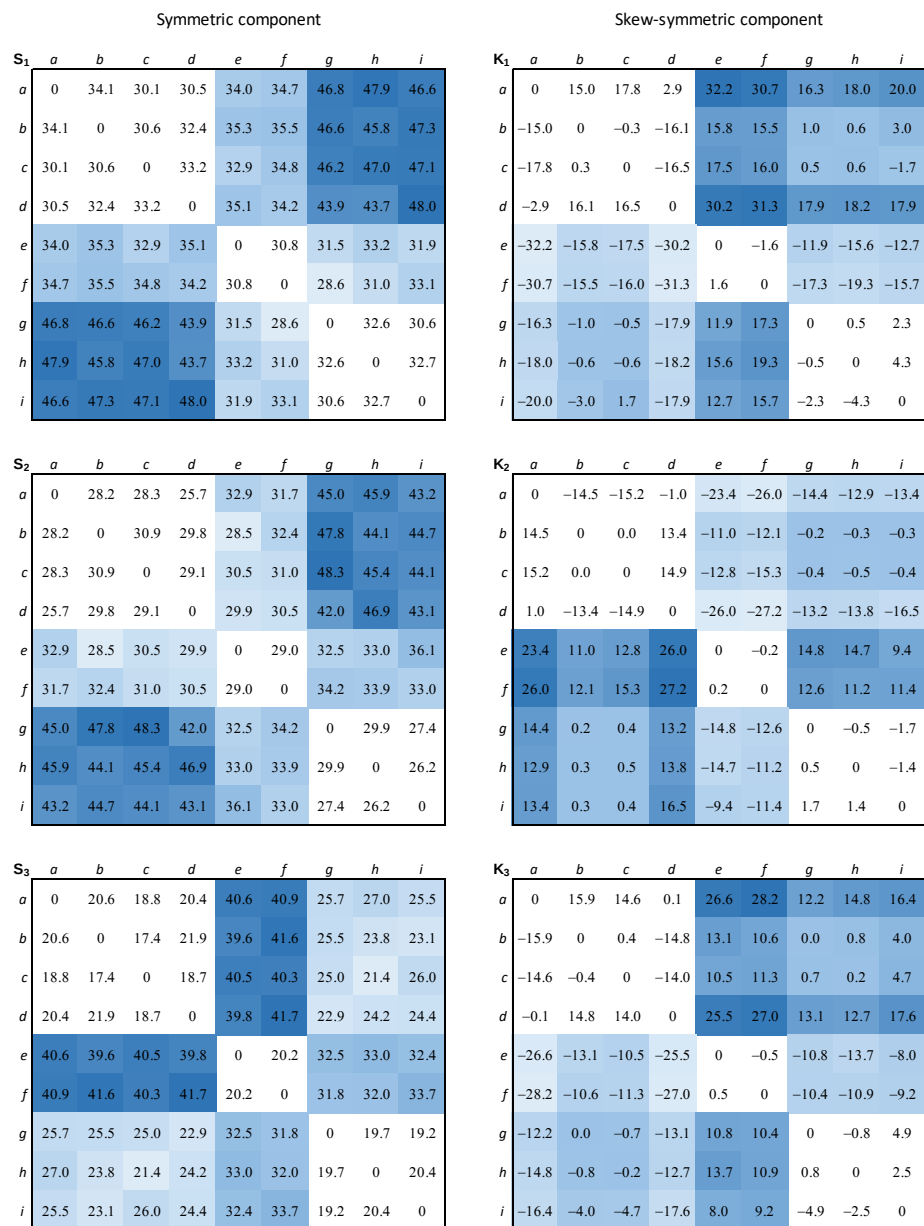


Figure 2. Artificial three-way dissimilarity data: symmetric and skew-symmetric components. Darker shades of blue represent higher magnitudes.

By analyzing the symmetric and skew-symmetric components of each data matrix, it becomes easier to disclose the pattern of the data and determine which groups of objects serve primarily as the origin or destination of the exchanges.

Objects in the same cluster have a common behavior towards objects in different clusters, in terms of outgoing and incoming exchanges, but the magnitude and direction of the exchanges change across occasions. As an example, if we look at the first and third occasions, it is clear that the outgoing exchanges from all objects in C_2 directed to objects in C_1 and C_3 are smaller than the incoming exchanges. This results in negative imbalances for the exchanges from all objects in C_2 to objects in either C_1 or C_3 . Cluster C_2 can therefore be considered as the *origin cluster* for these two occasions. Conversely, in the second occasion, C_2 serves as a *destination cluster*, as its outgoing amounts towards C_1 and C_3 are greater than its incoming ones, resulting in positive imbalances.

By inspecting all the symmetric and skew-symmetric components, we may also note that both the objects b and c (in cluster C_1) present a common behavior across occasions: they have almost symmetrical exchanges with the objects g and h (in cluster C_3) and small imbalances with all other objects belonging to either cluster C_2 or C_3 .

In order to have a better understanding of the underlying patterns and identify the common behaviors of the objects across the different occasions, the model formalized in Section 2 is fitted to the artificial data and the results analyzed in Section 4, together with the results from the Generalized INDCLUS model [20] for comparison.

2. The Model

Let us assume that \mathbf{X}_h ($h = 1, \dots, H$) is a square asymmetric matrix where the element x_{ilh} represents the pairwise dissimilarity between the objects i and l ($i, l = 1, \dots, N$) observed at the occasion h ($h = 1, \dots, H$) and is generally different from x_{lih} .

The model proposed here aims at clustering the N objects by accounting for both the symmetric and the skew-symmetric effects from the decomposition (1) of the observed asymmetries \mathbf{X}_h ($h = 1, \dots, H$).

In particular, the data are assumed to subsume two clustering structures that are common to all occasions: the first one defines a standard partitioning of all objects fitting the average amount of the exchanges; the second one, which fits the imbalances, defines an “incomplete” partitioning of the objects, where some of them are allowed to remain unassigned.

Specifically, the clustering structure consists of two nested partitions into J clusters $\{C_1, \dots, C_j, \dots, C_J\}$ and $\{G_1, \dots, G_j, \dots, G_J\}$, such that

- Every object belongs to one and only one non-empty cluster C_j ($j = 1, \dots, J$);
- If object i belongs to cluster C_j , it can either belong to cluster G_j or remain unassigned to any cluster of the partition $\{G_1, \dots, G_j, \dots, G_J\}$.

The partition $\{C_1, \dots, C_j, \dots, C_J\}$ is referred to as a *complete partition* because every object must be assigned to some cluster C_j , while $\{G_1, \dots, G_j, \dots, G_J\}$ is called an *incomplete partition* because a number of N_0 ($N_0 \leq N$) out of N objects are allowed to remain unassigned to any cluster. Note that the complete and the incomplete partitions are common to all occasions and linked together, the latter being constrained to be nested within the former ($G_j \subseteq C_j$ for $j = 1, \dots, J$).

The *complete partition* is uniquely identified by an $(N \times J)$ binary membership matrix $\mathbf{U} = [u_{ij}]$ ($u_{ij} = \{0, 1\}$ for $i = 1, \dots, N$ and $j = 1, \dots, J$ and $\sum_{j=1}^J u_{ij} = 1$ for $i = 1, \dots, N$), where $u_{ij} = 1$ if object i belongs to cluster C_j , and $u_{ij} = 0$ otherwise.

The *incomplete partition* is identified by an $(N \times J)$ binary membership matrix $\mathbf{V} = [v_{ij}]$ ($v_{ij} = \{0, 1\}$ for $i = 1, \dots, N$ and $j = 1, \dots, J$), where $v_{ij} \leq u_{ij}$ ($i = 1, \dots, N$ and $j = 1, \dots, J$); i.e., any object i can either remain unassigned to any cluster or belong to cluster G_j if it belongs to cluster C_j in the complete partition.

Hereafter, \mathbf{I}_N denotes the identity matrix of size N ; $\mathbf{1}_{AB}$ and $\mathbf{1}_A$ denote the matrix of size $(A \times B)$ of all ones and the column vector with A ones, respectively; $\bar{\mathbf{I}}_N = (\mathbf{1}_{NN} - \mathbf{I}_N)$ is

the $(N \times N)$ matrix of ones except for the zeros on the main diagonal; and $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_H]$ is the $(M \times PH)$ augmented matrix obtained by collecting the H matrices $\mathbf{Y}_1, \dots, \mathbf{Y}_H$ of size $(M \times P)$ next to each other.

Let us consider the Gower decomposition (1) of each matrix \mathbf{X}_h ($h = 1, \dots, H$) into the sum of its symmetric and skew-symmetric components \mathbf{S}_h and \mathbf{K}_h , respectively. Both of these components can be modeled by defining two clustering structures that depend on the matrices \mathbf{U} and \mathbf{V} , respectively, as introduced in Vicari [17] for a two-way asymmetric dissimilarity matrix.

Specifically, the symmetric component \mathbf{S}_h and the skew-symmetric component \mathbf{K}_h for the occasion h ($h = 1, \dots, H$) are modeled by the two clustering structures introduced in Vicari [16,18] and depend on the common complete and incomplete membership matrices \mathbf{U} and \mathbf{V} , respectively, as follows:

$$\mathbf{S}_h = \mathbf{U}\mathbf{R}_h\tilde{\mathbf{U}}' + \tilde{\mathbf{U}}\mathbf{R}_h\mathbf{U}' + \mathbf{E}_{hS}, \quad (h = 1, \dots, H), \quad (2)$$

$$\mathbf{K}_h = \mathbf{V}\mathbf{T}_h\tilde{\mathbf{V}}' - \tilde{\mathbf{V}}\mathbf{T}_h\mathbf{V}' + \mathbf{E}_{hK}, \quad (h = 1, \dots, H), \quad (3)$$

where

- $\tilde{\mathbf{U}} = \mathbf{1}_{NJ} - \mathbf{U}$ and $\tilde{\mathbf{V}} = \mathbf{1}_{NJ} - \mathbf{V}$;
- $\mathbf{R}_h = \text{diag}(\mathbf{r}_h)$ and $\mathbf{T}_h = \text{diag}(\mathbf{t}_h)$ with $\mathbf{r}_h = [r_{1h}, \dots, r_{Jh}]'$ and $\mathbf{t}_h = [t_{1h}, \dots, t_{Jh}]'$ being the occasion-specific weight vectors of size J associated with the clusters of the complete and incomplete partition, respectively;
- The error terms \mathbf{E}_{hS} and \mathbf{E}_{hK} represent the parts of \mathbf{S}_h and \mathbf{K}_h not accounted for by the model, respectively.

For identifiability reasons, any matrix $(\mathbf{V}\mathbf{T}_h)$ is constrained to sum to zero: i.e., $\mathbf{1}'_N(\mathbf{V}\mathbf{T}_h)\mathbf{1}_J = 0$ ($h = 1, \dots, H$).

Models (2) and (3) can be combined and plugged into (1) to specify the model accounting for the asymmetric dissimilarities between clusters at the occasion h , as follows:

$$\begin{aligned} \mathbf{X}_h &= \mathbf{S}_h + \mathbf{K}_h + b_h\tilde{\mathbf{I}}_N + \mathbf{E}_h \\ &= \left[\mathbf{U}\mathbf{R}_h\tilde{\mathbf{U}}' + \tilde{\mathbf{U}}\mathbf{R}_h\mathbf{U}' \right] + \left[\mathbf{V}\mathbf{T}_h\tilde{\mathbf{V}}' - \tilde{\mathbf{V}}\mathbf{T}_h\mathbf{V}' \right] + b_h\tilde{\mathbf{I}}_N + \mathbf{E}_h, \quad (h = 1, \dots, H), \quad (4) \end{aligned}$$

where b_h is an additive constant term and the general error term \mathbf{E}_h represents the part of \mathbf{X}_h not accounted for by the model.

Models (2) and (3) can be expressed in compact notation in terms of $\underline{\mathbf{S}} = [\mathbf{S}_1, \dots, \mathbf{S}_H]$ and $\underline{\mathbf{K}} = [\mathbf{K}_1, \dots, \mathbf{K}_H]$, which denote the $(N \times NH)$ augmented matrices obtained by collecting the H matrices \mathbf{S}_h and \mathbf{K}_h next to each other, respectively, as follows:

$$\underline{\mathbf{S}} = (\mathbf{1}'_H \otimes \mathbf{U})\mathbf{R}(\mathbf{I}_H \otimes \tilde{\mathbf{U}}') + (\mathbf{1}'_H \otimes \tilde{\mathbf{U}})\mathbf{R}(\mathbf{I}_H \otimes \mathbf{U}') + \underline{\mathbf{E}}_S, \quad (5)$$

$$\underline{\mathbf{K}} = (\mathbf{1}'_H \otimes \mathbf{V})\mathbf{T}(\mathbf{I}_H \otimes \tilde{\mathbf{V}}') - (\mathbf{1}'_H \otimes \tilde{\mathbf{V}})\mathbf{T}(\mathbf{I}_H \otimes \mathbf{V}') + \underline{\mathbf{E}}_K, \quad (6)$$

where \otimes denotes the Kronecker product, and \mathbf{R} and \mathbf{T} are the two $(HJ \times HJ)$ diagonal matrices with the HJ -vectors $\mathbf{r} = [r'_1, \dots, r'_h, \dots, r'_H]'$ and $\mathbf{t} = [t'_1, \dots, t'_h, \dots, t'_H]'$ as main diagonals, respectively, $\underline{\mathbf{E}}_S = [\mathbf{E}_{1S}, \dots, \mathbf{E}_{hS}, \dots, \mathbf{E}_{HS}]$ and $\underline{\mathbf{E}}_K = [\mathbf{E}_{1K}, \dots, \mathbf{E}_{hK}, \dots, \mathbf{E}_{HK}]$.

Recall that, given any two matrices $\mathbf{A} = [a_{ij}]$ and \mathbf{B} of sizes $(N \times J)$ and $(M \times P)$, respectively, the Kronecker product between \mathbf{A} and \mathbf{B} is the $(NM \times JP)$ matrix, as follows:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ \vdots & & \vdots \\ a_{N1}\mathbf{B} & \cdots & a_{NJ}\mathbf{B} \end{bmatrix}.$$

Finally, model (4) can be expressed in compact notation in terms of the augmented matrix $\underline{\mathbf{X}} = [\mathbf{X}_1, \dots, \mathbf{X}_H]$ by combining models (5) and (6) as follows:

$$\begin{aligned}
\mathbf{X} &= \mathbf{S} + \mathbf{K} + (\mathbf{b}' \otimes \tilde{\mathbf{I}}_N) + \mathbf{E} \\
&= (\mathbf{1}'_H \otimes \mathbf{U})\mathbf{R}(\mathbf{I}_H \otimes \tilde{\mathbf{U}}') + (\mathbf{1}'_H \otimes \tilde{\mathbf{U}})\mathbf{R}(\mathbf{I}_H \otimes \mathbf{U}') + \\
&\quad (\mathbf{1}'_H \otimes \mathbf{V})\mathbf{T}(\mathbf{I}_H \otimes \tilde{\mathbf{V}}') - (\mathbf{1}'_H \otimes \tilde{\mathbf{V}})\mathbf{T}(\mathbf{I}_H \otimes \mathbf{V}') + \\
&\quad (\mathbf{b}' \otimes \tilde{\mathbf{I}}_N) + \mathbf{E},
\end{aligned} \tag{7}$$

where $\mathbf{b} = [b_1, \dots, b_h, \dots, b_H]'$ and $\mathbf{E} = [\mathbf{E}_1, \dots, \mathbf{E}_h, \dots, \mathbf{E}_H]$.

It is important to note that, in the model, in addition to a common clustering structure, occasion-specific weight vectors \mathbf{r}_h and \mathbf{t}_h are assumed to account for the heterogeneity of the occasions. These weights allow for measuring the extent to which exchanges vary across occasions, providing quantifications of the exchanges between clusters at the occasion h in terms of magnitude and direction.

3. The Algorithm

In model (4), the complete and the incomplete membership matrices \mathbf{U} and \mathbf{V} , the weight vectors \mathbf{r}_h and \mathbf{t}_h , and the constants b_h ($h = 1, \dots, H$) can be estimated by solving the following least-squares fitting problem:

$$\begin{aligned}
\min F(\mathbf{U}, \mathbf{V}, \mathbf{r}_h, \mathbf{t}_h, b_h) &= \\
\frac{\sum_{h=1}^H \left\| \mathbf{X}_h - (\mathbf{U}\mathbf{R}_h\tilde{\mathbf{U}}' + \tilde{\mathbf{U}}\mathbf{R}_h\mathbf{U}') - (\mathbf{V}\mathbf{T}_h\tilde{\mathbf{V}}' - \tilde{\mathbf{V}}\mathbf{T}_h\mathbf{V}') - b_h\tilde{\mathbf{I}}_N \right\|^2}{\sum_{h=1}^H \left\| \mathbf{X}_h \right\|^2}
\end{aligned} \tag{8}$$

subject to

$$u_{ij} = \{0, 1\} \quad (i = 1, \dots, N; j = 1, \dots, J) \text{ and } \sum_{j=1}^J u_{ij} = 1 \quad (i = 1, \dots, N), \tag{9}$$

$$v_{ij} = \{0, 1\} \quad (i = 1, \dots, N; j = 1, \dots, J) \text{ and } v_{ij} \leq u_{ij} \quad (i = 1, \dots, N), \tag{10}$$

$$\mathbf{1}'_N(\mathbf{V}\mathbf{T}_h)\mathbf{1}_J = 0 \quad (h = 1, \dots, H). \tag{11}$$

Problem (8), subject to constraints (9)–(11), can be reformulated in compact form in terms of model (7) as follows:

$$\begin{aligned}
\min F(\mathbf{U}, \mathbf{V}, \mathbf{R}, \mathbf{T}, \mathbf{b}) &= \\
\frac{1}{\left\| \mathbf{X} \right\|^2} &\left\| \mathbf{X} - (\mathbf{1}'_H \otimes \mathbf{U})\mathbf{R}(\mathbf{I}_H \otimes \tilde{\mathbf{U}}') - (\mathbf{1}'_H \otimes \tilde{\mathbf{U}})\mathbf{R}(\mathbf{I}_H \otimes \mathbf{U}') - \right. \\
&\quad \left. (\mathbf{1}'_H \otimes \mathbf{V})\mathbf{T}(\mathbf{I}_H \otimes \tilde{\mathbf{V}}') + (\mathbf{1}'_H \otimes \tilde{\mathbf{V}})\mathbf{T}(\mathbf{I}_H \otimes \mathbf{V}') - (\mathbf{b}' \otimes \tilde{\mathbf{I}}_N) \right\|^2 = \\
\frac{\left\| \mathbf{S} - (\mathbf{1}'_H \otimes \mathbf{U})\mathbf{R}(\mathbf{I}_H \otimes \tilde{\mathbf{U}}') - (\mathbf{1}'_H \otimes \tilde{\mathbf{U}})\mathbf{R}(\mathbf{I}_H \otimes \mathbf{U}') - (\mathbf{b}' \otimes \tilde{\mathbf{I}}_N) \right\|^2}{\left\| \mathbf{X} \right\|^2} &+ \\
\frac{\left\| \mathbf{K} - (\mathbf{1}'_H \otimes \mathbf{V})\mathbf{T}(\mathbf{I}_H \otimes \tilde{\mathbf{V}}') + (\mathbf{1}'_H \otimes \tilde{\mathbf{V}})\mathbf{T}(\mathbf{I}_H \otimes \mathbf{V}') \right\|^2}{\left\| \mathbf{X} \right\|^2} &= \\
F^S(\mathbf{U}, \mathbf{R}, \mathbf{b}) + F^K(\mathbf{V}, \mathbf{T}), &
\end{aligned} \tag{12}$$

where the equivalence is due to the orthogonality of \mathbf{S} and \mathbf{K} .

The equivalent constrained optimization problems (8) and (12) can be solved by using an Alternating Least-Squares (ALS) algorithm, which alternates the estimation of a set of parameters while maintaining all the others fixed as detailed below.

After an initialization step in which all parameters satisfying the constraints are chosen, the algorithm alternates between two main steps.

In the first step, in order to ensure that the relative loss function (12) is non-increasing, the two membership matrices \mathbf{U} and \mathbf{V} are jointly updated row by row in N substeps by solving assignment problems for the different rows of \mathbf{U} and \mathbf{V} satisfying the constraints (9) and (10). Specifically, for a given row i , setting $u_{ij} = 1$ for $j = 1, \dots, J$ implies that either $v_{ij} = u_{ij}$ or $v_{ij} = 0$; i.e., object i in matrix \mathbf{V} can either be assigned to the same cluster j as in matrix \mathbf{U} or remain unassigned. For the row i and given the remaining rows of \mathbf{U} and \mathbf{V} , all possible $2J$ assignments of the object i are considered: either to the corresponding clusters C_j and G_j ($u_{ij} = 1$ and $v_{ij} = 1$) or to the cluster C_j only ($u_{ij} = 1$ and $v_{ij} = 0$). For each of them, the weight vectors \mathbf{r}_h and \mathbf{t}_h ($h = 1, \dots, H$) are also estimated as optimal solutions of constrained regression problems. Finally, by evaluating the relative loss function (12) for all possible potential assignments and selecting the one corresponding to the minimum loss value, the assignment of the object i in \mathbf{U} and \mathbf{V} is chosen. Note that the relative loss function cannot increase at each substep, since the whole space of feasible solutions for both \mathbf{U} and \mathbf{V} is explored for each object.

In the same step, the weight vectors \mathbf{r}_h and \mathbf{t}_h are estimated as solutions of constrained regression problems for each possible choice of the different rows of \mathbf{U} and \mathbf{V} , respectively.

In the second step, the constant \mathbf{b} is then estimated by successive residualizations of the three-way data matrix.

The two main steps are alternated and iterated until convergence. The relative loss function (12) does not increase at each step, and the algorithm stops when the loss decreases less than a fixed arbitrary positive and small threshold.

In order to increase the chance of finding the global minimum, the best solution over different random starting parameters is retained.

Moreover, in order to estimate the matrices \mathbf{R} and \mathbf{T} , model (7) is reformulated as a regression problem with respect to the unknown vectors \mathbf{r} and \mathbf{t} , as follows:

$$\begin{aligned}\underline{\mathbf{x}} &= \underline{\mathbf{s}} + \underline{\mathbf{k}} + (\mathbf{b} \otimes \tilde{\mathbf{i}}) + \underline{\mathbf{e}} \\ &= \mathbf{Q}_U \mathbf{r} + \mathbf{Q}_V \mathbf{t} + (\mathbf{b} \otimes \tilde{\mathbf{i}}) + \underline{\mathbf{e}},\end{aligned}\quad (13)$$

where

- $\underline{\mathbf{x}}$ is the column vector of size HN^2 of the vectorized matrix \mathbf{X} , i.e., $\underline{\mathbf{x}} = \text{vec}(\mathbf{X}) = [x_{111}, \dots, x_{N11}, \dots, x_{11h}, \dots, x_{N1h}, \dots, x_{1NH}, \dots, x_{NNH}]'$;
- $\underline{\mathbf{s}} = \text{vec}(\mathbf{S})$ and $\underline{\mathbf{k}} = \text{vec}(\mathbf{K})$ are the column vectors of size HN^2 of the vectorized matrices \mathbf{S} and \mathbf{K} , respectively;
- $\mathbf{Q}_U = [(\mathbf{I}_H \otimes \tilde{\mathbf{U}}) | \otimes | (\mathbf{1}'_H \otimes \mathbf{U})] + [(\mathbf{I}_H \otimes \mathbf{U}) | \otimes | (\mathbf{1}'_H \otimes \tilde{\mathbf{U}})]$ is a matrix of size $(HN^2 \times HJ)$, where $| \otimes |$ denotes the Khatri–Rao product [23,24];
- $\mathbf{Q}_V = [(\mathbf{I}_H \otimes \tilde{\mathbf{V}}) | \otimes | (\mathbf{1}'_H \otimes \mathbf{V})] - [(\mathbf{I}_H \otimes \mathbf{V}) | \otimes | (\mathbf{1}'_H \otimes \tilde{\mathbf{V}})]$ is a matrix of size $(HN^2 \times HJ)$;
- $\tilde{\mathbf{i}}$ is the column vector of size N^2 of the vectorized matrix $\tilde{\mathbf{I}}_N$;
- $\underline{\mathbf{e}} = \text{vec}(\mathbf{E})$ is the column vector of size HN^2 of the error term.

Recall that, given any the two matrices \mathbf{A} and \mathbf{B} with the same number J of columns, the Khatri–Rao product of \mathbf{A} and \mathbf{B} is the column-wise Kronecker product, i.e., $\mathbf{A} | \otimes | \mathbf{B} = (\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_j \otimes \mathbf{b}_j, \dots, \mathbf{a}_J \otimes \mathbf{b}_J)$, where \mathbf{a}_j and \mathbf{b}_j are the j -th ($j = 1, \dots, J$) column of \mathbf{A} and \mathbf{B} , respectively.

Therefore, by taking into account (13), the relative loss function (12) becomes as follows:

$$\begin{aligned}
 F(\mathbf{U}, \mathbf{V}, \mathbf{r}, \mathbf{t}, \mathbf{b}) &= F^S(\mathbf{U}, \mathbf{r}, \mathbf{b}) + F^K(\mathbf{V}, \mathbf{t}) \\
 &= \frac{\|\underline{\mathbf{s}} - \mathbf{Q}_U \mathbf{r} - (\mathbf{b} \otimes \tilde{\mathbf{i}})\|^2}{\|\underline{\mathbf{x}}\|^2} + \frac{\|\underline{\mathbf{k}} - \mathbf{Q}_V \mathbf{t}\|^2}{\|\underline{\mathbf{x}}\|^2}.
 \end{aligned} \tag{14}$$

A detailed description of the steps of the algorithm, implemented in MATLAB R2023a, is given below.

Initialization step.

Initial estimates of the parameters $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$, $\hat{\mathbf{r}}$, $\hat{\mathbf{t}}$, and $\hat{\mathbf{b}}$ are chosen randomly or in a rational way, but they are required to satisfy the set of constraints (9)–(11).

Step 1. Updating the membership matrices \mathbf{U} and \mathbf{V} and weight-vectors \mathbf{r} and \mathbf{t} . (see Algorithm 1)

Given the current estimates of $\hat{\mathbf{b}}$, \mathbf{U} , and \mathbf{V} , the weight vectors \mathbf{r} , and \mathbf{t} are estimated by minimizing (14) subject to constraints (9)–(11).

The loss function (14) is minimized sequentially for the different rows of \mathbf{U} and \mathbf{V} by solving N assignment problems. Finally, the column sums of the estimated $\hat{\mathbf{U}}$ are checked to avoid empty clusters.

Furthermore, the weight vectors \mathbf{r} and \mathbf{t} are estimated in two nested substeps as follows. Given the current $\hat{\mathbf{U}}$ and $\hat{\mathbf{b}}$, for every possible binary choice for the different rows of \mathbf{U} , the vector \mathbf{r} is estimated by solving the following regression problem:

$$F^S(\mathbf{r}; \hat{\mathbf{U}}, \hat{\mathbf{b}}) = \frac{\|\underline{\mathbf{s}} - \mathbf{Q}_{\hat{\mathbf{U}}} \mathbf{r} - (\hat{\mathbf{b}} \otimes \tilde{\mathbf{i}})\|^2}{\|\underline{\mathbf{x}}\|^2}. \tag{15}$$

Similarly, given the current $\hat{\mathbf{V}}$, for every admissible choice for the different rows of \mathbf{V} , the weight vector \mathbf{t} is obtained as the solution of the following constrained regression problem:

$$F^K(\mathbf{t}; \hat{\mathbf{V}}) = \frac{\|\underline{\mathbf{k}} - \mathbf{Q}_{\hat{\mathbf{V}}} \mathbf{t}\|^2}{\|\underline{\mathbf{x}}\|^2}, \tag{16}$$

subject to constraints (11).

The pseudocode for updating the i -th row of \mathbf{U} and \mathbf{V} , holding all other rows constant, and for updating \mathbf{r} and \mathbf{t} is as follows.

In the following, let $\mathbf{0}$ be the J -column vector of zeros; $\mathbf{p}^{(j)}$ be the J -dimensional vector with all entries equal to 0 except for the j -th one, which is 1; and \mathbf{u}_i and \mathbf{v}_i denote the vectors corresponding to the i -th row of \mathbf{U} and \mathbf{V} , respectively.

Algorithm 1 Step 1.

```

begin
  for  $i := 1$  to  $N$  do
    for  $j := 1$  to  $J$  do
       $\mathbf{u}_i^{(j)} = \mathbf{p}^{(j)}$ ;
       $\mathbf{U}^{(j)} = [\mathbf{u}_1, \dots, \mathbf{u}_i^{(j)}, \dots, \mathbf{u}_N]'$ ;
       $\mathbf{r}^{(j)} = (\mathbf{Q}'_{\mathbf{U}^{(j)}} \mathbf{Q}_{\mathbf{U}^{(j)}})^{-1} \mathbf{Q}'_{\mathbf{U}^{(j)}} (\mathbf{s} - (\hat{\mathbf{b}} \otimes \tilde{\mathbf{i}}))$ ; (17)
      comment: solution of the regression problem (15) corresponding to the possible
              assignment of object  $i$  to cluster  $j$  of the complete partition ( $\mathbf{U}$ );
      for  $w := 1$  to 2 do
        if  $w = 1$  then  $\mathbf{v}_i^{(j,w)} = \mathbf{u}_i^{(j)}$ , else if  $w = 2$  then  $\mathbf{v}_i^{(j,w)} = \mathbf{0}$ ; end;
         $\mathbf{V}^{(j,w)} = [\mathbf{v}_1, \dots, \mathbf{v}_i^{(j,w)}, \dots, \mathbf{v}_N]'$ ;
         $N_V = (\mathbf{1}'_N \mathbf{V}^{(j,w)} \mathbf{1}_J)$ ;
        comment: number of assigned objects in the incomplete partition;
         $\tilde{\mathbf{t}}^{(j,w)} = (\mathbf{Q}'_{\mathbf{V}^{(j,w)}} \mathbf{Q}_{\mathbf{V}^{(j,w)}})^{-1} \mathbf{Q}'_{\mathbf{V}^{(j,w)}} \mathbf{k}$ ; (18)
        comment: solution of the regression problem (16) corresponding to the possible
              assignment of object  $i$  to the cluster  $j$  of the incomplete partition ( $\mathbf{V}$ );
        for  $h := 1$  to  $H$  do
           $\mathbf{t}_h^{(j,w)} = \mathbf{V}^{(j,w)+} (\mathbf{V}^{(j,w)} \tilde{\mathbf{t}}^{(j,w)} - \frac{\mathbf{1}'_N \mathbf{V}^{(j,w)} \tilde{\mathbf{t}}^{(j,w)}}{N_V} \mathbf{V}^{(j,w)} \mathbf{1}_J)$ ; (19)
          comment: constraint (11) is imposed;  $\mathbf{V}^{(j,w)+}$  is the Moore–Penrose
                  inverse of  $\mathbf{V}^{(j,w)}$ ;
        end
         $\mathbf{t}^{(j,w)} = [\mathbf{t}_1^{(j,w)'}, \dots, \mathbf{t}_h^{(j,w)'}, \dots, \mathbf{t}_H^{(j,w)'}]'$ ;
         $f^{(j,w)}(\mathbf{u}_i^{(j)}, \mathbf{r}^{(j)}, \mathbf{v}_i^{(j,w)}, \mathbf{t}^{(j,w)}) = F^S(\mathbf{U}^{(j)}, \mathbf{r}^{(j)}; \hat{\mathbf{b}}) + F^K(\mathbf{V}^{(j,w)}, \mathbf{t}^{(j,w)})$ ;
      end
    end
  end
   $(l, g) = \arg \min_{1 \leq j \leq J} (\arg \min_{w \in \{1,2\}} (f^{(j,w)}))$ ;
   $\hat{\mathbf{u}}_i = \mathbf{p}^{(l)}$ ;
   $\hat{\mathbf{r}} = \mathbf{r}^{(l)}$ ;
  if  $g = 1$  then  $\hat{\mathbf{v}}_i = \hat{\mathbf{u}}_i$ ,  $\hat{\mathbf{t}} = \mathbf{t}^{(l,1)}$ , else if  $g = 2$  then  $\hat{\mathbf{v}}_i = \mathbf{0}$ ,  $\hat{\mathbf{t}} = \mathbf{t}^{(l,2)}$ ; end;
end

```

Step 2. Updating constant \mathbf{b} .

Given the current estimates of $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$, $\hat{\mathbf{r}}$, and $\hat{\mathbf{t}}$, the estimate of \mathbf{b} is given by the following:

$$\hat{\mathbf{b}} = [(\mathbf{I}_H \otimes \tilde{\mathbf{t}}')(\mathbf{I}_H \otimes \tilde{\mathbf{i}})]^{-1} (\mathbf{I}_H \otimes \tilde{\mathbf{t}}') (\mathbf{s} - \mathbf{Q}_{\hat{\mathbf{U}}} \hat{\mathbf{r}}), \quad (20)$$

where the inverse always exists, since $[(\mathbf{I}_H \otimes \tilde{\mathbf{t}}')(\mathbf{I}_H \otimes \tilde{\mathbf{i}})]$ is a full-rank diagonal matrix of size H with diagonal elements all equal to $N(N-1)$.

Stopping rule.

The relative loss function value is computed for the current values of $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$, $\hat{\mathbf{r}}$, $\hat{\mathbf{t}}$, and $\hat{\mathbf{b}}$ and since $F(\hat{\mathbf{U}}, \hat{\mathbf{V}}, \hat{\mathbf{r}}, \hat{\mathbf{t}}, \hat{\mathbf{b}})$ is bounded from below, it converges to a point that is expected to be at least a local minimum. When the loss function (14) has not decreased considerably with respect to a tolerance value, the process is assumed to be converged. Otherwise, steps 1 and 2 are repeated in turn.

3.1. Meaning of the Parameter Estimates

In order to evaluate the meaning of the estimated weights $\hat{\mathbf{r}}$ of the *complete partition*, let us consider the membership matrix $\hat{\mathbf{U}} = [\hat{u}_{ij}]_{(i=1,\dots,N; j=1,\dots,J)}$ that uniquely identifies the clusters of the complete partition $\{C_1, \dots, C_j, \dots, C_J\}$.

From Equation (17), the estimated weight \hat{r}_{jh} of the cluster C_j in the h -th occasion is the average amount of the exchanges between objects in cluster C_j and objects in clusters different from C_j , corrected for the mean of the average amounts between all clusters different from C_j . Hence, large (small) values of \hat{r}_{jh} indicate clusters characterized by large (small) amounts of exchanges on average.

Similarly, let us consider the *incomplete partition* $\{G_1, \dots, G_j, \dots, G_J\}$ identified by the matrix $\hat{\mathbf{V}} = [\hat{v}_{ij}]_{(i=1,\dots,N; j=1,\dots,J)}$. Then, from (18) imposing the constraint (11), the weight \hat{t}_{jh} of the cluster G_j in the occasion h is as follows:

$$\hat{t}_{jh} = \frac{1}{N \times N_{G_j}} \sum_{i=1}^N \sum_{l=1}^N k_{ilh} \hat{v}_{ij} \tilde{\delta}_{lj}, \quad (j = 1, \dots, J; h = 1, \dots, H),$$

where N_{G_j} is the number of objects assigned to the cluster G_j .

Note that \hat{t}_{jh} is the average imbalance from all objects in the cluster G_j towards all objects in clusters other than G_j , at the occasion h , corrected for the average imbalance originating from all clusters different from G_j . Therefore, a positive (negative) weight \hat{t}_{jh} qualifies the cluster G_j as a “destination” (“origin”) cluster of the exchanges at the occasion h , and the objects belonging to such a cluster have a similar pattern in terms of exchanges directed towards the other clusters.

Given the occasion h and due to (3), all objects belonging to the same cluster G_j have the same weight \hat{t}_{jh} , and the weighted sum of them over all clusters is zero due to constraint (11). Constraint (11) is necessary to guarantee the identifiability of the model and the uniqueness of the solution: this is because the weights are defined by difference. In a general and widely applicable framework, such a choice implies that the model defines a closed exchange system in the sense that the total imbalance between clusters within each occasion is zero. Note that, in specific contexts, a value other than zero could generally be chosen to handle appropriate assumptions. Only (19) in the algorithm should be modified accordingly.

Furthermore, the objects that remain unassigned to any cluster of the incomplete partition are actually objects that generate (almost) zero mutual imbalances and, equivalently, almost symmetric exchanges in all occasions.

The constant term \hat{b}_h is assumed to be added to all dissimilarities in the h -th occasion and, thus, only affects the average amounts of exchanges between objects (symmetric component S_h). Therefore, the additive constant \hat{b}_h in (20) represents the baseline average dissimilarity, independent of any clustering and direction, and plays the same role as the intercept of a linear regression model.

Finally, it is worth noting that, for each occasion h , model (7) accounts for the between-cluster effects, while the exchanges within clusters are fitted only by the constant term \hat{b}_h , which actually also represents the average exchange between objects within clusters.

4. Illustrative Example: Results

In order to detect the underlying clustering structure common to all occasions and identify “origin” and “destination” clusters, the model proposed here was fitted to the artificial data of Figure 1, together with the Generalized INDCLUS model [20] for comparison.

Let us briefly recall the generalization of the INDCLUS model [20] to fit three-way asymmetric similarity data \mathbf{Y}_h ($h = 1, \dots, H$). The model assumes that there exist two sets of J overlapping clusters (i.e., two coverings) of the same set of N objects in row and column, and that they are common to all occasions, namely, $\mathbf{P} = [p_{ij}]$ and $\mathbf{Q} = [q_{ij}]$, where p_{ij} and q_{ij} assume values in $\{0, 1\}$ (for $i = 1, \dots, N$ and $j = 1, \dots, J$); i.e., each object is allowed to belong to more than one cluster or none at all. In addition, a single set of weights

is assumed for both set clusters, but different for each occasion. The model can be written as follows:

$$Y_h = PW_hQ' + c_h\mathbf{1}_{NN} + E_h^*, \quad (h = 1, \dots, H), \tag{21}$$

where W_h is the non-negative diagonal weight matrix of order J for the occasion h , c_h is a real-valued additive constant for the occasion h , and E_h^* is the error term.

Since model (21) fits similarities, the original artificial dissimilarities in $[0, 100]$ of Figure 1 were simply converted into similarities by taking $Y_h = 100 - X_h$ ($h = 1, 2, 3$).

The best solution in $J = 3$ clusters was retained over 100 random starts of the algorithm.

None of the two best coverings in three overlapping clusters obtained by the Generalized INDCLUS model (relative loss function equal to 0.0228) identify the true generated partition of the objects. The common coverings P and Q for the sets of objects in the rows and in the columns, respectively, consist of three row clusters, $C_1^{Ir} = \{a, b, c, d\}$, $C_2^{Ir} = \{e, f\}$, and $C_3^{Ir} = \{b, c, g, h, i\}$, and three column clusters, $C_1^{Ic} = \{b, c, e, f\}$, $C_2^{Ic} = \{a, b, c, d, e, f, g, h, i\}$, and $C_3^{Ic} = \{a, b, c, d, g, h, i\}$. The first covering identifies row clusters of objects with similar outgoing exchanges to the column clusters, which contain groups of objects with similar incoming exchanges.

The estimated weights of the three clusters of both coverings in each occasion and the constants are reported in Table 1. The constants represent the average exchange between all objects within each occasion, regardless of the direction. Thus, the second occasion has the highest average level of exchange ($c_2 = 59.7$) and the first occasion the lowest ($c_2 = 49.6$).

Table 1 displays the weights of the outgoing row clusters and incoming column clusters, where a large positive weight for any two corresponding clusters C_j^{Ir} and C_j^{Ic} ($j = 1, 2, 3$) qualifies C_j^{Ir} as the origin of the exchanges directed to the destination C_j^{Ic} . Therefore, in the second occasion, $C_1^{Ir} = \{a, b, c, d\}$ is an origin cluster to $C_1^{Ic} = \{b, c, e, f\}$. In addition, it can be noted that, in occasion 2, Generalized INDCLUS fails to identify the true behavior of objects $\{g, h, i\}$ in C_3 as destination from $C_1 = \{a, b, c, d\}$ and origin to $C_2 = \{e, f\}$ (Figure 1). Instead, objects $\{g, h, i\}$ are estimated here to have mutual exchanges with all other objects due to their membership in several column clusters.

Table 1. Artificial data: occasion-specific weights and constants from Generalized INDCLUS.

	Cluster			Constant
	Row	C_1^{Ir}	C_2^{Ir}	
Column	C_1^{Ic}	C_2^{Ic}	C_3^{Ic}	
	w_{1h}	w_{2h}	w_{3h}	c_h
Occasion 1	0	38.0	18.5	49.6
Occasion 2	25.1	0	0	59.7
Occasion 3	0	22.8	26.8	55.7

Moreover, in the first and third occasions, the clusters $C_2^{Ir} = \{e, f\}$ and $C_3^{Ir} = \{b, c, g, h, i\}$ are origins toward C_2^{Ic} , which contains all objects, and $C_3^{Ic} = \{a, b, c, d, g, h, i\}$, respectively.

We can observe that, in the special case when the clusters from Generalized INDCLUS form a partition (they do not overlap), model (21) reduces to estimating only the exchanges from any C_j^{Ir} row cluster to its corresponding C_j^{Ic} column cluster. Conversely, in the presence of overlapping clusters, objects common to different clusters also contribute to the estimation of the exchange between different clusters. That is why, in Table 1, due to the large cluster overlap, it is very cumbersome to extract the directions of the exchange from any row cluster C_j^{Ir} to any other column cluster C_j^{Ic} . It becomes necessary to look at the whole full-size ($N \times N$) estimated matrices and analyze the estimated exchanges between pairs of objects.

For the sake of clarity, we observe that, in this well-structured artificial situation, the lowest weights are exactly zero, which is generally not the case.

Model (4) proposed here was fitted to the artificial data of Figure 1, and the best solution in $J = 3$ clusters was retained over 100 random starts of the algorithm.

The best resulting partition (relative loss equal to 0.0026) correctly identifies the *complete partition* from the symmetric components consisting of clusters $C_1 = \{a, b, c, d\}$, $C_2 = \{e, f\}$, $C_3 = \{g, h, i\}$, and the *incomplete partition* formed by clusters $G_1 = \{a, d\}$, $G_2 = \{e, f\}$, $G_3 = \{i\}$, which are nested in the corresponding clusters of the former.

The objects $b, c, g,$ and h remain unassigned to any cluster in the incomplete partition from the skew-symmetric components due to their almost zero mutual imbalances and almost symmetric exchanges at each occasion.

Table 2 reports the estimates of the weight vectors \hat{r} and \hat{t} and the constant \hat{b} . In addition, for each occasion h , the estimated between-cluster components, both the symmetric (\hat{S}_h) from the complete partition and the skew-symmetric (\hat{K}_h) from the incomplete partition, are shown in Figure 3.

Table 2. Artificial data: occasion-specific weights from the proposed model.

	Complete Partition			Incomplete Partition			Constant
	C_1	C_2	C_3	G_1	G_2	G_3	
	\hat{r}_{1h}	\hat{r}_{2h}	\hat{r}_{3h}	\hat{t}_{1h}	\hat{t}_{2h}	\hat{t}_{3h}	
Occasion 1	8.8	−6.0	5.8	16.2	−15.3	−1.9	31.8
Occasion 2	6.9	−4.4	9.7	−13.2	12.3	1.8	28.4
Occasion 3	6.4	14.5	−1.6	14.1	−12.5	−3.4	19.7

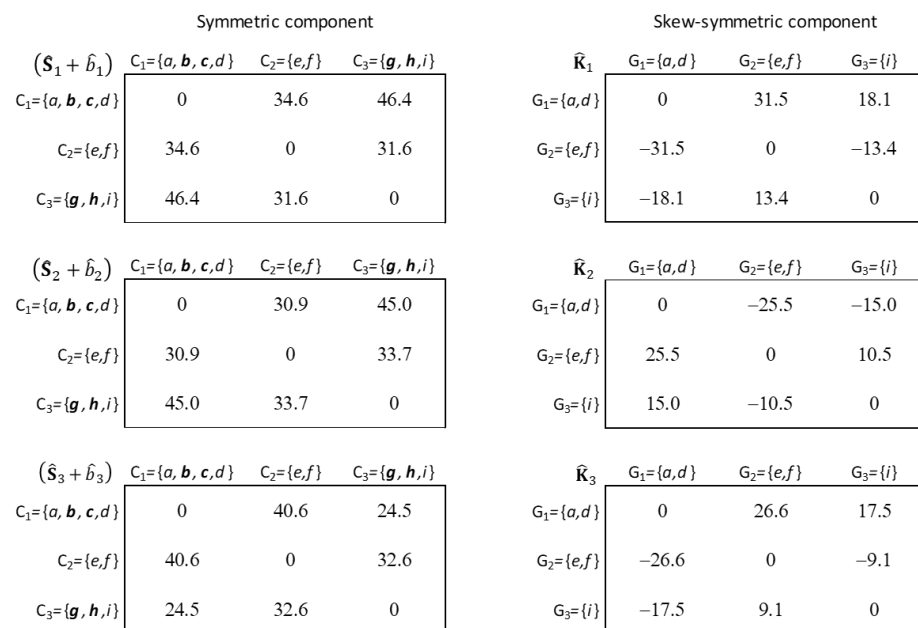


Figure 3. Artificial three-way dissimilarity data: estimated symmetric and skew-symmetric components from the proposed model.

The estimated constants \hat{b}_h (Table 2) represent the baseline average exchange level at each occasion, regardless of any clustering, and take on different values in the occasions: the highest ($\hat{b}_1 = 31.8$) in the first, while the third has the lowest ($\hat{b}_3 = 19.7$).

Given the occasion h , the weight \hat{r}_{jh} of the cluster C_j ($j = 1, 2, 3$) represents the average exchange between the cluster C_j and every other cluster, in addition to the additive constant. Thus, from the occasion-specific weights \hat{r}_h ($h = 1, 2, 3$) in Table 2, we observe that, in occasion 1, C_1 subsumes the highest average exchange with the other two clusters ($\hat{r}_{11} = 8.8$), while in occasion 2, the highest average exchange concerns C_3 ($\hat{r}_{32} = 9.7$). In

contrast, in occasion 3, cluster C_3 has the lowest weight ($\hat{r}_{33} = -1.6$), while C_2 has the highest ($\hat{r}_{23} = 14.5$). This is consistent with what can be seen from the data in Figure 2.

In addition, the different role of each cluster can be identified from the occasion-specific weights \hat{t}_h ($h = 1, 2, 3$) of the clusters of the incomplete partition, which allow for qualifying origins and destinations. Since each \hat{t}_{jh} represents the weighted average imbalance at the occasion h from the cluster G_j to all other clusters, in Table 2, the negative weight of G_1 in the second occasion ($\hat{t}_{12} = -13.2$) qualifies G_1 as an origin cluster, while it acts as a destination cluster in occasions 1 and 3, where its corresponding weights are both positive ($\hat{t}_{11} = 16.2$ and $\hat{t}_{13} = 14.1$, respectively). The reverse is true for both clusters G_2 and G_3 .

As an example, we can see from Figure 3 that the estimated symmetric part of the exchange between the clusters C_1 and C_2 in occasion 1 is 34.6, which is the sum of the estimated weights of C_1 and C_2 ($\hat{r}_{11} = 8.8$ and $\hat{r}_{21} = -6.0$, respectively) plus the constant value ($\hat{b}_1 = 31.8$), i.e., $(\hat{s}_{121} + \hat{b}_1) = 8.8 - 6.0 + 31.8 = 34.6$.

Moreover, the corresponding estimated imbalance between G_1 and G_2 yields $\hat{k}_{121} = 31.5$, which is the difference of the estimated weights of the clusters G_1 and G_2 ($\hat{t}_{11} = 16.2$ and $\hat{t}_{21} = -15.3$, respectively), i.e., $\hat{k}_{121} = 16.2 + 15.3 = 31.5$.

Finally, the estimated between-cluster dissimilarity between C_1 and C_2 in the first occasion is the sum of $(\hat{s}_{121} + \hat{b}_1)$ and \hat{k}_{121} , hence $\hat{x}_{121} = 66.1$.

Remark 1. From Figure 3, we can easily derive the summary of what emerges from the detailed results, since the full-size exchange matrices between objects (Figure 2) are synthesized into reduced-size exchange matrices between clusters with a very limited loss of information. This makes it possible to show the role played by each cluster of objects as an origin or destination of exchange, exactly reflecting the original data in Figure 1.

5. Application to Student Mobility Data

The data analyzed in this application have been taken from the OECD (Organization for Economic Co-operation and Development) Education Statistics Database, which collects data annually on international student exchanges in tertiary education.

The three-way data array consists of $H = 5$ asymmetric matrices of exchanges of international student mobility observed in $N = 20$ countries (objects) from 2016 to 2020 ($H = 5$ occasions). The data report the number of international tertiary students who received their prior education in the origin country (the rows of each year matrix) enrolled in the host country (the columns of each year matrix). The countries are the twenty founding members of the OECD: Austria (AT), Belgium (BE), Canada (CA), Denmark (DK), France (FR), Germany (DE), Greece (EL), Iceland (IS), Ireland (IE), Italy (IT), Luxembourg (LU), Netherlands (NL), Norway (NO), Portugal (PT), Spain (ES), Sweden (SE), Switzerland (CH), Turkey (TR), United Kingdom (UK), and United States of America (US).

Between 2016 and 2020, the total population of mobile international students in the twenty founding members of the OECD experienced a remarkable growth of 16.8%. Specifically, the number jumped from 555,920 students in 2016 to 649,376 students in 2020, reflecting a significant increase in the number of tertiary students enrolled abroad. As highlighted in [25], the COVID-19 pandemic had a very uneven impact on international student exchanges across countries during 2019–2020. Nevertheless, the percentage of tertiary students enrolled abroad increased in several of the twenty OECD countries and remained unchanged in many others. Overall, there were about three international students for every national student studying abroad in OECD countries in 2016–2020, but this ratio is equal to or greater than ten in the United Kingdom and the United States [25,26]. In contrast, Luxembourg is among the twenty OECD founder countries with the lowest ratio of international students to national students abroad [25].

The aim here is to explore both the magnitude and the direction of the international mobility of the tertiary students to identify clusters of countries that are the main origins or destinations of the student exchanges across 5 years (2016–2020).

First, the data were transformed by calculating for each country the percentage share of outgoing mobile students moving to each of the other countries for tertiary education.

5.1. Mobility Data: Results from Generalized INDCLUS

The Generalized INDCLUS model (21) was first fitted directly to the original similarity data.

The algorithm was run by varying the number of clusters J from 2 to 7, and the best solution in 100 runs from different random starting partitions was retained to prevent from falling into local minima.

The choice of the optimal number of clusters was determined by looking at the decrease in the relative loss function as J increases. From the scree plot (Figure 4(a)), the partition in $J = 5$ clusters was chosen as the best solution for Generalized INDCLUS.

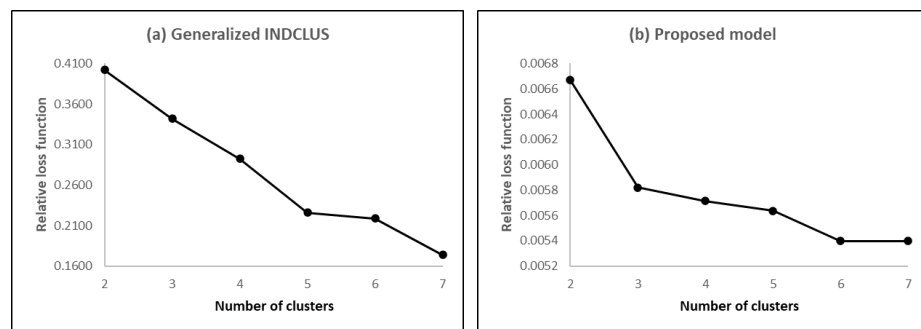


Figure 4. Mobility data: scree plot of the loss values of (a) Generalized INDCLUS and (b) the proposed model.

Table 3 shows the row and column clusters, while the occasion-specific weights are reported in Table 4.

Table 3. Mobility data: row and column clusters from Generalized INDCLUS.

Row Cluster		Column Cluster	
C_1^{Ir}	Austria, Belgium, Denmark, Greece, Ireland, Italy, Luxembourg, Norway, Portugal, Spain, Switzerland, United States	C_1^{Ic}	Germany, United Kingdom
C_2^{Ir}	Canada, Denmark, Iceland, Ireland, Norway, Sweden	C_2^{Ic}	Denmark, United Kingdom, United States
C_3^{Ir}	Belgium, Canada, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States	C_3^{Ic}	France, Netherlands, United Kingdom, United States
C_4^{Ir}	Austria, France, Germany, Luxembourg, Netherlands, Switzerland	C_4^{Ic}	Austria, Belgium, Germany, Switzerland
C_5^{Ir}	Austria, Canada, Turkey, United States	C_5^{Ic}	Germany, United States

All countries belong to at least one cluster of the row covering, but while Iceland and the United Kingdom belong to only one cluster (C_2^{Ir} and C_3^{Ir} , respectively), Austria, Canada, Ireland, Switzerland, and the United States belong to three clusters. The strong overlap of the row clusters indicates a very heterogeneous situation of student mobility with, in principle, many countries from which students leave.

On the other hand, only nine countries belong to at least one column cluster, while the remaining countries are not assigned to any cluster; i.e., they correspond to zero row profiles in the membership matrix Q . This reveals a greater concentration of incoming flows in a few countries than that of outgoing flows from the countries of origin.

The cluster weights of the two coverings (Table 4) are quite similar over time, indicating that the mobility pattern is quite stable over the years considered.

Table 4. Mobility data: occasion-specific weights and constants from Generalized INDCLUS.

Year	Cluster					Constant
	Row	C_1^{Ir}	C_2^{Ir}	C_3^{Ir}	C_4^{Ir}	C_5^{Ir}
	Column	C_1^{Ic}	C_2^{Ic}	C_3^{Ic}	C_4^{Ic}	C_5^{Ic}
	w_{1h}	w_{2h}	w_{3h}	w_{4h}	w_{5h}	c_h
2016	15.398	17.778	9.298	9.852	20.682	1.145
2017	15.145	17.496	9.357	9.689	20.175	1.185
2018	16.961	16.366	8.234	9.545	21.392	1.268
2019	18.168	15.490	7.929	9.836	21.011	1.268
2020	17.361	15.808	8.010	9.859	21.213	1.288

Every year, student mobility is directed from almost all European countries (C_1^{Ir} and C_2^{Ir}) to two main European destinations (Germany and UK), as well as to the US from Canada and Northern Europe.

From the solution, we can say, for example, that Norway and Sweden, together in cluster C_2^{Ir} and separately in clusters C_1^{Ir} and C_3^{Ir} , respectively, have outgoing flows mainly to Denmark, UK, and the US. Swedish students also move to France and the Netherlands, while Norwegian students to Germany. Swedish and Norwegian mobility to all other countries results to being reciprocal to the same extent.

5.2. Mobility Data: Results from the Proposed Model

In order to fit our proposed model that fits dissimilarity data, the proportions of outgoing students in each year were converted to pairwise dissimilarities by taking their complements to 100.

As before, the algorithm was run with J varying from 2 to 7, and the best solution in 100 runs from different random starting partitions was retained. The scree plot of the loss values in Figure 4(b) shows an elbow at the partition with $J = 6$ clusters selected for analysis.

The *complete partition* and the *incomplete partition* estimated from the symmetric and skew-symmetric components of the exchanges, respectively, are reported in Table 5, where countries not assigned to the corresponding clusters of the incomplete partition are indicated in italics.

Table 5. Mobility data: complete and incomplete partition from the proposed model. Italics indicate unassigned countries in the incomplete partition.

	Complete Partition		Incomplete Partition
C_1	Germany, United States	G_1	Germany, United States
C_2	Greece, Ireland, Turkey	G_2	Greece, Ireland, Turkey
C_3	<i>Denmark</i> , Iceland, Norway, Sweden	G_3	Iceland, Norway, Sweden
C_4	<i>Belgium, France, Luxembourg, Netherlands, Portugal, Spain</i>	G_4	Luxembourg, Portugal
C_5	<i>Austria</i> , Canada, Italy, Switzerland	G_5	Canada, Italy, Switzerland
C_6	United Kingdom	G_6	United Kingdom

The six clusters are shown in Figure 5, where different colors indicate different clusters and the unassigned countries of the incomplete partition are shown in light colors on the color scale.

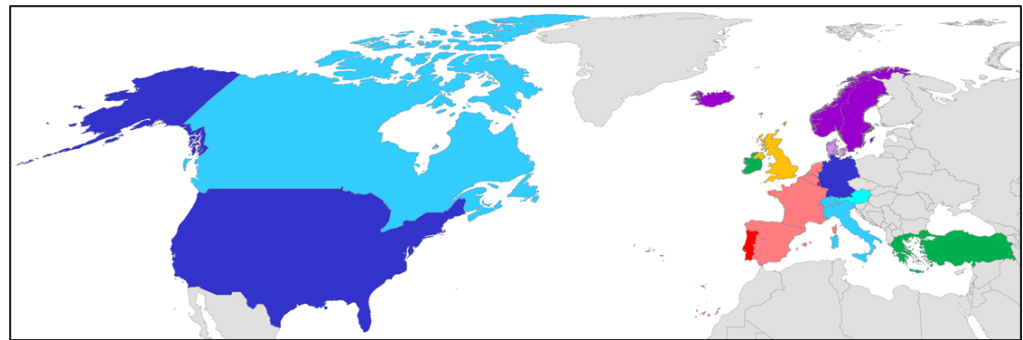


Figure 5. Mobility data: complete and incomplete partition from the proposed model. Different colors indicate different clusters: blue = C_1 , green = C_2 , violet = C_3 , red = C_4 , sky blue = C_5 , yellow = C_6 , and the unassigned countries of the incomplete partition are shown in light colors on the color scale.

The clusters C_1 , C_2 , and C_6 coincide with G_1 , G_2 , and G_6 , respectively, while six countries (Austria, Belgium, Denmark, France, Netherlands, and Spain) do not belong to any cluster of the incomplete partition because their mutual mobility is basically symmetrical.

The two nested partitions highlight the strength of proximity factors, such as language, historical ties, geographical distance, bilateral relationships, and political framework conditions (e.g., the European Higher Education Area) as key determinants for international student mobility. As an example, the Scandinavian countries in cluster C_3 share the same patterns of international student mobility directed to the other OECD countries.

Austria and Belgium, which belong to different clusters (C_3 and C_4 , respectively) but remain unassigned in the incomplete partition, show small mutual mobility exchanges over time (0.7195% in 2016 and 0.8485% in 2020) with minimal imbalances (0.3077% in 2016 and 0.5077% in 2020).

The constants \hat{b}_h ($h = 1, \dots, 5$) increase slightly over the years (Table 6 and Figure 6): this means that since the model has been fitted to the complements to 100 of the share of outgoing students in each year, the baseline average mutual exchange between countries ($100 - \hat{b}_h$, $h = 1, \dots, 5$) shows a slight decrease over the years. This also implies that the average mutual mobility within clusters tends to decrease slightly over time, as the baseline fits the within-cluster mobility in each year.

Table 6. Mobility data: year-specific weights for the complete partition and constants from the proposed model.

Year	Complete Partition						Constant
	C_1	C_2	C_3	C_4	C_5	C_6	
	\hat{r}_{1h}	\hat{r}_{2h}	\hat{r}_{3h}	\hat{r}_{4h}	\hat{r}_{5h}	\hat{r}_{6h}	\hat{b}_h
2016	−5.653	2.757	3.676	2.913	1.273	−10.221	93.135
2017	−5.483	2.684	3.655	2.818	1.267	−10.069	93.158
2018	−6.232	2.496	3.519	2.602	1.074	−10.192	93.570
2019	−5.473	2.447	3.478	2.518	0.987	−10.892	93.589
2020	−6.122	2.461	3.534	2.453	1.174	−10.087	93.575

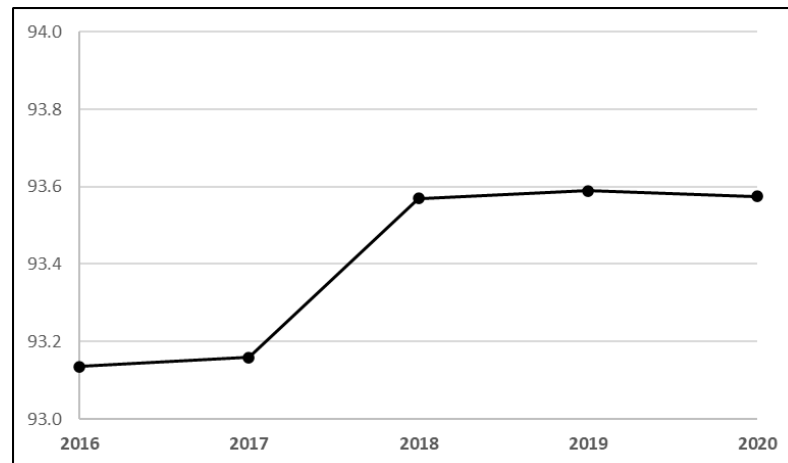


Figure 6. Mobility data: year-specific constants from the proposed model.

In addition to the baseline exchange (\hat{b}_h), the estimated average annual mobility between clusters of countries depends on the weights \hat{r}_h (Table 6). The average mobility has remained relatively stable over time, with very small increases (i.e., very small decreases in the weights \hat{r}_{jh} ($h = 1, \dots, 5$), taking into account the transformation of the percentages to obtain dissimilarities) in almost all clusters until 2019. Fluctuations can be observed for some clusters: in 2019, a slight increase in the average mobility was observed for the United Kingdom ($\hat{r}_{6,2019} = -10.892$ vs. $\hat{r}_{6,2018} = -10.192$), in contrast with a slight decrease for the cluster C_1 with Germany and the United States ($\hat{r}_{1,2019} = -5.473$ vs. $\hat{r}_{1,2018} = -6.232$). Conversely, the situation is reversed in 2020, probably due to the impact of Brexit.

Each year, the ranking of the clusters based on the increasing weights \hat{r}_h consistently ranks the United Kingdom (C_6) with the highest average share of reciprocal mobility (regardless of direction), followed by Germany and the United States (C_1), while Scandinavian countries (C_3) have the lowest, as can be seen in Figure 7.

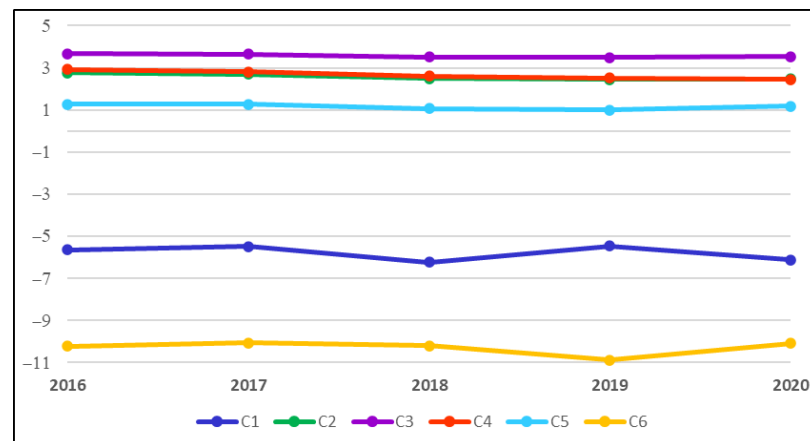


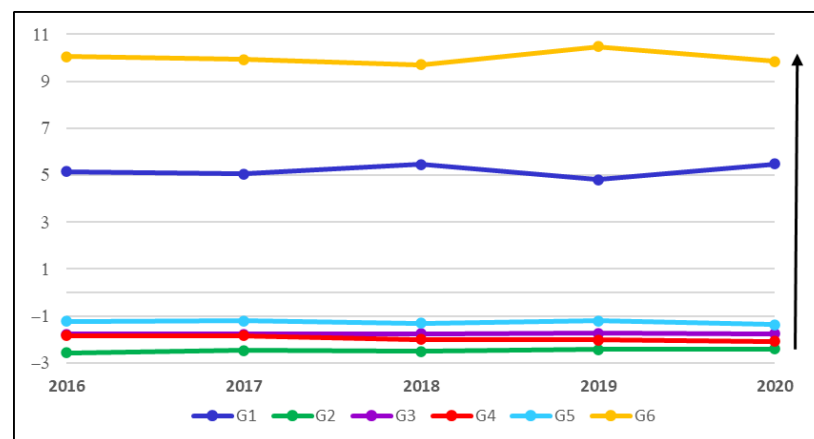
Figure 7. Mobility data: year-specific weights of the complete partition from the proposed model.

As for the directions of the mobility, the analysis of the cluster weights of the incomplete partition (Table 7) reveals a consistent pattern over the years: throughout the 2016–2020 period, the mobility originates from countries in the clusters G_2 , G_3 , G_4 , and G_5 (the weights \hat{t}_{jh} assigned to these clusters are always negative over time) mainly directed to Germany and the United States (G_1), as well as to the United Kingdom (G_6).

Table 7. Mobility data: year-specific weights of the incomplete partition from the proposed model.

Year	Incomplete Partition					
	G_1	G_2	G_3	G_4	G_5	G_6
	\hat{t}_{1h}	\hat{t}_{2h}	\hat{t}_{3h}	\hat{t}_{4h}	\hat{t}_{5h}	\hat{t}_{6h}
2016	5.158	−2.568	−1.772	−1.836	−1.223	10.048
2017	5.051	−2.469	−1.770	−1.842	−1.211	9.932
2018	5.467	−2.497	−1.743	−2.000	−1.308	9.711
2019	4.814	−2.413	−1.738	−2.009	−1.208	10.466
2020	5.470	−2.403	−1.755	−2.087	−1.377	9.839

In all years, the cluster ranking based on the increasing weights \hat{t}_h places Greece, Ireland, and Turkey (G_2), followed by Luxembourg and Portugal (G_4), as the main countries of origin of the international students. As can be seen from the direction of the arrow in Figure 8, the cluster G_5 , which includes Canada, Italy, and Switzerland, is an origin cluster of mobile students toward G_1 and G_6 and, to a greater extent, a destination cluster from G_2 , G_3 , and G_4 . Similarly, the Scandinavian students (G_3) move to Germany, the US, and the UK (G_1 and G_6) and to G_5 , while the Scandinavian countries have incoming mobility from G_2 . This differs in part from the results of Generalized INDCLUS, which does not capture the imbalance of the international mobility from Greece, Ireland, and Turkey directed to the Scandinavian countries.

**Figure 8.** Mobility data: year-specific weights of the incomplete partition from the proposed model (arrow indicates the direction of mobility).

The countries in the two clusters C_2 and C_4 have comparable average mobility over time (the green (C_2) and red (C_4) lines overlap in Figure 7), but with different levels of imbalances: in fact, the student outflows from Greece, Ireland, and Turkey (G_2) are always larger than those from Luxembourg and Portugal (G_4), as can be seen in Figure 8.

English is the most widely spoken language in the globalized world, with one in four people worldwide using it [27]. Not surprisingly, English-speaking countries are always the most attractive study destinations overall: the UK is the top destination in Europe; the United States and Germany are destinations for international students from Europe and Canada and are the source of student mobility to the United Kingdom. Among the top 3 destinations, Germany is the major recipient country in the European Union.

A few countries (clusters G_1 and G_6) are net “importers” of students; that is, they have more students coming in to study than those leaving to study abroad. In contrast, some clusters of countries, such as G_2 and G_4 , are net “exporters” of students.

6. Results and Discussion

In a complex situation where the available data represent asymmetric proximities between pairs of objects measured or observed in different occasions, the proposed model aims to unveil a common clustering structure that accounts for the systematic differences between occasions.

The model proposed here proved to be effective in identifying clusters of objects that share a common pattern of exchange across different occasions, even with a different role as origin or destination. The artificial data and the real application have shown as main results (1) the capability of the model to identify clusters of objects with similar exchange behavior in magnitude and direction (origin/destination clusters) and (2) the flexibility to capture possible differences in the directions of such cluster exchanges across occasions, compared with the Generalized INDCLUS model.

Actually, the goal of Generalized INDCLUS is not to summarize the average exchanges of each cluster, and compared with our model, it provides much more complicated solutions in order to derive the behavior of each cluster because of the overlap between clusters, as can also be seen from the simple and small illustrative example. We can observe that, in the special case where the clusters from Generalized INDCLUS form a partition (they do not overlap), as in our model, it actually accounts for the variability *within* clusters and is not able to capture the exchanges *between* clusters, which is the very goal of our proposal. However, as shown in Sections 4 and 5, we experienced that, in the general case, the clusters resulting from Generalized INDCLUS have (often many) objects in common, which, on the one hand, guarantees flexibility and the possibility to estimate the exchange between pairs of objects belonging to different clusters. On the other hand, this comes at the expense of ease of interpretation and synthesis, since the behavior of objects belonging to the same cluster is not the same regardless of the destination cluster.

Thus, within the scarcity of methods for clustering three-way asymmetric proximity data, our proposal is effective, more parsimonious, and promising when the aim is to obtain a synthesis of the average behavior of exchange between clusters across occasions (which could be useful for policy makers, for example). If, instead, the estimation of any pairwise exchange is the main concern, Generalized INDCLUS might be more flexible in general due to the possible overlap.

Further developments may consider the possibility of assuming a fuzzy clustering, where, instead of a crisp membership of an object to each cluster (u_{ij} in $\{0, 1\}$), a membership degree in $[0, 1]$ is allowed. In addition, the inclusion of possible covariates in the model, if available, could also be taken into consideration to better investigate the determinants of the exchanges.

Author Contributions: Conceptualization, L.B. and D.V.; methodology, L.B. and D.V.; software, D.V.; validation, L.B. and D.V.; formal analysis, L.B.; investigation, L.B.; resources, D.V.; data curation, D.V.; writing—original draft preparation, L.B. and D.V.; visualization, L.B.; supervision, D.V.; project administration, D.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article, with the exception of the data analyzed in Section 5, which are available on the OECD Education Statistics Database, at https://stats.oecd.org/Index.aspx?DataSetCode=EDU_ENRL_MOBILE (accessed on 1 February 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Saito, T.; Yadohisa, H. *Data Analysis of Asymmetric Structures. Advanced Approaches in Computational Statistics*; Marcel Dekker: New York, NY, USA, 2005.
2. Bove, G.; Okada, A. Methods for the analysis of asymmetric pairwise relationships. *Adv. Data Anal. Classif.* **2018**, *12*, 5–31.
3. Bove, G.; Okada, A.; Vicari, D. *Methods for the Analysis of Asymmetric Proximity Data*; Springer Nature Singapore: Singapore, 2021.
4. Furnas, G.W. *Objects and Their Features: The Metric Representation of Two Class Data*. Unpublished Doctoral Dissertation, Stanford University, Stanford, CA, USA, 1980.

5. DeSarbo, W.S.; De Soete, G. On the Use of Hierarchical Clustering for the Analysis of Nonsymmetric Proximities. *J. Consum. Res.* **1984**, *11*, 601–610.
6. DeSarbo, W.S.; Manrai, A.K.; Burke, R.R. A Nonspatial Methodology for the Analysis of Two-Way Proximity Data Incorporating the Distance–Density Hypothesis. *Psychometrika* **1990**, *55*, 229–253.
7. De Soete, G.; DeSarbo, W.S.; Furnas, G.W.; Carroll, J.D. The Estimation of Ultrametric and Path Length Trees from Rectangular Proximity Data. *Psychometrika* **1984**, *49*, 289–310.
8. DeSarbo, W.S. GENNCLUS: New models for general nonmetric clustering analysis. *Psychometrika* **1982**, *47*, 449–475.
9. Hubert, L. Min and max hierarchical clustering using asymmetric similarity measures. *Psychometrika* **1973**, *38*, 63–72.
10. Fujiwara, H. Methods for Cluster Analysis Using Asymmetric Measures and Homogeneity Coefficient. *Jpn. J. Behav.* **1980**, *7*, 12–21.
11. Yadohisa, H. Formulation of Asymmetric Agglomerative Hierarchical Clustering and Graphical Representation of Its Result. *Bull. Comput. Stat. Jpn.* **2002**, *15*, 309–316.
12. Takeuchi, A.; Saito, T.; Yadohisa, H. Asymmetric agglomerative hierarchical clustering algorithms and their evaluations. *J. Classif.* **2007**, *24*, 123–143.
13. Olszewski, D. Asymmetric k-means algorithm. In *Lecture Notes in Computer Science, Part II, Proceedings of the International Conference on Adaptive and Natural Computing Algorithm (ICANNGA 2011), Ljubljana, Slovenia, 14–16 April 2011*; Dovnikar, A., Lotric, U., Ster, B., Eds.; Springer: Heidelberg, Germany, 2011; Volume 6594, pp. 1–10.
14. Olszewski, D. K-means clustering of asymmetric data. In *Hybrid Artificial Intelligent Systems, Proceedings of the HAIS 2012, Salamanca, Spain, 28–30 March 2012, Lecture Notes in Computer Science*; Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.B., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7208, pp. 243–254.
15. Olszewski, D.; Ster, B. Asymmetric clustering using the Alpha-Beta divergence. *Pattern Recognit.* **2014**, *47*, 2031–2041.
16. Vicari, D. Classification of asymmetric proximity data. *J. Classif.* **2014**, *31*, 386–420.
17. Vicari, D. Modeling Asymmetric Exchanges Between Clusters. In *Advanced Studies in Behaviormetrics and Data Science*; Imaizumi, T., Nakayama, A., Yokoyama, S., Eds.; Springer Nature Singapore: Singapore, 2020; pp. 297–313.
18. Vicari, D. CLUSKEXT: CLUstering model for SKew-symmetric data including EXternal information. *Adv. Data Anal. Classif.* **2018**, *12*, 43–64.
19. Vicari, D.; Di Nuzzo, C. A between-cluster approach for clustering skew-symmetric data. *Adv. Data Anal. Classif.* **2024**, *12*, 163–192.
20. Chaturvedi, A.; Carroll, J.D. An alternating combinatorial optimization approach to fitting the INDCLUS and Generalized INDCLUS models. *J. Classif.* **1994**, *11*, 155–170.
21. Carroll, J. D.; Arabie, P. INDCLUS: An individual differences generalization of ADCLUS model and the MAPCLUS algorithm. *Psychometrika* **1983**, *48*, 157–169.
22. Gower, J.C. The analysis of asymmetry and orthogonality. In *Recent Developments in Statistics*; Van Cutsem, B., Barra, J.R., Brodeau, F., Romier, G., Eds.; North Holland: Amsterdam, The Netherlands, 1977; pp. 109–123.
23. McDonald, R.P. A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *Br. J. Math. Stat. Psychol.* **1980**, *33*, 161–183.
24. Rao, C.R.; Mitra, S. *Generalized Inverse of Matrices and Its Applications*; Wiley: New York, NY, USA, 1971.
25. OECD. *Education at a Glance 2023: OECD Indicators*; OECD Publishing: Paris, France, 2023.
26. OECD. *Education at a Glance 2019: OECD Indicators*; OECD Publishing: Paris, France, 2019.
27. Sharifian, F. Globalisation and developing metacultural competence in learning English as an international language. *Multiling. Educ.* **2013**, *3*, 7.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.