HAL
open science

Analysing a quality of life survey using a co-clustering model for ordinal data and some dynamic implications

Margot Selosse, Julien Jacques, Christophe Biernacki, Florence Cousson-Gélie

To cite this version:

Margot Selosse, Julien Jacques, Christophe Biernacki, Florence Cousson-Gélie. Analysing a quality of life survey using a co-clustering model for ordinal data and some dynamic implications. Journal of the Royal Statistical Society: Series C Applied Statistics, 2019, 68 (Part 5), pp.1327-1349. 10.1111/rssc.12365 . hal-01643910v3

HAL Id: hal-01643910

https://hal.science/hal-01643910v3

Submitted on 9 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysing a quality of life survey using a co-clustering model for ordinal data and some dynamic implications

Margot Selosse

Université de Lyon, Lyon 2, ERIC EA 3083, Lyon, France.

Julien Jacques

Université de Lyon, Lyon 2, ERIC EA 3083, Lyon, France.

Christophe Biernacki

Inria, Université de Lille, CNRS, Lille, France.

Florence Cousson-Gélie

Université Paul Valéry Montpellier 3, Université Montpellier, EPSLYON EA 4556, F34000, Montpellier, France.

Summary. The dataset that motivated this work is a psychological survey on women affected by a breast tumour. Patients replied at different stages of their treatment to questionnaires with answers on an ordinal scale. The questions relate to aspects of their life referred to as "dimensions". To assist psychologists in analysing the results, it is useful to highlight the structure of the dataset. The clustering method achieves this by creating groups of individuals that are depicted by a representative of the group. From a psychological position, it is also useful to observe how questions may be clustered. The simultaneous clustering of both patients and questions is called "co-clustering". However, placing questions in the same group when they are not related to the same dimension does not make sense from a psychological perspective. Therefore, constrained co-clustering was performed to prevent questions of different dimensions from being placed in the same column-cluster. The evolution of co-clusters over time was then investigated. The method uses a constrained Latent Block Model embedding a probability distribution for ordinal data. Parameter estimation relies on a stochastic EM algorithm associated with a Gibbs sampler, and the ICL-BIC criterion is used to select the number of co-clusters.

1. Introduction

The aim of this work is to provide an efficient tool for exploring ordinal data from psychological surveys. Indeed, when psychology experts set up surveys for a study, they often collect a large quantity of data, both in terms of the number of individuals (people) and the number of variables (questions). This is advantageous because the more data they collect, the more reliable their conclusions will be. Nevertheless, shortly after collecting the data, a first exploratory phase is necessary. This phase of understanding makes it possible to synthesize the data, to distinguish structures inherent to the data and to detect anomalies if they exist. It also allows better visualization and overall understanding of the data. Unsupervised algorithms for clustering or pattern detection help provide a global overview of a dataset. In high dimensions, when the number of variables is large, it is often useful to simultaneously cluster the rows and columns of the

data (Govaert and Nadif (2003)). Indeed, as it is necessary to summarize the individuals into homogeneous groups, it is also interesting to summarize the variables. The result of this type of simultaneous clustering, referred to as "co-clustering", provides a more refined synthesis by summarizing the dataset using blocks, as illustrated by Figure 1. This work presents a novel method for co-clustering ordinal data, as the psychological surveys that motivated this work are based on questionnaires with answers on ordinal scales.

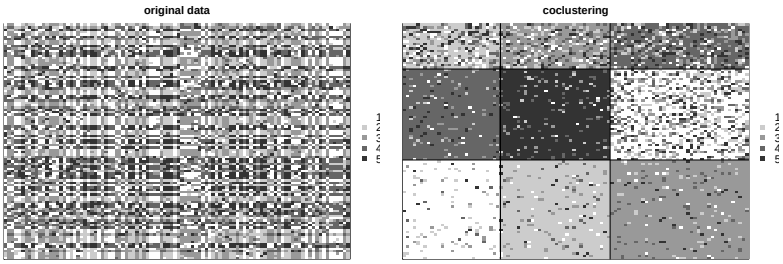

Fig. 1. Original dataset (left) and co-clustering results (right).

The dataset that initiated this work is a survey carried out among women affected by breast cancer (Cousson-Gélie (2014)). Individuals with cancer usually experience traumatizing hardships such as chemotherapy and intense stress. The disease and its treatment have an impact on different domains of their environment, such as their social life or emotional state. In psychology, these domains are divided into dimensions. For example, in Table 1, the domain *quality of life* is divided into six dimensions (physical functioning, role functioning, social functioning, emotional functioning, cognitive functioning and global health evaluation). In contrast, the domain *emotional state* is defined using two dimensions: anxiety and depression (Zigmond and Snaith (1983)). Other psychological dimensions have been identified as a quality of life predictor, such as perceived control of the illness, which corresponds to the general belief whereby evolution of the disease depends either on internal factors (action, effort or personal abilities) or on external factors (luck or destiny) (Cousson-Gélie (2014)), or social support, which assesses perceived availability (number of people on whom the individual thinks they can count if necessary) and the degree of satisfaction relating to this support (Sarason et al. (1983)).

The patients were asked to reply to various questionnaires related to distinct dimensions, the answers being of the ordinal kind with different numbers of levels (Agresti (2010)). They repeated this task at six different stages of their treatment. Therefore, the resulting dataset comprises a set of six tables, with the rows representing the patients and the columns representing the questions. First of all, the psychologists sought to identify psychological profiles. In particular, they wanted to analyse the mutual influence of the different dimensions for each profile. To help them with this task, constrained co-clustering was performed. As mentioned above, co-clustering is a technique that performs simultaneous clustering of the rows and columns of a matrix. As a result, co-clustering highlights the internal structure of the dataset, which in this case makes it possible to detect typical psychological profiles and the groups of questions that differentiate them. The term "constrained" is used because the co-clustering operation was

set up as to ensure that the questions (columns) that did not relate to a common dimension were kept separate. This is referred to as a "cannot-link constraint" (Wagstaff et al. (2001)), here it applies only on the column-clusters. Next, the experts sought to investigate how their patients' answers evolved over certain characteristic periods of time (stages). Indeed, they also wanted to focus on the changes in their psychological state, which is known as the "trajectory" (Annema et al. (2017)). Performing co-clustering at each stage at which the patients had to answer the questionnaires gives a better idea of the changes among patients from different perspectives. On a global scale, it shows how the groups of individuals evolved, and how the replies changed during the study period. On a more precise scale, co-clustering made it possible to analyse the behaviour of a single patient, by allowing the observation of how her row-clusters changed over the time periods.

The dataset exclusively contains values of the ordinal type. Confronted with such data, practitioners often transform them into continuous data by associating an arbitrary number with each level (Kaufman and Rousseeuw (2008); Lewis et al. (2005)), or transform them into nominal data (Vermunt and Magidson (2005)). These choices make it possible to use well-known distributions, but result in either the loss of the information given by the existing order among levels (when considering them as nominal data) or the introduction of an arbitrary notion of distance between levels (when transforming them into continuous data).

Recent contributions have defined clustering algorithms specific to ordinal data. Several contributions use Gaussian latent variables to model the data: in McParland and Gormley (2011), the observed data are viewed as discrete versions of an underlying latent Gaussian variable. In Ranalli and Rocci (2016), the observed categorical variables are considered as a discretization of an underlying finite mixture of Gaussians. Other contributions use the multinomial distribution to model the data. In Giordan and Diana (2011), the authors use the multinomial distribution and a cluster tree, while Jollois and Nadif (2009) use a constrained multinomial distribution. Another approach is to consider a mixture model. For instance, Corduas (2008) proposes a clustering algorithm based on a mixture of CUB models (D'Elia and Piccolo (2005)). In the CUB model, an answer is interpreted as the result of a cognitive process where the decision is intrinsically continuous but is expressed on a discrete scale of m levels. This approach interprets the choice of the respondent as a weighted combination of two components. The first reflects a personal feeling and is expressed by a shifted binomial random variable. The second component reflects an intrinsic uncertainty and is expressed by a uniform random variable. More recently, Biernacki and Jacques (2016) have defined a new distribution for ordinal data, referred to as " BOS", which is used through a mixture model to perform ordinal data clustering. The BOS distribution is defined with two Gaussian-like parameters (μ,π), μ being a position parameter (the mode of the distribution), and π being a precision parameter indicating the spread of the data around the mode. One of the advantages of the BOS model is that its parameters are easy to interpret, which is very important when working with non-statistician professionals. The BOS model will be described in more detail in Section 3.1.

In a co-clustering context, Jacques and Biernacki (2018) define a model-based algorithm relying on the Latent Block Model (LBM, Govaert and Nadif (2003)) embedding

the BOS distribution. Nevertheless, the weakness of this model is its inability to treat variables with different numbers of levels. Indeed, the LBM relies on the assumption that data in a block are independent and identically distributed. This means that to be identically distributed, two variables should at least share the same distribution support. Consequently, the co-clustering model of Jacques and Biernacki (2018) assumes that all the ordinal features have the same number of levels. This is an issue for the psychological survey studied in this paper since the number of levels can be different. Furthermore, as explained before, the questions of the survey being studied are related to psychological dimensions. This means that the variables are already grouped according to the dimension. While it is still interesting to perform co-clustering that will detect smaller groups, it is also important not to group together questions that are not related to the same dimension.

In the present work, co-clustering is performed through a constrained version of the Latent Block Model, so that certain questions cannot be part of the same column-cluster. This extension solves the two issues discussed. Firstly, it allows separation of the questions that do not have the same number of levels.It also makes it possible to constrain the column-clusters so that they are formed with questions regarding the same psychological dimension.

The paper is organized as follows: Section 2 presents the dataset and the notation, while Section 3 explains the statistical models that were used. Section 4 describes the results obtained on the psychological dataset. Lastly, Section 5 concludes this paper.

2. Materials

2.1. Dataset

2.1.1. Description of survey population

Several questionnaires were given to $N = 161$ women having their first surgery for suspicious breast tumour. These patients were between 31 and 77 years old with an average age of 56.25 years (standard deviation = 9.99). Most were married or living maritally (77.0%). Nearly half of the patients were active professionally (49.7%) and 38.5% were retired when they started the study. These 161 patients were asked to answer several questionnaires at different stages of their treatment: one at their first surgery, and followed by a questionnaire 1, 4, 7, 10, 13 months after this assessment. This means that the patients replied six times to 134 questions and each answer was given on an ordinal scale (with between four and seven levels). Therefore, the dataset comprises a set of six matrices of ordinal data such that the observations (rows) correspond to the patients, and the variables (columns) correspond to the questions.

The dataset also contains missing values, for which we distinguish two types. The first type occurred when some patients did not answer any of the questions at one of the six stages (i.e. they did not return the questionnaire at this stage). In this case, when co-clustering was performed solely on the answers for this stage, the rows corresponding to these patients were placed in a special row-cluster called "did not answer" (see Figure 8). Co-clustering was then performed without taking them into account. The second type occurred when some patients failed to answer to only a couple of questions (i.e. they returned an incomplete questionnaire). In this case, the patient was taken into account

Table 1. Table of domains and dimensions raised in the questionnaires.

Domains				
Quality of life (Aaronson et al. (1993))	Social Support (Sarason et al. (1983))	Specific Social Support (Pierce et al. (1997))	Emotional State (Zigmond and Snaith (1983))	Control perception (Cousson-Gélie (2014))
Dimensions				
Physical functioning, Role functioning, Emotional functioning, Cognitive functioning, Social functioning, Global health evaluation.	Satisfaction, Quantity.	Intensity, Perception of availability, Conflict.	Anxiety, Depression.	Causal attribution, Control perception, Religion control.

for co-clustering, and the missing values (18 values in total) were estimated by the algorithm. The way the algorithm deals with this type of missing data is described later on.

2.1.2. Psychological dimensions

The questionnaires given to the patients were detailed; indeed, the design of questionnaires is a highly specialized undertaking in psychology. Each questionnaire relates to domains of life, and each domain is itself divided into dimensions (e.g. MaloneBeach and Zarit (1995)). Table 1 lists the domains and the corresponding dimensions included in the study. In the questionnaires, most of the questions are associated with a dimension. The few questions that are not related to one of these psychological dimensions concern the symptoms of the disease and its treatment (nausea, tiredness, etc.).

2.2. Data representation and conventions

First of all, the dataset was recoded so that for all the questions, the most positive answer was given the level "1". For example, for the question: *"Have you had trouble sleeping?"* with possible responses: *"Not at all"*, *"A little"*, *"Quite a bit"* and *"Very much"*, the following levels were assigned to the replies: 1 *"Not at all"*, 2 *"A little"*, 3 *"Quite a bit"* and 4 *"Very much"*, because it is perceived as more positive not to have had trouble sleeping.

Secondly, a graphical way of representing the data was defined, as shown in Figure 2: the women are projected onto rows and the questions are projected onto columns. Therefore, the cell (i, j) is the reply of patient i to question j. The shades of grey indicate how positively the individual replied. For example, for the question *"Have you had trouble sleeping?"*, if the patient answers *"Not at all"*, the corresponding cell will be white, whereas a response such as *"Very much"* will correspond to a black cell.

2.3. Notation

Firstly, an ordinal variable x with m levels $\{l_1, ..., l_m\}$ is a categorical variable for which the order of levels is significant. The order of levels is indicated by "<": $l_1 < ... < l_m$. For simplicity the levels are numbered $\{1, ..., m\}$ according to their order. Following this notation, an ordinal variable x is an element of $\{1, ..., m\}$.

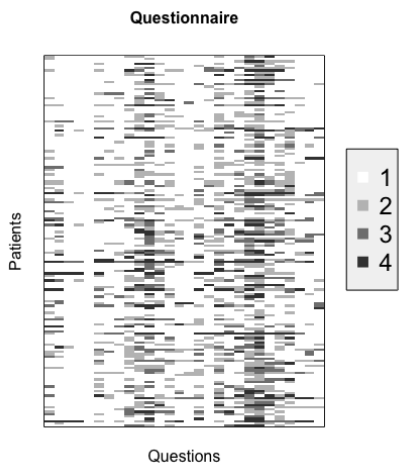

Fig. 2. Graphical representation of the patients' replies. The women are represented by rows and the questions by columns. A cell represents an individual's answer to a question. The darker the cell, the more pessimistically the patient responded.

The representation of the questionnaire responses at a given time will now be detailed. The questions are separated according to two criteria: the number of levels m and the dimension to which they are related. This means that variables with a different number of levels and variables related to different dimensions are separated. This results in a matrix split up into D tables, such that the dth table is an $N \times J_d$ matrix written $\boldsymbol{x}^d$, where N is the number of observations (in this case patients) and J_d the number of questions in the dth table. The matrix $\boldsymbol{x}^d$ comprises ordinal data with a number of levels m_d. Figure 3 illustrates this notation.

$$\boldsymbol{x} = \left[\left[\begin{array}{c} \\ \boldsymbol{x}^1 \\ \\ \end{array} \right] \ \ldots \ \left[\begin{array}{c} \\ \boldsymbol{x}^D \\ \\ \end{array} \right] \right], \text{ with } \boldsymbol{x}^d = (x_{ij}^d)_{i=1,\ldots,N;\ j=1,\ldots,J_d}$$

Fig. 3. Representation of the patients and questions at a given time. Questions related to different dimensions or with a different number of levels m are separated.

The goal of co-clustering is to partition the rows of $\boldsymbol{x}$ into G row-clusters, and the column of each submatrix $\boldsymbol{x}^d$ into H_d column-clusters.

The dataset contains missing data. The whole dataset will be written $\boldsymbol{x} = (\check{\boldsymbol{x}}, with \hat{\boldsymbol{x}})$, $\check{\boldsymbol{x}}$ being the observed data and $\hat{\boldsymbol{x}}$ the missing data. Consequently, a cell of $\boldsymbol{x}$ will be annotated as follows: $\check{x}_{ij}$, if x_{ij} is observed, $\hat{x}_{ij}$ otherwise.

Finally, the bounds for the indices i, j, g, h: $1 \leq i \leq N$, $1 \leq j \leq J$, $1 \leq g \leq G$, $1 \leq h \leq H$ (or $1 \leq h \leq H_d$ from Section 3.2.2) will not be written explicitly. Therefore,

the sums and products relating to rows, columns, row-clusters and column-clusters will be subscripted respectively by the letters i, j, g, and h, meaning that the sums and products will be written $\sum_i$, $\sum_j$, $\sum_g$ and $\sum_h$, and $\prod_i$, $\prod_j$, $\prod_g$ and $\prod_h$.

3. Methods

3.1. The BOS distribution for ordinal data

The binary ordinal search (BOS) model (Biernacki and Jacques (2016)) is a probability distribution for ordinal data parametrized by a position parameter $\mu \in \{1, ..., m\}$ and a precision parameter $\pi \in [0, 1]$. This distribution rises from the uniform distribution when $\pi = 0$ to a more peaked distribution around the mode μ when π increases, and reaches a Dirac distribution at the mode μ when $\pi = 1$. Figure 4 illustrates the shape of the BOS distribution with different values of μ and π. It is shown in Biernacki and Jacques (2016) that the BOS distribution is a polynomial function of π with degree $m-1$, whose coefficients depend on the position parameter μ. Therefore, $p(x_{ij}|\mu, \pi)$ can be written as:

$$p(x_{ij}|\mu, \pi) = \sum_{p=0}^{m-1} a_p(m, \mu, x_{ij})\pi^p.$$

For a univariate ordinal variable, the path in the stochastic binary search can be seen as a latent variable. Therefore, maximum likelihood estimation of model parameters can be performed simply using an EM algorithm (Dempster et al. (1977)).

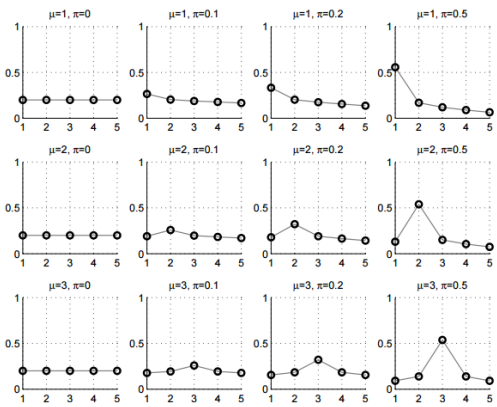

Fig. 4. BOS distribution $p(x; \mu, \pi)$: shape for $m = 5$ and for different values of μ and π.

3.2. Latent block model extension

In this section, the constrained latent block model is described after a brief summary of the latent block model concept (Govaert and Nadif (2013)).

3.2.1. Latent block model

Let $\boldsymbol{x}$ be a data matrix. It is assumed that there exists a partition $\boldsymbol{v}$ and a partition $\boldsymbol{w}$ such that each element x_{ij} is generated under a parameterized probability density function $f(x_{ij};\alpha_{gh})$, where g denotes the cluster of row i while h denotes the cluster of column j. The univariate random variables x_{ij} are assumed to be conditionally independent given the row and column partitions $\boldsymbol{v}$ and $\boldsymbol{w}$. Therefore, the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{v}$ and $\boldsymbol{w}$ can be expressed in the following form:

$$p(\boldsymbol{x}|\boldsymbol{v},\boldsymbol{w};\boldsymbol{\alpha}) \;=\; \prod_{i,j,g,h} f(x_{ij};\alpha_{gh})^{v_{ig}w_{jh}},$$

considering that $v_{ig}=1$ if i belongs to cluster g, whereas $v_{ig}=0$ otherwise, and that $w_{jh}=1$ when j belongs to cluster h, but $w_{jh}=0$ otherwise.

Different univariate distributions can be used depending on the type of data (e.g. Gaussian, Bernoulli, Poisson, etc.). In the present case, the BOS distribution has been chosen. The label for row i is called v_i and belongs to $\{1,...,G\}$. Similarly, the label for column j is called w_j and belongs to $\{1,...,H\}$. They are latent variables, and as is usual in latent variable theory, they are assumed to be independent (Everitt (1984)). So we have $p(\boldsymbol{v},\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\rho}) \;=\; p(\boldsymbol{v};\boldsymbol{\gamma})p(\boldsymbol{w};\boldsymbol{\rho})$ with:

$$p(\boldsymbol{v};\boldsymbol{\gamma}) \;=\; \prod_i p(v_i;\boldsymbol{\gamma}) \;=\; \prod_{i,g}\gamma_g^{v_{ig}} \text{ and } p(\boldsymbol{w};\boldsymbol{\rho}) \;=\; \prod_j p(w_j;\boldsymbol{\rho}) \;=\; \prod_{j,h}\rho_h^{w_{jh}},$$

knowing that $\gamma_g = p(v_{ig}=1)$ with $g\in\{1,...,G\}$ and $\rho_h = p(w_{jh}=1)$ with $h\in\{1,...,H\}$. This implies that, for all i, the distribution of v_i is the multinomial distribution $\mathcal{M}(\gamma_1,...,\gamma_G)$ and does not depend on i. Similarly for all j, the distribution of w_j is the multinomial distribution $\mathcal{M}(\rho_1,...,\rho_H)$ and does not depend on j. Based on these considerations, the parameter of the latent block model is defined as $\boldsymbol{\theta}=(\boldsymbol{\gamma},\boldsymbol{\rho},\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}=(\alpha_{gh})_{g,h}$, with $\alpha_{gh}=(\mu_{gh},\pi_{gh})$ being the position and precision BOS parameters of the distribution of block (g,h). Additionally, $\boldsymbol{\gamma}=(\gamma_1,...,\gamma_G)$ and $\boldsymbol{\rho}=(\rho_1,...,\rho_H)$ are the mixing proportions. Therefore, if V and W are the sets of all possible labels $\boldsymbol{v}$ and $\boldsymbol{w}$, the probability density function $p(\boldsymbol{x};\boldsymbol{\theta})$ of $\boldsymbol{x}$ can be written:

$$p(\boldsymbol{x};\boldsymbol{\theta}) \;=\; \sum_{(v,w)\in V\times W} \prod_{ig}\gamma_g^{v_{ig}} \prod_{jh}\rho_h^{w_{jh}} \prod_{i,j,g,h} f(x_{ij};\alpha_{gh})^{v_{ig}w_{jh}}.$$

3.2.2. Constrained latent block model

In this section, the latent block model is extended as in Robert (2017) so that the questions from different dimensions are kept separate, as are the questions with a different number of levels. Note that this extension can also be used in the case of ordinal data with different numbers of levels, by requiring variables with a different number of levels to be kept separate.

In this co-clustering framework, $\boldsymbol{x}$ is an $N\times(J_1+...+J_D)$ matrix, and it is seen as matrices $\boldsymbol{x}^1,...,\boldsymbol{x}^D$ stored side by side as explained in Section 2.3. It is assumed that there is a row-partition $\boldsymbol{v}$, and that for all $d\in\{1,...,D\}$ there exists a column-partition $\boldsymbol{w}^d$ such that each element x_{ij}^d is generated under a parameterized probability density function $f(x_{ij}^d;\alpha_{gh})$. Here, h denotes the cluster of column j, with $j\in\{1,...,J_d\}$ and

$h \in \{1, ..., H_d\}$. The univariate random variables x_{ij}^d are assumed to be conditionally independent given the row and column partitions $\boldsymbol{v}$ and $\boldsymbol{w^d}$. Therefore, the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{v}$ and $\boldsymbol{w} = (\boldsymbol{w}^d)_{d\in\{1,...,D\}}$ can be written in the following form:

$$p(\boldsymbol{x}|\boldsymbol{v}, \boldsymbol{w}; \boldsymbol{\alpha}) \;=\; \prod_{i,j,d} f(x_{ij}^d; \alpha_{v_i w_j^d}) \;=\; \prod_{d,i,j,g,h} f(x_{ij}^d; \alpha_{gh})^{v_{ig} w_{jh}^d},$$

knowing that: $\forall d \in \{1, ..., D\}$, $w_{jh}^d = 1$ when j belongs to cluster h, but $w_{jh}^d = 0$ otherwise.

The labels $v_1, ..., v_N, (w_1^d, ..., w_{J_d}^d)_{d\in\{1,...,D\}}$ are latent variables assumed to be independent: $p(\boldsymbol{v}, \boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\rho}) \;=\; p(\boldsymbol{v}; \boldsymbol{\gamma}) \prod_d p(\boldsymbol{w}^d; \boldsymbol{\rho}^d)$ with:

$$p(\boldsymbol{v}; \boldsymbol{\gamma}) \;=\; \prod_i p(v_i; \boldsymbol{\gamma}) \;=\; \prod_{i,g} \gamma_g^{v_{ig}} \text{ and } p(\boldsymbol{w}^d; \boldsymbol{\rho}^d) \;=\; \prod_j p(w_j^d; \boldsymbol{\rho}^d) \;=\; \prod_{j,h} \rho_h^{d\, w_{jh}^d},$$

knowing that $\rho_h^d \;=\; p(w_{jh} = 1)$ with $h \in \{1, ..., H\}$. Again, for all i, the distribution of v_i is the multinomial distribution $\mathcal{M}(\gamma_1, ..., \gamma_G)$ and does not depend on i. Equally for all j and for all d, the distribution of w_j^d is the multinomial distribution $\mathcal{M}(\rho_1^d, ..., \rho_{H_d}^d)$ and does not depend on j. By analogy with the classical latent block model, the probability density function $p(\boldsymbol{x}; \boldsymbol{\theta})$ is written:

$$p(\boldsymbol{x}; \boldsymbol{\theta}) \;=\; \sum_{(v, w^1, .., w^D) \in V \times W_1 \times ... \times W_D} \prod_{i,g} \gamma_g^{v_{ig}} \prod_{d,j,h} \rho_h^{d\, w_{jh}^d} \prod_{i,j,g,d,h} f(x_{ij}^d; \alpha_{gh})^{v_{ig} w_{jh}^d}.$$

3.3. Model inference with an SEM-Gibbs algorithm

This section details the model inference in the case of the constrained latent block model. The aim is to estimate $\boldsymbol{\theta}$ by maximizing the observed log-likelihood $l(\boldsymbol{\theta}; \check{\boldsymbol{x}}) = \sum_{\hat{x}} \log p(\boldsymbol{x}; \boldsymbol{\theta})$. In a co-clustering context, the EM algorithm is not computationally feasible (see Govaert and Nadif (2013)). Indeed, the E-step requires the calculation of the joint conditional probability of the missing labels $p(v_{ig} = 1, w_{jh} = 1|\boldsymbol{x}; \boldsymbol{\theta}^{(q)})$ for $1 \leq i \leq N$, $1 \leq g \leq G$, $1 \leq d \leq D$, $1 \leq j \leq J_d$, $1 \leq h \leq H_d$, with $\boldsymbol{\theta}^{(q)}$ the current value of the parameter. Therefore, this step involves computing $N \times G \times (J_1 \times H_1 + ... + J_D \times H_D)$ terms that cannot be factorized as for a standard mixture, due to the dependence of the row and column labels conditionally on the observations. There exist several alternatives to the EM algorithm, such as the variational EM algorithm (Govaert and Nadif (2005)), the SEM-Gibbs algorithm or Bayesian inference (Govaert and Nadif (2013)). The SEM-Gibbs algorithm is known to avoid spurious solutions (Keribin et al. (2010)), which is why it is used in this paper.

3.3.1. SEM-Gibbs algorithm

Starting from an initial value for the parameter $\boldsymbol{\theta}^{(0)}$, the qth iteration of the algorithm is composed of two steps.

SE-step The SE-step consists in simulating the latent variables according to their joint conditional probability using Gibbs sampling. Therefore, it repeats, for a given number of iterations, the generation of the row partitions conditionally on the column partitions and the generation of the column partitions conditionally on the row partitions. The generation of the row partitions $v_{ig}^{(q+1)} \mid \boldsymbol{x}, \boldsymbol{w}^{(q)}$ is performed according to:

$$p(v_{ig}^{(q+1)} = 1 \mid \boldsymbol{x}^d, \boldsymbol{w}^{(q)}; \boldsymbol{\theta}^{(q)}) \propto \gamma_g^{(q)} \times \prod_d t_g^d(\boldsymbol{x}_{i.}^d | \boldsymbol{w}^{d^{(q)}}; \boldsymbol{\alpha}^{(q)}),$$

where $t_g^d(\boldsymbol{x}_{i.}^d | \boldsymbol{w}^{d^{(q)}}; \boldsymbol{\alpha}^{(q)}) = \prod_{j,h} f(x_{ij}^d; \mu_{gh}^d, \pi_{gh}^d)^{w_{jh}^{d\ (q)}}$ with $\boldsymbol{x}_{i.}^d = (x_{ij}^d)_j$. The generation of the column partitions $w_{jh}^d \mid \boldsymbol{x}, \boldsymbol{v}^{(q+1)}$ for the dth table $\boldsymbol{x}^d$ ($d \in \{1, ..., D\}$) is performed according to:

$$p(w_{jh}^d = 1 \mid \boldsymbol{x}^d, \boldsymbol{v}^{(q+1)}; \boldsymbol{\theta}^{(q)}) \propto \rho_h^{d\,(q)} \times s_h^d(\boldsymbol{x}_{.j}^d | \boldsymbol{v}^{(q+1)}; \boldsymbol{\alpha}^{(q)}),$$

where $s_h^d(\boldsymbol{x}_{.j}^d | \boldsymbol{v}^{(q+1)}; \boldsymbol{\alpha}^{(q)}) = \prod_{i,g} f(x_{ij}^d; \mu_{gh}^d, \pi_{gh}^d)^{v_{ig}^{(q+1)}}$ with $\boldsymbol{x}_{.j}^d = (x_{ij}^d)_i$.

M-step The M-step consists in maximizing the completed log-likelihood by updating the co-cluster parameters according to the results of the last SE-step. It relies on the EM algorithm used in Biernacki and Jacques (2016) for the estimation of the BOS distribution on each block.

3.3.2. Imputation of missing values

The SEM algorithm is able to take into account the missing data and estimate them. It is assumed that the whole missing process is "missing at random" (see Little and Rubin (1986)). Firstly, the notation of $\boldsymbol{x}$ becomes $\boldsymbol{x}^{(q)}$ since the missing variables are going to be imputed. Then, a third step is added to the SE-step. For all $d \in \{1, .., D\}$, it generates the missing data $\hat{x}_{ij}^{d(q+1)} | \check{\boldsymbol{x}}^d, \boldsymbol{v}^{(q+1)}, \boldsymbol{w}^{d^{(q+1)}}$ as follows:

$$p(\hat{x}_{ij}^{d(q+1)} | \check{\boldsymbol{x}}^d, \boldsymbol{v}^{(q+1)}, \boldsymbol{w}^{d^{(q+1)}}; \boldsymbol{\theta}^{(q)}) = \prod_{g,h} p(\hat{x}_{ij}^{d(q+1)}; \mu_{gh}^{d\ (q)}, \pi_{gh}^{d\ (q)})^{v_{ig}^{(q+1)} w_{gh}^{d\ (q+1)}}.$$

3.3.3. Estimation of partitions and model parameters

The SEM algorithm repeats the aforementioned steps several times. The first iterations are called the burn-in period, which means the parameters are not yet stable. Consequently, the iterations that occur after this burn-in period are taken into account; they are called the sample distribution. The final estimate of the position parameter μ_{gh} is the mode of the sampling distribution. The final estimate of the continuous parameters $(\pi_{gh}^d, \boldsymbol{\gamma}_g, \boldsymbol{\rho}_h^d)_d$ is the median of the sample distribution. This corresponds to a final estimate of $\boldsymbol{\theta}$ that is called $\hat{\boldsymbol{\theta}}$. Next, a sample of $(\hat{\boldsymbol{x}}, \boldsymbol{v}, (\boldsymbol{w}^d)_d)$ is generated by an SE-step with $\boldsymbol{\theta}$ fixed to $\hat{\boldsymbol{\theta}}$. The final partitions $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}})$ and the missing observation $\hat{\boldsymbol{x}}$ are estimated by the mode of their sample distribution.

3.4. Model selection

To select the number of clusters, G in rows and $H_1, ..., H_D$ in columns, a model selection criterion must be used. The most classical, such as BIC (Schwarz (1978)) rely on penalizing the maximum log-likelihood value $l(\hat{\boldsymbol{\theta}}; \boldsymbol{x})$. However, due to the dependency of the observed data, this value is not available in a co-clustering context.

Alternatively, an approximation of the ICL information criterion (Biernacki et al. (2000)), referred to here as the ICL-BIC, can be invoked as it makes it possible to overcome the previous problem due to the dependency structure in $\check{\mathbf{x}}$. The key point is that this latter vanishes as the ICL relies on the complete latent block information $(\mathbf{v}, \mathbf{w})$, instead of integrating it out as is the case for BIC. In particular, Keribin et al. (2015) detail how to express the ICL-BIC for the general case of categorical data and Jacques and Biernacki (2018) for the specific case of ordinal data using the BOS model. In the present work, the ICL-BIC is therefore adapted for the constrained latent block model:

$$\begin{aligned}\text{ICL-BIC}(G, H_1, ..., H_D) &= \log p(\check{\boldsymbol{x}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}^1, \ldots, \hat{\boldsymbol{w}}^D; \hat{\boldsymbol{\theta}}) \\ &\quad - \frac{G-1}{2} \log N - \sum_d \frac{H_d - 1}{2} \log J_d - \sum_d \frac{G \times H_d}{2} \log(N \times J_d),\end{aligned}$$

where $\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}^1 ..., \hat{\boldsymbol{w}}^D$ are the row and column partitions discovered by the SEM algorithm, and $\hat{\boldsymbol{\theta}}$ is the corresponding estimated model parameter.

It should be noticed that co-clustering has to be performed for each possible value of G and H_d, $d \in \{1, ..., D\}$, then the result with the highest ICL-BIC retained. Let n_G be the number of candidate values for G, while n_{H_d} is the number of candidate values for H_d, $d \in \{1, ..., D\}$. Thus, the number of co-clustering processes to execute is $n_G \times n_{H_1} \times ... \times n_{H_D}$. As an example, if $D = 3$ and the user wants to try 3 values for G and for each H_d, then it would be necessary to execute $3^4 = 81$ co-clustering operations. Depending on the dataset, it might take too much time to find the best solution.

We propose the following heuristic search. Let us start by computing the ICL-BIC with minimum values $(G_{min}, H_{1_{min}}, ..., H_{D_{min}})$, then adding 1 to each number of clusters, step by step, and computing the ICL-BIC. We can then retain the best solution (the highest ICL-BIC) and continue the same process until the ICL-BIC stops increasing.

4. Application to the survey dataset

4.1. Constrained co-clustering with different dimensions

Several constrained co-clustering operations were performed on the dataset, with different dimensions and at different times. This section presents some significant results that were obtained. In the following experiments, the heuristic search described in Section 3.4 was executed with $G_{min} = 3$ and $H_{d_{min}} = 1$ to select the number of row-clusters and column-clusters (G and $(H_d), d \in \{1, ..., D\}$). All the ICL-BIC values are available in the appendix. The choice of a sufficient number of iterations for the SEM algorithm and for the burn-in period was made empirically. It was noticed that the parameters would stabilize after 150 iterations (or fewer). Therefore, the burn-in period was set to 400 iterations and the total number of iterations was fixed at 500. In the appendix, Figure A1 shows how some example parameters converge during iterations.

Table 2. Co-clustering result on anxiety, depression and symptom dimensions: estimated BOS parameters (μ_g

	Anxiety		Depression		Symptom	
	Col.-cluster 1	Col.-cluster 2	Col.-cluster 1	Col.-cluster 2	Col.-cluster 1	Col.-cl
Row-cluster 1	(2,0.77)	(2,0.77)	(1,0.70)	(2,0.83)	(2,0.46)	(1,0
Row-cluster 2	(2,0.68)	(3,0.72)	(2,0.47)	(2,0.79)	(3,0.39)	(1,0
Row-cluster 3	(1,0.64)	(2,0.44)	(1,0.77)	(2,0.70)	(2,0.58)	(1,0
Row-cluster 4	(1,0.67)	(2,0.47)	(1,0.79)	(2,0.71)	(1,0.80)	(1,0
Row-cluster 5	(2,0.72)	(3,0.55)	(2,0.64)	(2,0.75)	(2,0.66)	(1,0

4.1.1. Anxiety, depression and symptom.

As a first experiment, it was decided to investigate the responses that were given at time $T5$, at the end of the treatment. The questions regarding the symptoms of the treatment are interesting at this time because it marks the point at which the patients had been receiving chemotherapy for one year. Constrained co-clustering was performed by taking the questions related to the anxiety, depression and symptom dimensions. In this case, all the questions have a number of levels m equal to 4. Therefore, the only constraint is the separation of the questions that are related to different dimensions. The execution time of this set-up is about 12 seconds with a 2.00GHz Intel Xeon E5 2620 CPU and 8 Go of RAM. The result of the constrained co-clustering operation is illustrated by Figure 5. For all the figures, clusters are read from left to right and from top to bottom. Table 2 details the estimated BOS parameters (μ_{gh} and π_{gh}) for $g \in \{1, ..., G\}$ and $h \in \{1, ..., H_d\}, \forall d \in \{1, ..., D\}$.

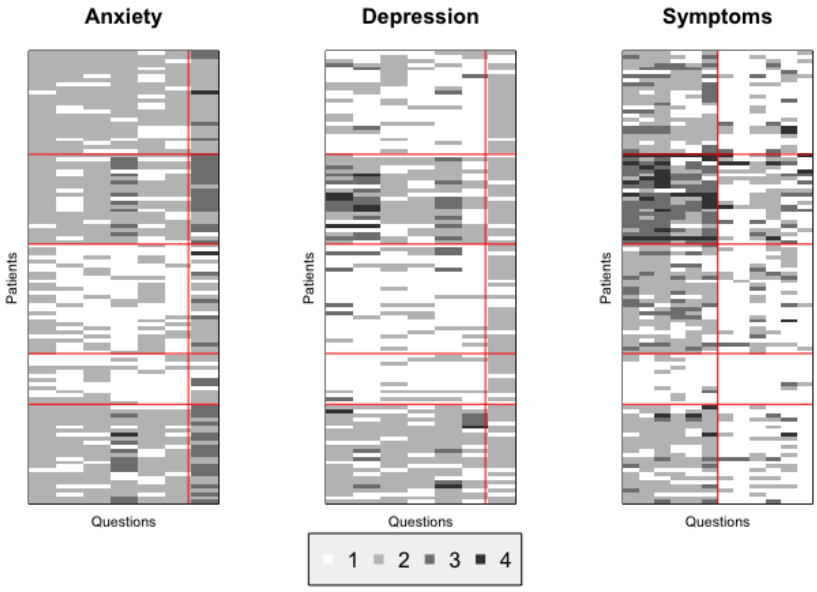

Fig. 5. Results of constrained co-clustering on anxiety, depression and symptom dimensions.

Five row-clusters are highlighted by the co-clustering results. Table 2 shows that the position parameters of the second row-cluster $(\mu_{2h})_{d,h}$ are generally greater than (or equal to) those of the other row-clusters. This means that the second group feels more anxiety

Table 3. Co-clustering result on social support dimensions: estimated BOS parameters (μ_{gh}, π_{gh}) for each cluster (g, h).

	Satisfaction	Availability		Intensity		Conflict		
	Col.-cluster 1	Col.-cluster 1	Col.-cluster 2	Col.-cluster 1	Col.-cluster 2	Col.-cluster 1	Col.-cluster 2	Col.-cluster 3
Row-cluster 1	(2,0.90)	(1,0.72)	(1,0.96)	(3,0.48)	(1,0.59)	(4,0.80)	(3,0.24)	(1,0.62)
Row-cluster 2	(3,0.87)	(1,0.64)	(1,0.50)	(3,0.46)	(2,0.48)	(4,0.47)	(3,0.42)	(1,0.49)
Row-cluster 3	(1,0.73)	(2,0.72)	(1,0.86)	(3,0.52)	(2,0.63)	(4,0.59)	(3,0.51)	(1,0.44)
Row-cluster 4	(2,0.79)	(1,0.61)	(2,0.64)	(3,0.31)	(2,0.50)	(3,0.27)	(3,0.32)	(1,0.44)
Row-cluster 5	(1,0.93)	(1,0.78)	(1,0.91)	(3,0.27)	(1,0.68)	(4,0.71)	(3,0.18)	(1,0.63)

and depression, and feels the disease symptoms more intensively than the other groups. It can also be seen that the fourth row-cluster is less inclined to anxiety and depression and suffers less from the symptoms than the others groups; indeed, parameters $(\mu_{4h})_{d,h}$ are generally the lowest. Furthermore, the precision parameters $(\pi_{4h})_{d,h}$ are quite high for this row-cluster, which means that the answers show limited spread around the position $(\mu_{4h})_{d,h}$. By observing the results for these two groups, it is possible to establish that the degree to which symptoms are felt is closely linked to signs of anxiety and depression, which is a fairly logical and intuitive result. However, the first, third and fifth groups provide more information. They are effectively very similar in terms of the degree to which they suffer the disease symptom dimensions. However, they differ a great deal in the first column-cluster of anxiety, and the first of depression. This means that even if a link between symptoms, anxiety and depression can be deduced from initial observations, it is not fully confirmed when people do not experience the symptoms at the extremes ("Very much" or "Not much"). The column-clusters offer interesting results as well: there is a clear separation between the symptoms. By examining the questions in each cluster, it becomes clear that the questions in the first cluster exclusively deal with pain and fatigue, while the second cluster deals with other symptoms such as nausea or loss of appetite. The co-clustering operation therefore detected two sub-dimensions for the symptoms dimension. Furthermore, there is a big difference in how the patients feel these two clusters: it is clearly noticeable that the position parameters $(\mu_{g1})_{(symptoms)}$ are generally higher than $(\mu_{g2})_{(symptoms)}$. Therefore, all the patients in general suffer more from pain and fatigue than the other symptoms.

4.1.2. Social support: satisfaction, availability, intensity and conflict.

As a second experiment, the questions related to social support were used. The responses are taken from the fourth stage of the survey. This is in the middle of the patients' treatment, so they have already experienced a great deal, but know that they have to keep going for a few more months. Their perception of the social support they receive is therefore interesting at this point. This aspect includes questions relating to four dimensions: satisfaction (where the number of levels $m = 6$), perception of availability, intensity and conflict (where the number of levels $m = 4$). The questions that relate to the same dimension have the same number of levels. Again, the only constraint is the separation of the questions that are not related to the same dimensions. The result of the constrained co-clustering operation is illustrated by Figure 6. Furthermore, Table 3 details the estimated BOS parameters (μ and π) for each co-cluster.

The co-clustering operation detected five row-clusters. The third and the fifth are clearly satisfied with the social support they receive. Indeed, their position parameters

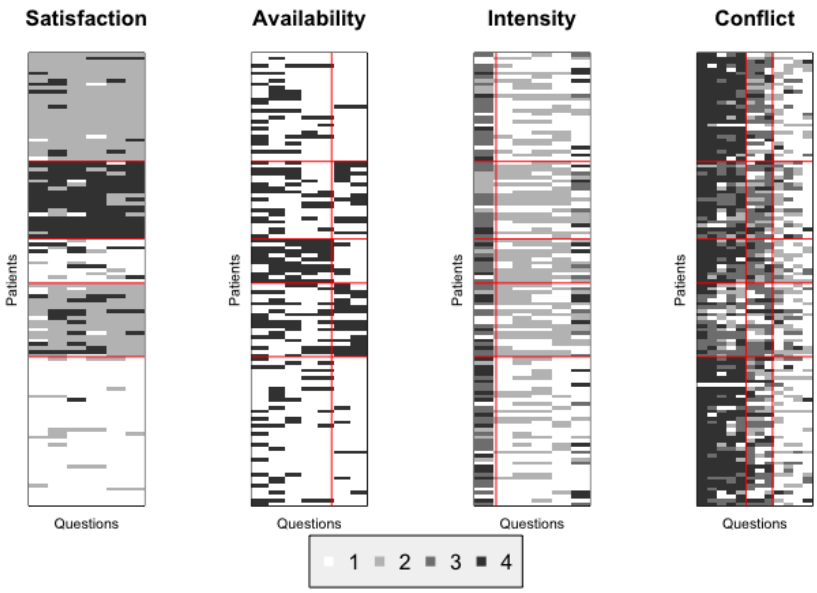

Fig. 6. Result of constrained co-clustering on dimensions related to social support.

$\mu_{31(satisfaction)}$ and $\mu_{51(satisfaction)}$ are equal to 1. Furthermore, the precision parameters $\pi_{31(satisfaction)}$ and $\pi_{51(satisfaction)}$ are very high, which means that most of the patients effectively gave the most positive response to the questions regarding their satisfaction. In contrast, the women in the first group are quite dissatisfied with their social support compared to the others. Another result is that the third group, one of the most satisfied, has one of the worst levels of perception of availability among their close family and friends ($\mu_{31(availability)} \geq \mu_{g1(availability)}$). Furthermore, it is also interesting to observe the column-clusters that were detected by the co-clustering operation for the conflict dimension. The first group of questions is about the effort the patient has to make to avoid conflict with their loved ones. The second group comprises questions about changes in relationships, while the last cluster concerns feelings of anger towards close family and friends.

4.1.3. Symptoms at different times

In this experiment, the questions related to symptoms were selected at different stages (at times $T0$, $T2$ and $T5$). The constraint is therefore not to separate the questions from different dimensions, but to separate the questions that are from different stages. The point of performing co-clustering on such a dataset is that the row-clusters group together individuals who evolved in a similar way regarding this dimension. Furthermore, the column-clusters provide information about how the patients' symptoms generally worsened (or improved) throughout the treatment. BOS parameters for this experiment are available in Table 4, and Figure 7 illustrates the results.

The co-clustering operation highlights three row-clusters. The third groups together

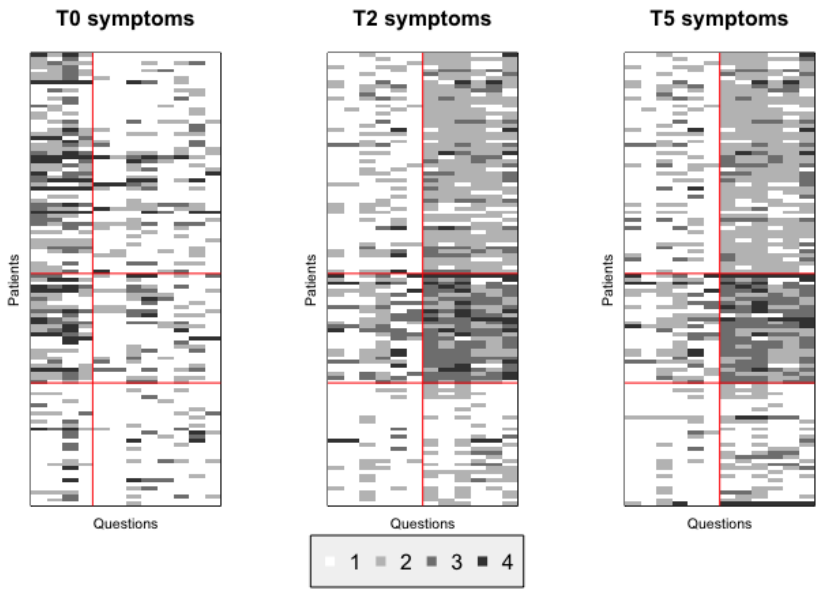

Fig. 7. Co-clustering results for questions related to symptoms, at three different times.

Table 4. Co-clustering results for the symptoms dimension, at three different times: estimated BOS parameters $(\mu_{gh}, \pi$ each co-cluster (g, h).

	T0 Symptoms		T2 Symptoms		T25 Symptoms	
	Col.-cluster 1	Col.-cluster 2	Col.-cluster 1	Col.-cluster 2	Col.-cluster 1	Col.-cluster 2
Row-cluster 1	(2,0.20)	(1,0.67)	(2,0.62)	(1,0.72)	(2,0.64)	(1,0.74)
Row-cluster 2	(2,0.09)	(1,0.62)	(3,0.43)	(1,0.40)	(3,0.42)	(1,0.46)
Row-cluster 3	(1,0.66)	(1,0.84)	(1,0.58)	(1,0.85)	(1,0.54)	(1,0.84)

people who felt the symptoms of the disease to a lesser degree than the others: the position parameters $(\mu_{3h})_{d,h}$ are all equal to 1. Furthermore, the precision parameters $(\pi_{3h})_{d,h}$ are fairly high, which implies that the responses show limited spread around the value 1. It is also interesting to investigate how the column-clusters evolve. To begin with, for each time, the symptoms are separated into two column-clusters: the first is systematically worse overall than the second, because $(\mu_{g1})_{(T0,T2,T5)} \geq (\mu_{g2})_{(T0,T2,T5)}, \forall g \in \{1, ..., G\}$. It is observed that at time $T0$ there are fewer symptoms in column-cluster 1 than in column-cluster 2, whereas they are equally shared at times $T2$ and $T5$.

4.2. Handling of the dynamical aspect of the data

The patients answered the same questionnaires at six different stages of their treatment and the way in which the responses change is clearly of interest. Defining a model to study this evolution is essential, but it is not the purpose of this paper and will be covered elsewhere. Here, we focus on providing a tool that allows the psychologists to visualize the evolution of the row-clusters without establishing the mathematical reasoning. To this end, visualizations were created to provide the psychologists could have a first impression of this evolution, using `Javascript` library `D3js`. First of all, the

dimensions of questionnaire EORTC QLQ-C30 were selected. Then, co-clustering was performed, using a similar method to that in Section 4.1, for each time $T0$, $T1$, $T2$, $T3$, $T4$ and $T5$. The visualization shows the row-clusters on the y-axis, and the timeline on the x-axis: Figure 8 illustrates the home page of a visualization that was created with the dimensions dealing with quality of life and emotional state. If the expert wishes to observe the evolution of a single patient, they can click on the list of patients on the right to see the row-clusters she belongs to through time, as shown in Figure 9. In addition, if the expert wants to know the co-cluster BOS parameters, they can click on the row-clusters, and read the (μ, π) of the corresponding co-clusters, as shown in Figure 10.

These visualizations showed that, overall, the patients became more stable with time. Indeed, it can be observed that whereas many patients change row-clusters at the three first stages $T0$, $T1$ and $T2$, these transitions get rarer after time $T2$.

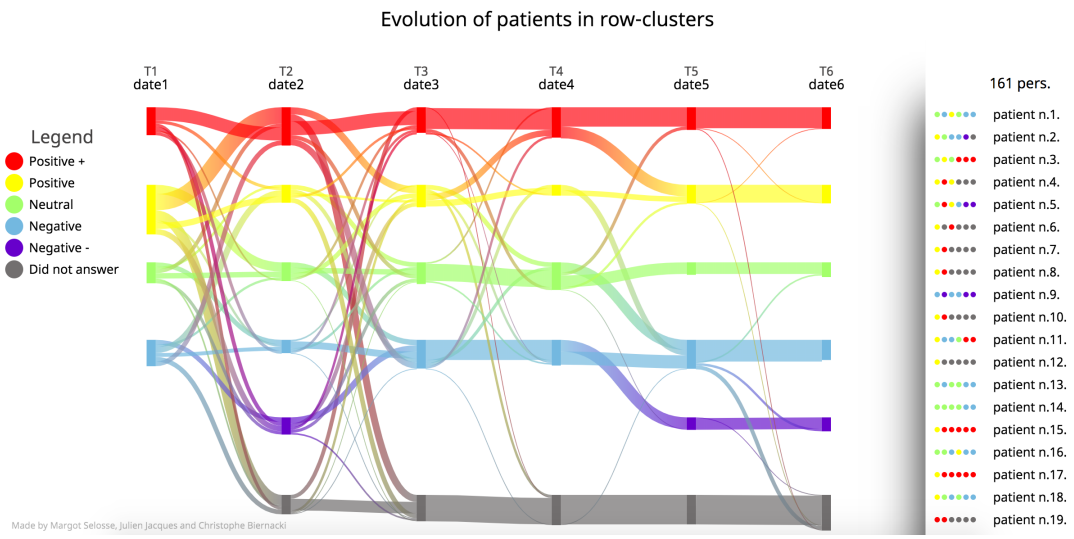

Fig. 8. Home page: the row-clusters shown on the y-axis, and the timeline on the x-axis.

5. Conclusion

In this paper, a co-clustering algorithm is proposed to analyse psychological questionnaires given to women affected by breast cancer. This dataset has many specific features, which makes it difficult to use classical techniques without changing the information. Firstly, it comprises questionnaires with answers on an ordinal scale. In addition, it includes a temporal aspect because the patients answered these questionnaires six times. The questions are also linked to psychological dimensions, which cannot be ignored. Finally, just like many real datasets, this one contains some missing values.

To adapt to the particularities of the survey, an extension of the latent block model is defined, and the parsimonious BOS distribution for ordinal data is employed. The reader should be aware that the classic latent block model is symmetric in nature. It means

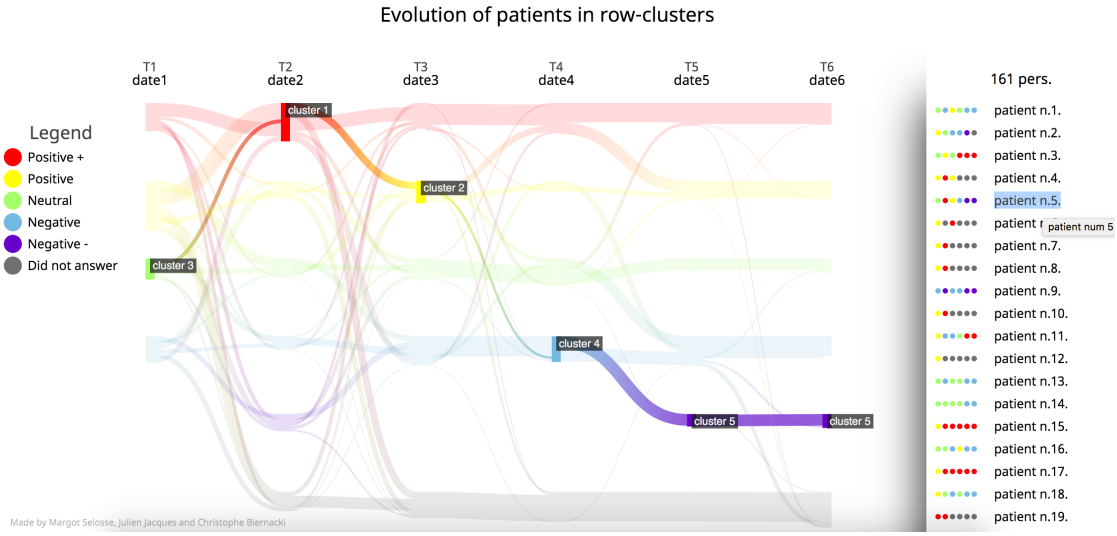

Fig. 9. When the user clicks on a patient in he right-hand list, they can observe the psychological trajectory of this patient.

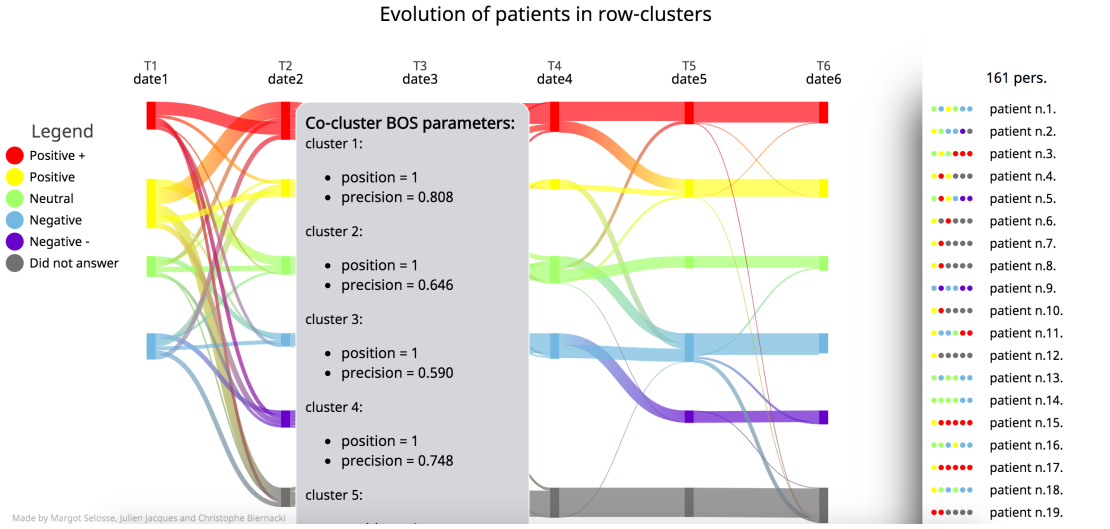

Fig. 10. When the user clicks on a row-cluster they are able to see the BOS parameters of all corresponding co-clusters.

that the role of observations can be interchanged with the role of the variables. In this work, the hypotheses added break this symmetry. The model inference is performed with an SEM-Gibbs algorithm, which makes it possible to take missing values into account. An R package called *ordinalClust* with the full implementation of this work is available on CRAN. Lastly, visualizations have been created to help psychologists observe the evolution of their patients.

The results were particularly satisfying for the psychologists. The proposed technique provides a parsimonious way to cluster the patients by placing the questions in a small number of groups, and the meaningfulness of the BOS parameters makes it easy to interpret the resulting co-cluster. Furthermore, the constrained co-clustering method resolves two issues: the different numbers of levels for the questions, and the fact that the questions refer to different psychological dimensions.

With the proposed approach, features with different numbers of levels can be processed using the co-clustering method, without allowing them to be part of the same column-cluster. In future work, it would be interesting to examine that possibility, and ideally to perform co-clustering operations with data of different kinds (continuous, functional, etc.). Finally, although the dynamical aspect of the data has been approached with visualizations, it would be advantageous to define a mathematical model in this respect.

6. Acknowledgements

We would like to thank INCa (Institut National du Cancer), Institut Lilly, Institut Bergonié, Centre Régional de Lutte Contre le Cancer de Bordeaux (C. Tunon de Lara, J. Delefortrie, A. Rousovoal, A. Avril and E. Bussières) and Laboratoire de Psychologie de l'Université de Bordeaux (C. Quintrinc and S. de Castro-Lévèque).

References

Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. d., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M. and Takeda, F. (1993) The european organization for research and treatment of cancer qlq-c30: A quality-of-life instrument for use in international clinical trials in oncology. *JNCI: Journal of the National Cancer Institute*, **85**, 365–376.

Agresti, A. (2010) *Analysis of Ordinal Categorical Data, 2nd Ed.* John Wiley & Sons, Inc.

Annema, C., Roodbol, P. F., Van den Heuvel, E. R., Metselaar, H. J., Van Hoek, B., Porte, R. J. and Ranchor, A. V. (2017) Trajectories of anxiety and depression in liver transplant candidates during the waiting-list period. *British Journal of Health Psychology*, **22**, 481–501.

Biernacki, C., Celeux, G. and Govaert, G. (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.*, **22**, 719–725.

Biernacki, C. and Jacques, J. (2016) Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm. *Statistics and Computing*, **26**, 929–943.

Corduas, M. (2008) A statistical procedure for clustering ordinal data. *Quaderni di statistica*, **10**, 177–189.

Cousson-Gélie, F. (2014) Évolution du contrôle religieux la première année suivant l'annonce d'un cancer du sein : quels liens avec les stratégies de coping, l'anxiété, la dépression et la qualité de vie ? *Psychologie Francaise*, **59**, 331 – 341.

D'Elia, A. and Piccolo, D. (2005) A mixture model for preferences data analysis. *Computational Statistics & Data Analysis*, **49**, 917–934.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, series B*, **39**, 1–38.

Everitt, B. S. (1984) *Introduction to Latent Variable Models.* Chapman and Hall.

Giordan, M. and Diana, G. (2011) A clustering method for categorical ordinal data. *Communications in Statistics - Theory and Methods*, **40**, 1315–1334.

Govaert, G. and Nadif, M. (2003) Clustering with block mixture models. *Pattern Recognition*, **36**, 463–473.

Govaert, G. and Nadif, M. (2005) An em algorithm for the block mixture model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 643–647.

Govaert, G. and Nadif, M. (2013) *Co-Clustering: models, algorithms and applications.* Computing Engineering series. ISTE-Wiley.

Jacques, J. and Biernacki, C. (2018) Model-Based Co-clustering for Ordinal Data. *Computational Statistics and Data Analysis*, **123**, 101–115.

Jollois, F.-X. and Nadif, M. (2009) Classification de données ordinales : modèles et algorithmes. In *41èmes Journées de Statistique, SFdS, Bordeaux*. Bordeaux, France.

Kaufman, L. and Rousseeuw, P. J. (2008) *Introduction*, 1–67. John Wiley and Sons, Inc.

Keribin, C., Brault, V., Celeux, G. and Govaert, G. (2015) Estimation and selection for the latent block model on categorical data. *Statistics and Computing*, **25**, 1201–1216.

Keribin, C., Govaert, G. and Celeux, G. (2010) Estimation d'un modèle à blocs latents par l'algorithme SEM. In *42èmes Journées de Statistique*. Marseille, France.

Lewis, S. J. G., Foltynie, T., Blackwell, A. D., Robbins, T. W., Owen, A. M. and Barker, R. A. (2005) Heterogeneity of parkinson's disease in the early clinical stages using a data driven approach. *Journal of Neurology, Neurosurgery & Psychiatry*, **76**, 343–348.

Little, R. J. A. and Rubin, D. B. (1986) *Statistical Analysis with Missing Data*. New York, NY, USA: John Wiley & Sons, Inc.

MaloneBeach, E. E. and Zarit, S. H. (1995) Dimensions of social support and social conflict as predictors of caregiver depression. *International Psychogeriatrics*, **7**, 2538.

McParland, D. and Gormley, I. C. (2011) Clustering ordinal data via latent variable models. *Berthold Lausen, Dirk Van den Poel, Alfred Ultsch (eds.). Algorithms from and for Nature and Life : Classification and Data Analysis.*

Pierce, G. R., Sarason, I. G., Sarason, B. R., Solky-Butzel, J. A. and Nagle, L. C. (1997) Assessing the quality of personal relationships. *Journal of Social and Personal Relationships*, **14**, 339–356.

Ranalli, M. and Rocci, R. (2016) Mixture models for ordinal data: A pairwise likelihood approach. *Statistics and Computing*, **26**, 529–547.

Robert, V. (2017) *Classification croisée pour l'analyse de bases de données de grandes dimensions de pharmacovigilance*. Ph.D. thesis. Thèse de doctorat dirigée par Celeux, Gilles Mathématiques appliquées Paris Saclay 2017.

Sarason, I. G., Levine, H. M., Basham, R. B. and Sarason, B. R. (1983) Assessing social support: The social support questionnaire. *Journal of Personality and Social Psychology*, 139.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Vermunt, J. and Magidson, J. (2005) Latent gold 4.0 user's guide. belmont, massachusetts:statistical innovations inc.

Wagstaff, K., Cardie, C., Rogers, S. and Schrödl, S. (2001) Constrained k-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, 577–584. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. URL: `http://dl.acm.org/citation.cfm?id=645530.655669`.

Zigmond, A. S. and Snaith, R. P. (1983) The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, **67**, 361–370.

Table 1. ICL-BIC values for experiment with anxiety, depression and symptom dimensions.

Iteration number	Tested set	ICL-BIC value
0	**3111**	**-3182.636**
1	3211	-3202.003
	4111	-3178.63
	3121	-3135.95
	3112	**-2981.384**
2	4112	-2955.255
	3212	-3012.662
	3122	**-2931.426**
	3113	-3003.892
3	4122	-2909.457
	3222	**-2907.208**
	3132	-2937.025
	3123	-2941.77

Iteration number	Tested set	ICL-BIC value
4	**4222**	**-2897.574**
	3322	-2899.502
	3232	-2915.243
	3223	-2914.43
5	**5222**	**-2887.428**
	4322	-2901.103
	4232	-2902.765
	4223	-2911.143
6	6222	-2890.224
	5322	-2890.043
	5232	-2898.423
	5223	-2900.492

Table 2. ICL-BIC values for experiment with social support dimensions.

Iteration number	Tested set	ICL-BIC value
0	**31111**	**-4159.044**
1	41111	-4148.478
	32111	-4169.625
	31211	-4167.269
	31121	-4109.996
	31112	**-3890.939**
2	41112	-3861.417
	32112	-3901.316
	31212	-3966.906
	31122	-3847.206
	31113	**-3792.995**
3	**41113**	**-3759.687**
	32113	-3800.504
	31213	-3793.978
	31123	-3760.164
	51113	-3803.808

Iteration number	Tested set	ICL-BIC value
4	51113	-3745.543
	42113	-3792.785
	41213	-3807.943
	41123	**-3708.452**
	41114	-3782.372
5	**51123**	**-3684.643**
	42123	-3723.946
	41223	-3710.339
	41133	-3814.453
	41124	-3807.466
6	61123	-3672.869
	52123	-3689.939
	51223	**-3646.392**
	51133	-3670.815
	51124	-3674.937

Iteration number	Tested set	ICL-B
7	61223	-365
	52223	-366
	51323	-366
	51233	-367
	51224	-366

Appendix

The following tables present the ICL-BIC obtained by executing the heuristic search described in Section 3.4 for the applications described in Section 4. At each iteration, the values in bold represent the highest ICL-BIC values of the iteration. The underlined values are the final chosen values for $(G, H_1, ..., H_D)$.

Figure A1 presents the evolution of some parameters through the SEM algorithm iterations.

Table 3. ICL-BIC values for experiment with symptom dimensions at times ($T0$, $T1$, $T2$).

Iteration number	Tested set	ICL-BIC value
0	**3111**	**-4479.951**
1	4111	-4473.039
	3211	-4419.312
	3121	-4278.773
	3112	**-4259.51**
2	4112	-4206.192
	3212	-4200.733
	3122	**-4021.611**
	3113	-4262.249
3	4122	-4033.372
	3222	**-3967.913**
	3132	-4012.178
	3123	-4103.417
4	4222	-3981.241
	3322	-4082.588
	3232	-4080.828
	3223	-4046.097

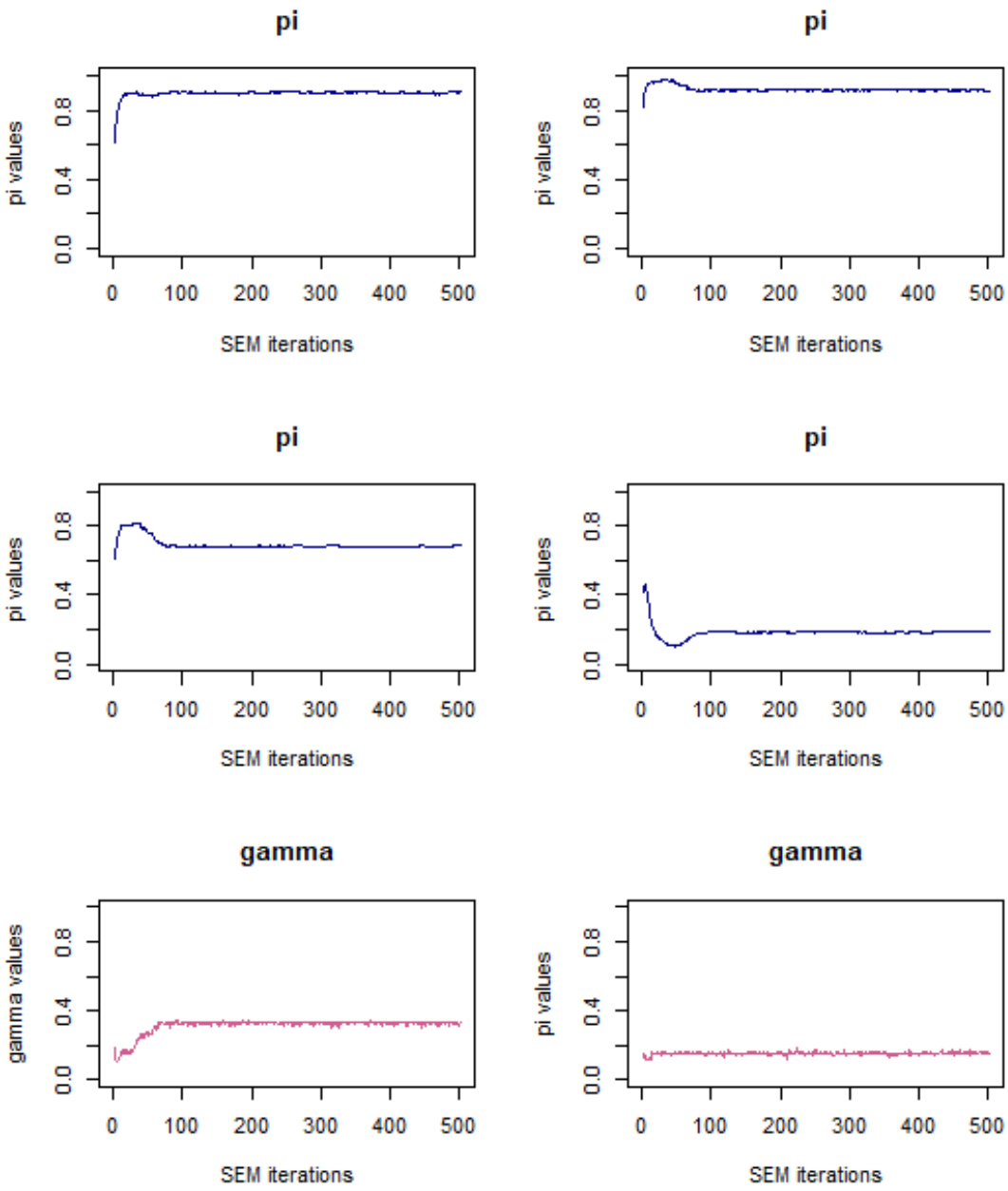

Fig. A1. Evolution of parameters over time in the SEM algorithm.