GINDCLUS: Generalized INDCLUS with External Information Supplementary Material

Laura Bocci and Donatella Vicari

Sapienza University of Rome, Italy

S.1 GINDCLUS: Special Cases

An interesting framework where the GINDCLUS model may be positioned, together with other existing INDCLUS-related models, can be considered and a classification of the clustering models for three-way proximity data can be sketched.

The Local INDCLUS model, presented in Sect. 2.1 of the paper, represents a constrained version of the INDCLUS model since it adds the clustering of the subjects to INDCLUS.

The Local INDCLUS model is defined as

$$\underline{\mathbf{S}} = \mathbf{P}\underline{\mathbf{W}}(\mathbf{U} \otimes \mathbf{P})' + \left(\mathbf{C}' \otimes \mathbf{1}_N \mathbf{1}'_N\right) + \underline{\mathbf{E}} \tag{S.1}$$

where:

- $\underline{\mathbf{S}} = [\mathbf{S}_1, \ldots, \mathbf{S}_h, \ldots, \mathbf{S}_H]$ denotes the $N \times NH$ supermatrix obtained by collecting the H matrices $\mathbf{S}_1, \ldots, \mathbf{S}_H$ of order $N \times N$ next to each other;
- $\mathbf{P}$ is a $N \times J$ binary matrix defining the covering of the N objects into J overlapping clusters;

- $\underline{\mathbf{W}} = [\mathbf{W}_1, \ldots, \mathbf{W}_k, \ldots, \mathbf{W}_K]$ is the $J \times JK$ supermatrix formed by collecting the K diagonal non-negative class weight matrices $\mathbf{W}_k$ of order J next to each other;

- $\mathbf{U}$ is the $H \times K$ binary membership matrix identifying the partition of the H subjects into K classes;

- $\mathbf{C} = \mathbf{U}[c_1, \ldots, c_k, \ldots, c_K]'$ and c_k is a real-valued additive constant for class k $(k = 1, \ldots, K)$;

- $\mathbf{1}_N$ denotes the column vector with N ones;

- $\underline{\mathbf{E}}$ is the error matrix which explains the part of $\underline{\mathbf{S}}$ not accounted for by the model.

Hereafter, for the sake of brevity, $\mathbf{1}_A$ and $\mathbf{I}_A$ generally denote the column vector with A ones and the $A \times A$ identity matrix, respectively, and $\underline{\mathbf{Y}} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_a, \ldots, \mathbf{Y}_A]$ denotes the $R \times TA$ supermatrix obtained by collecting the A matrices $\mathbf{Y}_1, \ldots, \mathbf{Y}_A$ of order $R \times T$ next to each other.

A generalization of both INDCLUS and Local INDCLUS is represented by the GINDCLUS model in terms of both clustering of the subjects and including external variables.

In GINDCLUS (see Sect. 2.1 in the paper) the set of non-negative class weight matrices $\mathbf{W}_k$ are modeled as the product of two terms g_k and $\mathbf{B}_k$ which depend on the set of object and subject external variables $\mathbf{X}$ and $\mathbf{Z}$, respectively. The class weight matrices $\mathbf{W}_k$ are modeled as follows

$$\mathbf{W}_k = g_k \mathbf{B}_k = g_k diag(\mathbf{b}_k), \qquad (k = 1, \ldots, K), \tag{S.2}$$

where:

- $\mathbf{B}_k = diag(\mathbf{b}_k)$ is a diagonal matrix of size J whose main diagonal is the vector $\mathbf{b}_k$ of the non-negative class-conditional weights assigned to the J groups of objects

$$\mathbf{b}_k = \overline{\mathbf{X}}\boldsymbol{\beta}_k + \mathbf{e}_1, \qquad (k = 1, \ldots, K), \tag{S.3}$$

subject to

$$\mathbf{b}_k \geq \mathbf{0} \quad \text{and} \quad \mathbf{b}_k' \mathbf{1}_J = 1 \qquad (k = 1, \ldots, K), \tag{S.4}$$

where $\overline{\mathbf{X}}$ is the $J \times M$ matrix of the centroids of the J groups of objects (i.e., the j-th row of $\overline{\mathbf{X}}$ is the mean of the $N \times M$ matrix $\mathbf{X}$ of the object external variables over the objects belonging to group j), $\boldsymbol{\beta}_k$ is a vector of coefficients of size M and $\mathbf{e}_1$ is a vector of error components;

- g_k is a non-negative class-specific weight defined as

$$g_k = \overline{\mathbf{z}}_k' \boldsymbol{\gamma} + \mathbf{e}_2, \quad \text{subject to} \quad g_k \geq 0, \quad (k = 1, \ldots, K), \tag{S.5}$$

where $\overline{\mathbf{z}}_k$ is the column vector of the V coordinates of the centroid of the k-th class of subjects (i.e., the mean of the $H \times V$ matrix $\mathbf{Z}$ of the subject external variables over the subjects belonging to class k), $\boldsymbol{\gamma}$ is a vector of coefficients of size V and $\mathbf{e}_2$ is a vector of error components.

Therefore, GINDCLUS is defined as

$$\underline{\mathbf{S}} = \mathbf{P}\underline{\mathbf{B}}\left((\mathbf{U}\mathbf{G}) \otimes \mathbf{P}\right)' + \left(\mathbf{C}' \otimes \mathbf{1}_N \mathbf{1}_N'\right) + \underline{\mathbf{E}} \tag{S.6}$$

where $\underline{\mathbf{B}} = [\mathbf{B}_1, \ldots, \mathbf{B}_k, \ldots, \mathbf{B}_K]$ is the $J \times JK$ supermatrix formed by the K matrices $\mathbf{B}_k = diag(\mathbf{b}_k)$ and $\mathbf{G} = diag(\mathbf{g})$ is the diagonal matrix of size K obtained from

$\mathbf{g} = [g_1, \ldots, g_k, \ldots, g_K]'$.

In the design of a general model for clustering two modes of three-way proximity data taking into account the subject heterogeneity, there are two key elements: the number n_C of classes of subjects (and therefore distinct weight matrices) and the number n_G of groups of objects allowed in the model. Table 1 shows all the conceivable cases of the clustering models for three-way two-mode proximity data.

Table 1: Classification of clustering models for three-way proximity data

Groups of objects	**Classes of subjects (n_C)**				
(n_G)	H	$\cdots$	K	$\cdots$	1
N	(A) No classification		(B) Subjects in K classes (S.7) and (S.8)		(C) Subjects in one class (S.9) and (S.10)
$\vdots$					
J	(D) INDCLUS (S.11) and (S.12)		(E) GINDCLUS (S.1) and (S.6)		(F) ADCLUS-type model (S.13) and (S.14)
$\vdots$					
1	(G) Mean profile model (S.15) and (S.16)		(H) Class-conditional mean model (S.17) and (S.18)		(I) Extreme GINDCLUS (S.19)

Specifically:

(A) $\underline{n_G = N \text{ and } n_C = H}$. It is worth noting that no classification is done.

(B) $\underline{n_G = N \text{ and } n_C = K}$. This corresponds to a one-mode classification model, because only subjects are partitioned. Here, $\mathbf{P} = \mathbf{I}_N$ and model (S.1) becomes

$$\underline{\mathbf{S}} = \underline{\mathbf{W}}\,(\mathbf{U} \otimes \mathbf{I}_N)' + \left(\mathbf{C}' \otimes \mathbf{1}_N \mathbf{1}_N'\right) + \underline{\mathbf{E}}. \tag{S.7}$$

where $\underline{\mathbf{W}}$ is the $N \times NK$ supermatrix formed by the $N \times N$ matrices $\mathbf{W}_k$ ($k = 1, \ldots, K$). When $\mathbf{W}_k$ is modeled as in (S.2)-(S.5), model (S.6) becomes

$$\underline{\mathbf{S}} = \underline{\mathbf{B}}\,((\mathbf{UG}) \otimes \mathbf{I}_N)' + \left(\mathbf{C}' \otimes \mathbf{1}_N \mathbf{1}_N'\right) + \underline{\mathbf{E}}. \tag{S.8}$$

where $\underline{\mathbf{B}} = [\mathbf{B}_1, \ldots, \mathbf{B}_k, \ldots, \mathbf{B}_K]$ with $\mathbf{B}_k = diag(\mathbf{b}_k) = diag(\mathbf{X}\boldsymbol{\beta}_k)$ $(k = 1, \ldots, K)$.

(C) $\underline{n_G = N \text{ and } n_C = 1}$. Since $\mathbf{P} = \mathbf{I}_N$ and $\mathbf{U} = \mathbf{1}_H$, model (S.1) is

$$\underline{\mathbf{S}} = \left(\mathbf{1}'_H \otimes \mathbf{W}\right) + c\mathbf{1}_N\mathbf{1}'_{HN} + \underline{\mathbf{E}} \tag{S.9}$$

where $\mathbf{W}$ is the unique $N \times N$ matrix of weights. When $\mathbf{W}$ is modeled as in (S.2)-(S.5), model (S.6) reduces to

$$\underline{\mathbf{S}} = \left(\mathbf{1}'_H \otimes \left(\bar{\mathbf{z}}'\boldsymbol{\gamma}\right)\mathbf{B}\right) + c\mathbf{1}_N\mathbf{1}'_{HN} + \underline{\mathbf{E}} \tag{S.10}$$

being $\mathbf{B} = diag(\mathbf{X}\boldsymbol{\beta})$ and $\bar{\mathbf{z}}$ the mean vector of $\mathbf{Z}$.

(D) $\underline{n_G = J \text{ and } n_C = H}$. Since $\mathbf{U} = \mathbf{I}_H$, only one-mode (the objects) is classified and model (S.1) reduces to the INDCLUS model

$$\underline{\mathbf{S}} = \mathbf{P}\underline{\tilde{\mathbf{W}}}\left(\mathbf{I}_H \otimes \mathbf{P}'\right) + \left(\tilde{\mathbf{C}}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \underline{\mathbf{E}} \tag{S.11}$$

where $\underline{\tilde{\mathbf{W}}} = [\tilde{\mathbf{W}}_1, \ldots, \tilde{\mathbf{W}}_h, \ldots, \tilde{\mathbf{W}}_H]$ has size $J \times JH$, $\tilde{\mathbf{C}} = [\tilde{c}_1, \ldots, \tilde{c}_h, \ldots, \tilde{c}_H]'$. However, when the individual weight matrices are modeled as in (S.2)-(S.5), INDCLUS becomes

$$\underline{\mathbf{S}} = \mathbf{P}\underline{\tilde{\mathbf{B}}}\left(\tilde{\mathbf{G}} \otimes \mathbf{P}\right) + \left(\tilde{\mathbf{C}}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \underline{\mathbf{E}} \tag{S.12}$$

where $\underline{\tilde{\mathbf{B}}} = [\tilde{\mathbf{B}}_1, \ldots, \tilde{\mathbf{B}}_h, \ldots, \tilde{\mathbf{B}}_H]$ being $\tilde{\mathbf{B}}_h = diag(\overline{\mathbf{X}}\boldsymbol{\beta}_h)$ and $\tilde{\mathbf{G}} = diag(\mathbf{Z}\boldsymbol{\gamma})$.

(E) $\underline{n_G = J \text{ and } n_C = K}$. This case corresponds to the Local INDCLUS model (S.1) or to the GINDCLUS model (S.6) when the assumptions (S.2)-(S.5) hold.

(F) $\underline{n_G = J \text{ and } n_C = 1}$. Since $\mathbf{U} = \mathbf{1}_H$, the model is an ADCLUS-type model equivalent to apply ADCLUS to the (two-way) mean matrix. In this case, model (S.1) becomes

$$\underline{\mathbf{S}} = \left(\mathbf{1}'_H \otimes \mathbf{PWP}'\right) + c\mathbf{1}_N\mathbf{1}'_{HN} + \underline{\mathbf{E}} \tag{S.13}$$

where $\mathbf{W}$ is the unique $J \times J$ matrix of weights. When $\mathbf{W}$ is modeled as in (S.2)-(S.5), model (S.6) reduces to

$$\underline{\mathbf{S}} = \bar{\mathbf{z}}'\boldsymbol{\gamma}\left(\mathbf{1}'_H \otimes \mathbf{PBP}'\right) + c\mathbf{1}_N\mathbf{1}'_{HN} + \underline{\mathbf{E}} \tag{S.14}$$

being $\mathbf{B} = diag(\overline{\mathbf{X}}\boldsymbol{\beta})$.

(G) $\underline{n_G = 1 \text{ and } n_C = H}$. Since $\mathbf{P} = \mathbf{1}_N$ and $\mathbf{U} = \mathbf{I}_H$, a *mean profile* model derives

$$\underline{\mathbf{S}} = \left(\tilde{\mathbf{w}}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \left(\tilde{\mathbf{C}}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \underline{\mathbf{E}} \tag{S.15}$$

where $\tilde{\mathbf{w}} = [\tilde{w}_1, \dots, \tilde{w}_h, \dots, \tilde{w}_H]'$ is the unique non-negative vector of the individual weights assigned to the whole group of objects by every subject. When the individual weights are modeled as in (S.2)-(S.5), and being $\tilde{b}_h = 1$ because of (S.4), $\tilde{w}_h = \tilde{g}_h$ $(h = 1, \dots, H)$ follows and model (S.6) reduces to

$$\underline{\mathbf{S}} = \left(\tilde{\mathbf{g}}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \left(\tilde{\mathbf{C}}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \underline{\mathbf{E}} \tag{S.16}$$

where $\tilde{\mathbf{g}} = [\tilde{g}_1, \dots, \tilde{g}_h, \dots, \tilde{g}_H]'$.

(H) $\underline{n_G = 1 \text{ and } n_C = K}$. Since $\mathbf{P} = \mathbf{1}_N$, a *class-conditional mean* model derives

$$\underline{\mathbf{S}} = \left((\mathbf{Uw})' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \left(\mathbf{C}' \otimes \mathbf{1}_N\mathbf{1}'_N\right) + \underline{\mathbf{E}} \tag{S.17}$$

where $\mathbf{w} = [w_1, \dots, w_k, \dots, w_K]'$ is the vector of the class-specific weights.

These weights modeled as in (S.2)-(S.5) become $b_k = 1$ $(k = 1, \dots, K)$ due to (S.4) and then $w_k = g_k$ $(k = 1, \dots, K)$ where g_k is defined as in (S.5). Therefore, model (S.6) becomes

$$\underline{\mathbf{S}} = \left((\mathbf{Ug})' \otimes \mathbf{1}_N \mathbf{1}_N'\right) + \left(\mathbf{C}' \otimes \mathbf{1}_N \mathbf{1}_N'\right) + \underline{\mathbf{E}}. \tag{S.18}$$

(I) $\underline{n_G = 1 \text{ and } n_C = 1}$. Since $\mathbf{P} = \mathbf{1}_N$ and $\mathbf{U} = \mathbf{1}_H$, the *Extreme* GINDCLUS model is obtained

$$\underline{\mathbf{S}} = (w + c)\,\mathbf{1}_N \mathbf{1}_{HN}' + \underline{\mathbf{E}} \tag{S.19}$$

where w is the unique non-negative weight assigned to the whole group of objects by all subjects. When $w = gb$ as in (S.2)-(S.5), it follows that $g = \overline{\mathbf{z}}'\boldsymbol{\gamma}$ and $b = 1$ for (S.4).

S.2 Assessing the Constrained Models

It can be useful to have for the GINDCLUS model (11) a decomposition of the loss function (15) which represents the GINDCLUS Error Sum of Squares (*GIESS*), so that it can be partitioned into components assessing different aspects of the lack of fit of the model due to the clustering of the subjects and the taking into account of the external variables.

Let $\underline{\mathbf{s}}$ be the $(N^2 \times H)$ matrix formed by collecting the H column vectors $\mathbf{s}_h$ next to each other, $\mathbf{P}$ and $\mathbf{U}$ be the classification matrices fitted by the GINDCLUS model and $\mathbf{T}$ be the $(N^2 \times J)$ matrix built from $\mathbf{P}$ as in step *c)* of the algorithm (Sect. 3 in the paper). Moreover, let $\mathbf{W}^{IND}$ denote the $(J \times H)$ matrix of the weights estimated from the INDCLUS model (2) and $\mathbf{W}^{LIND}$ and $\mathbf{W}^{GIND}$ be the weights fitted by the Local INDCLUS and GINDCLUS models, respectively, rearranged to form $(J \times K)$ weight matrices.

Without loss of generality and just for simplicity, we assume here that the additive constant parts of the INDCLUS, Local INDCLUS and GINDCLUS models are null.

Given $\mathbf{P}$ and $\mathbf{U}$, the INDCLUS model (2), Local INDCLUS model (4) and GINDCLUS model (11) can be equivalently rewritten as:

$$\begin{aligned} \underline{\mathbf{s}} &= \mathbf{T}\mathbf{W}^{IND} + \underline{\mathbf{e}}_1 && \text{(S.20)} \\ \underline{\mathbf{s}} &= \mathbf{T}\mathbf{W}^{LIND}\mathbf{U}' + \underline{\mathbf{e}}_2 && \text{(S.21)} \\ \underline{\mathbf{s}} &= \mathbf{T}\mathbf{W}^{GIND}\mathbf{U}' + \underline{\mathbf{e}}_3 && \text{(S.22)} \end{aligned}$$

where $\underline{\mathbf{e}}_1$, $\underline{\mathbf{e}}_2$ and $\underline{\mathbf{e}}_3$ are the error terms.

Therefore, the corresponding least-squares loss functions are

$$\begin{aligned} IESS &= \left\|\underline{\mathbf{s}} - \mathbf{T}\mathbf{W}^{IND}\right\|^2 && \text{(S.23)} \\ LIESS &= \left\|\underline{\mathbf{s}} - \mathbf{T}\mathbf{W}^{LIND}\mathbf{U}'\right\|^2 && \text{(S.24)} \\ GIESS &= \left\|\underline{\mathbf{s}} - \mathbf{T}\mathbf{W}^{GIND}\mathbf{U}'\right\|^2 && \text{(S.25)} \end{aligned}$$

where the terms *IESS*, *LIESS* and *GIESS* denote the Error Sum of Squares of the INDCLUS, Local INDCLUS and GINDCLUS models, respectively.

Given $\mathbf{P}$, the unconstrained least-squares solution minimizing *IESS* (S.23) is

$$\mathbf{W}^{IND} = \mathbf{T}^{+}\underline{\mathbf{s}} \qquad \text{(S.26)}$$

where $\mathbf{T}^{+}$ denotes the pseudo-inverse of $\mathbf{T}$.

Moreover, *given* $\mathbf{P}$ and $\mathbf{U}$, the unconstrained least-squares solution of *LIESS*

(S.24) is

$$\mathbf{W}^{LIND} = \mathbf{T}^{+}\underline{\mathbf{s}}\mathbf{U}^{+\prime} = \mathbf{W}^{IND}\mathbf{U}^{+\prime} \tag{S.27}$$

where $\mathbf{U}^{+}$ denotes the pseudo-inverse of $\mathbf{U}$ and the right-hand term derives from (S.26).

By subtracting and adding $\mathbf{T}\mathbf{W}^{IND}$ into *LIESS* (S.24) and using (S.27), it follows

$$\begin{aligned} LIESS &= \left\|\underline{\mathbf{s}} - \mathbf{T}\mathbf{W}^{IND}\right\|^2 + \left\|\mathbf{T}\left(\mathbf{W}^{IND} - \mathbf{W}^{LIND}\mathbf{U}'\right)\right\|^2 \\ &= IESS + \left\|\mathbf{T}\mathbf{W}^{IND}\left(\mathbf{I}_H - \mathbf{U}\mathbf{U}^{+}\right)\right\|^2 \\ &= IESS + LC \end{aligned} \tag{S.28}$$

where *LC* (Lack of Clustering) denotes the part of the INDCLUS variability not accounted for by the partition of subjects identified by $\mathbf{U}$.

Similarly, starting from *GIESS* (S.25), by subtracting and adding $\mathbf{T}\mathbf{W}^{LIND}\mathbf{U}'$ and using (S.27)

$$\begin{aligned} GIESS &= \left\|\underline{\mathbf{s}} - \mathbf{T}\mathbf{W}^{LIND}\mathbf{U}'\right\|^2 + \left\|\mathbf{T}\left(\mathbf{W}^{LIND} - \mathbf{W}^{GIND}\right)\mathbf{U}'\right\|^2 \\ &= LIESS + \left\|\mathbf{T}\left(\mathbf{W}^{IND}\mathbf{U}\mathbf{U}^{+} - \mathbf{W}^{GIND}\mathbf{U}'\right)\right\|^2 \\ &= LIESS + LR \end{aligned} \tag{S.29}$$

where *LR* denotes the Lack of Regression fit due to the external variables.

Finally, plugging (S.28) into (S.29), the following decomposition holds

$$GIESS = IESS + LC + LR \tag{S.30}$$

which is just the decomposition (23) in Sect. 4.1 of the paper. The last two terms on the right-hand side in (S.30) measure the lack of fit due to the (clustering and regression) constraints additionally imposed to the INDCLUS model and can be used to asses the different aspects of the model misfit.

Note that the decomposition (S.30) still holds even when $\mathbf{W}^{IND}$, $\mathbf{W}^{LIND}$ and $\mathbf{W}^{GIND}$ are constrained to be non-negative because the optimal (constrained) estimates are obtained by setting to zero the negative entries and deriving the unconstrained least-squares solutions for the remaining ones (as in active-set-type algorithms).