ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2017)
66, *Part* 3, *pp.* 555–580

Bayesian modelling of networks in complex business intelligence problems

Daniele Durante, Sally Paganin and Bruno Scarpa

University of Padova, Italy

and David B. Dunson

Duke University, Durham, USA

[Received October 2015. Final revision May 2016]

Summary. Complex network data problems are increasingly common in many fields of application. Our motivation is drawn from strategic marketing studies monitoring customer choices of specific products, along with co-subscription networks encoding multiple-purchasing behaviour. Data are available for several agencies within the same insurance company, and our goal is to exploit co-subscription networks efficiently to inform targeted advertising of cross-sell strategies to currently monoproduct customers. We address this goal by developing a Bayesian hierarchical model, which clusters agencies according to common monoproduct customer choices and co-subscription networks. Within each cluster, we efficiently model customer behaviour via a cluster-dependent mixture of latent eigenmodels. This formulation provides key information on monoproduct customer choices and multiple-purchasing behaviour within each cluster, informing targeted cross-sell strategies. We develop simple algorithms for tractable inference and assess performance in simulations and an application to business intelligence.

Keywords: Business intelligence; Chinese restaurant process; Co-subscription networks; Cross-sell marketing strategies; Mixture of latent eigenmodels; Monoproduct choices

1. Introduction

Increasing business competition and market saturation have led companies to shift the focus of their marketing strategies from the acquisition of new customers to an increased penetration of their customer base. Targeting existing customers via cross-sell campaigns instead of attracting new customers provides a more effective strategy for the growth of the company and enhances customer retention by increasing the switching costs (Kamakura *et al.*, 1991). Therefore, monoproduct customers purchasing a single product from a company represent a key segment of the customer base, and companies are naturally interested in expanding these customers' purchases to additional products.

Business statistics currently offer a wide set of procedures for cross-sell campaigns relying on shared acquisition patterns of products, based on customer ownership data. A first effort in addressing this aim can be found in the latent trait model that was developed by Kamakura *et al.* (1991) to estimate the propensity of a customer towards a particular product, based on its ownership of other products. This procedure was later improved by Kamakura *et al.* (2003) combining information from customer databases with survey data. Verhoef and Donkers

Address for correspondence: Daniele Durante, Department of Statistical Sciences, University of Padova, Via Cesare Battisti 241, 35121 Padova, Italy.
E-mail: durante@stat.unipd.it

 0035–9254/17/66555

(2001) focused instead on predicting the potential value of a current customer via a multivariate probit model and proposed a 2×2 segmentation to improve targeting. Finally, Thuring (2012) developed a multivariate credibility method to identify a profitable set of customers for cross-selling. This is accomplished by estimating a latent risk profile for each customer, exploiting information on claims. Refer also to Thuring *et al.* (2012) and Kaishev *et al.* (2013) for recently developed cross-sell strategies and for an overview on available methodologies.

Previous procedures exploit different sources, including customer demographics and survey data, to estimate co-subscription probabilities between pairs of products for each customer in a single agency. Differently from this setting, we do not observe customer demographic data for a single agency but monitor monoproduct customer choices along with co-subscription networks among $V = 15$ products for $n = 130$ agencies operating in the Italian insurance market. Customer relationship management is becoming increasingly important to operate effectively in the insurance market. This sector is mostly stable in developed countries, and rising customer expectations, along with tight competition between top corporations and low growth potentials, force companies to exploit their databases efficiently to create, manage and maintain their portfolio of profitable customers (Matiş and Ilieş, 2014).

We observe choice data $y_{is} \in \{1, \ldots, V\}$ denoting the product that is subscribed to by monoproduct customer $s = 1, \ldots, n_i$, within agency $i = 1, \ldots, n$. The co-subscription network for each agency $i = 1, \ldots, n$ is available via a $V \times V$ symmetric adjacency matrix A_i, with $A_{i[vu]} = A_{i[uv]} = 1$ if—in agency i—the number of customers subscribing to both products v and u exceeds 10% of the total number of multiproduct customers subscribed to at least one of the two, and $A_{i[vu]} = A_{i[uv]} = 0$ otherwise, for every $v = 2, \ldots, V$ and $u = 1, \ldots, v - 1$. The presence of an edge between two products suggests a preference of customers in agency i for that specific pair, controlling for the total number of multiproduct customers subscribed to at least one of the two products. Refer to Fig. 1 for an example of the available data.

Although the 10% threshold can be adapted to the company requirements, we found $A_1, \ldots, A_{130}$ robust to moderate changes in this value. In fact, the adjacency matrices constructed with a 7.5% threshold differ from those based on the 10% threshold for only 6% of the edges. This gap remains at similarly low values 4% and 8% when the threshold changes from 10% to 12.5% and to 15% respectively.

Each agency can define appropriate cross-sell strategies by exploiting its co-subscription network A_i to estimate the propensity of a customer who subscribed to product $v = 1, \ldots, V$ additionally to buy $u \neq v$. This leads to V different cross-sell strategies $u_{i1}, \ldots, u_{iV}$, with u_{iv} defining which additional product $u \neq v$ is the best offer to currently monoproduct customers subscribed to v in agency i. Hence, $u_{iv} = \arg\max_u \{\text{pr}(\mathcal{A}_{i[vu]} = 1) : u \neq v\}$, with $\mathcal{A}_{i[vu]}$ the random variable characterizing the presence or absence of an edge between products v and u in agency i. Efficiently targeting advertising by offering customers the product mostly complementary to their current choice can substantially improve performance relative to untargeted advertising, while increasing the satisfaction and reducing churn effects due to frequent and pointless cross-selling attempts (Azzalini and Scarpa, 2012; Kamakura *et al.*, 2003). Satisfied customers are a key to enhancing positive word-of-mouth communication and are less sensitive to competing brands and price (Matiş and Ilieş, 2014).

The effectiveness at increasing the number of multiproduct customers in agency i depends not only on the propensity of customers with v to subscribe to u also, measured by $\text{pr}(\mathcal{A}_{i[vu]} = 1)$, but also on the proportion of monoproduct customers with v, defined by $p_{iv} = \text{pr}(\mathcal{Y}_i = v)$, with $\mathcal{Y}_i$ the random variable denoting the choices of monoproduct customers in agency i. If p_{iv} is low, then strategy u_{iv} targets a small portion of the customer base of agency i, and hence has a low ceiling on effectiveness. To take into account the role of p_{iv}, we associate

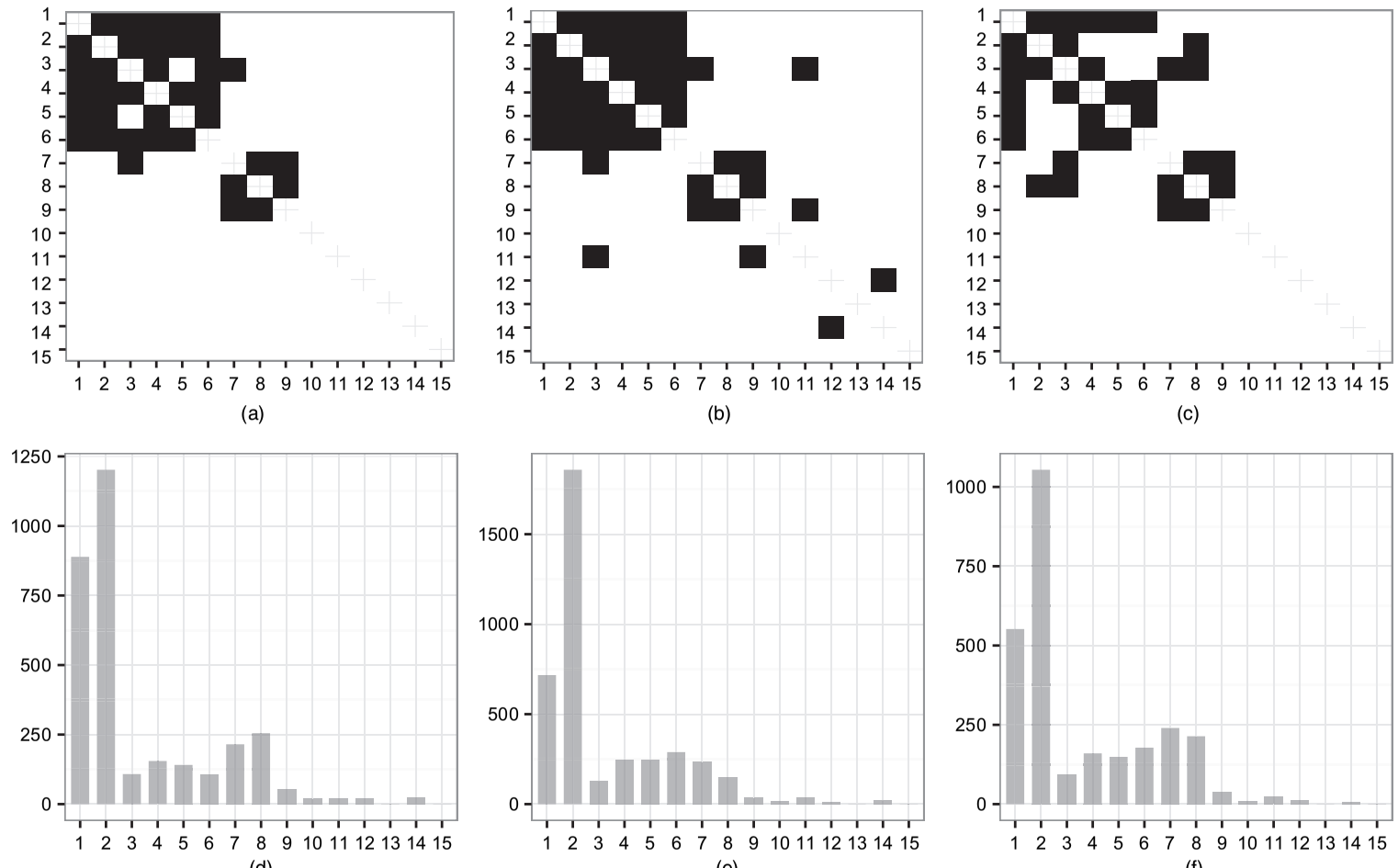

Fig. 1. Available data for three selected agencies: observed co-subscription networks for agencies (a) 6, (b) 7 and (c) 8 (■, edge; □, non-edge); total number of monoproduct customers for each product $v = 1, \ldots, 15$ based on choice data y_{is}, $s = 1, \ldots, n_i$, for agencies (d) 6, (e) 7 and (f) 8

each strategy u_{iv} with a performance indicator $e_{iv} = p_{iv} \max\{\text{pr}(\mathcal{A}_{i[vu]} = 1) : u \neq v\}$, for each $v = 1, \ldots, V$ and $i = 1, \ldots, n$. Strategies with a high e_{iv} will target a sizable proportion of the available monoproduct customers in agency i with advertising for a new product that is likely to be appealing to them.

In defining and evaluating cross-sell strategies, there are two important issues to take into consideration. Firstly, we are faced with statistical error in estimating the components underlying strategies u_{iv} and indicators e_{iv}, for each $v = 1, \ldots, V$ and $i = 1, \ldots, n$. This is a particular problem in estimating $\text{pr}(\mathcal{A}_{i[vu]} = 1)$ due to data sparsity. The second issue is that it is important to take into account the fact that administrative overheads can be reduced by considering the same strategy for different agencies within the same company. For groups of agencies having sufficiently similar customer bases, an identical strategy can be adopted to reduce administrative costs without decreasing effectiveness. Motivated by this notion, we propose to address both the statistical error and the administrative overheads issue through clustering of agencies according to the parameters characterizing their customer bases, and then administrating the same strategy to all agencies within a cluster.

As suggested by Fig. 1, it is reasonable to expect agencies operating in similar markets to exhibit clusters, corresponding to common patterns in the composition of their monoproduct customer choices and co-subscription networks. Efficient detection of such clusters allows adaptive reduction of the total number of strategies to be devised from $u_{i1}, \ldots, u_{iV}$, $i = 1, \ldots, n$, to $u_{k1}, \ldots, u_{kV}$, $k = 1, \ldots, K < n$, with each cluster-specific set of strategies $u_{k1}, \ldots, u_{kV}$ maintaining its effectiveness in targeting similar agencies. This higher level targeting and profiling represents a key to balance the need of the company to reduce costs and the importance of providing agencies with effective strategies that account for their specific structure. Providing agencies with sets of strategies that are suitably related to their structure is further important in increasing trust in the company and improving synergy.

We address this goal by developing a Bayesian hierarchical model, which adaptively associates shared sets of strategies $u_{k1}, \ldots, u_{kV}$ and performance indicators $e_{k1}, \ldots, e_{kV}$ to clusters of agencies characterized by monoproduct and multiproduct customers with similar purchasing behaviour. Each cluster-specific set of cross-sell strategies $u_{k1}, \ldots, u_{kV}$ is devised by learning cluster-specific co-subscription patterns between pairs of products, exploiting data on multiproduct customers. Joining this information with the estimated cluster-specific distribution of monoproduct customers across products, performance indicators $e_{k1}, \ldots, e_{kV}$ are constructed. To our knowledge, this is the first approach in the literature that considers a two-level cross-sell segmentation of the customer base, which clusters agencies with similar customer preferences, and profiles monoproduct and multiproduct purchasing behaviour within each cluster to define cross-sell strategies and related performance indicators.

1.1. Joint modelling of mixed domain data

As a step towards our goal of designing efficient cross-sell strategies, we first develop a joint model for the data $\{\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i}), A_i\}$, $i = 1, \ldots, n$, which characterizes the distribution across agencies of the monoproduct customer subscriptions along with the co-subscription network for multiproduct customers. The model is chosen to be flexible while automatically clustering different agencies that have similar monoproduct customer choices and co-subscription network profiles. This clustering facilitates borrowing of information across agencies and therefore improves efficiency in learning the joint distribution of the monoproduct and multiproduct purchasing behaviour of the customer bases. In addition, clustering provides a simplification, which is useful in the design of interpretable strategies.

There is an increasing statistical literature on joint modelling and coclustering of mixed domain data. Available procedures focus on learning the association between a univariate response variable and an object predictor, which is typically characterized by a function. Bigelow and Dunson (2009) favoured clustering among predictor trajectories, with each cluster associated with a specific offset in a generalized linear model for the response variable. Although providing an appealing procedure for interpretability and inference, their model lacks flexibility in constraining predictor and response clusters to be the same. This may require the introduction of many clusters to characterize the joint distribution of the mixed domain data appropriately, reducing the performance in estimating cluster-specific components and providing a biased overview of the underlying clustering structure.

Dunson *et al*. (2008) addressed the previous issue by modelling the conditional distribution of the response via a cluster-dependent mixture representation, rather than considering only a cluster-specific offset in the conditional expectation. In improving flexibility via dependent mixture modelling, their procedure can better identify the underlying clustering structure; see also Banerjee *et al*. (2013) for a recent overview of this topic. Although we are similar to the previous methods in looking for flexible and accurate joint modelling and coclustering procedures for mixed domain data, our motivating data set is substantially different in considering categorical monoproduct customer choices and network-valued co-subscription data. Flexible modelling of the conditional distribution of a network-valued random variable is still an on-going issue, which requires careful representations to borrow information across edges, to reduce the dimensionality and to maintain flexibility.

We propose a cluster-dependent mixture of latent eigenmodels, which allows the distribution of the co-subscription networks to change flexibly across clusters of agencies via cluster-specific mixing probabilities, while borrowing information among these agencies in learning the shared mixture components. Considering cluster dependence only in the mixing probabilities allows further dimensionality reduction, while providing simple and efficient computational methods. Differently from Dunson *et al*. (2008), we additionally avoid fixing the total number of clusters, but we learn this key quantity from our data via a Chinese restaurant process (CRP) prior for the cluster assignments.

The paper is organized as follows. In Section 2, we carefully describe our hierarchical formulation for joint modelling and coclustering of co-subscription networks and monoproduct choice data. Prior specification and summarized steps for posterior computation are outlined in Section 3. Section 4 considers simulation studies to evaluate the performance of our model in a scenario mimicking our motivating application. Results of the application to our business intelligence problem in an insurance company are presented and discussed in Section 5. Finally, Section 6 contains concluding remarks and suggests other possible fields of application along with further research directions.

2. Joint modelling of monoproduct choices and co-subscription networks

Let $\mathbf{C} = (C_1, \ldots, C_n)$ denote a vector of cluster assignments, with $C_i \in \{1, \ldots, K\}$ indicating the cluster membership of agency i. Agencies within the same cluster are characterized by a similar composition of their monoproduct customer choices as well as a comparable co-subscription network. To complete a specification of the joint model, we need to define a cluster-specific probabilistic representation of the monoproduct customer choices, as well as a cluster-specific probabilistic generative mechanism underlying the co-subscription networks. The latter is a key to define cross-sell strategies $u_{k1}, \ldots, u_{kV}$ in each cluster $k = 1, \ldots, K$, whereas the former provides the additional information to define the performance indicators $e_{k1}, \ldots, e_{kV}$, for $k = 1, \ldots, K$.

As a monoproduct customer can be associated with only one subscription $v = 1, \ldots, V$, it is straightforward to define a probabilistic representation for the monoproduct customer choices within each cluster. Let $\mathcal{Y}_k$ denote the categorical random variable characterizing the choices of the monoproduct customers in agencies within cluster k, for each $k = 1, \ldots, K$. To define the probability mass function that is associated with each $\mathcal{Y}_k$, we introduce a cluster-specific vector $\mathbf{p}_k = (p_{k1}, \ldots, p_{kV})$, with p_{kv} indicating the probability that a monoproduct customer in an agency within cluster k subscribes to product v, for each $v = 1, \ldots, V$. Assuming independence of the monoproduct customer choices, the joint probability for data $\mathbf{y}_i$ in agency i given its membership to cluster k is

$$\mathrm{pr}(\mathcal{Y}_k = y_{i1}) \, \mathrm{pr}(\mathcal{Y}_k = y_{i2}) \ldots \mathrm{pr}(\mathcal{Y}_k = y_{in_i}) = \prod_{v=1}^{V} p_{kv}^{n_{iv}}, \tag{1}$$

with n_{iv} the number of monoproduct customers in agency i who subscribed to product v.

Within each cluster k, the co-subscription networks are realizations from a network-valued random variable. As our network data are undirected and self-relations are not of interest, each symmetric adjacency matrix A_i, $i = 1, \ldots, n$, is uniquely characterized by its lower triangular elements comprising $\mathcal{L}(A_i) = (A_{i[21]}, A_{i[31]}, \ldots, A_{i[V1]}, A_{i[32]}, \ldots, A_{i[V2]}, \ldots, A_{i[V(V-1)]})$. Hence, in defining a probabilistic generative mechanism for the co-subscription networks within each cluster k, we can focus on the multivariate random variable $\mathcal{L}(\mathcal{A}_k)$ with binary elements $\mathcal{L}(\mathcal{A}_k)_l \in \{0, 1\}$, measuring the presence or absence of an edge among each pair of products $l = 1, \ldots, V(V-1)/2$ for agencies in cluster k. Note that in our notation $\mathcal{L}(\cdot)$ is an operator vectorizing the lower triangular elements of a given symmetric matrix.

As there are $2^{V(V-1)/2}$ possible configurations of co-subscription networks among V products, we cannot estimate the probability mass function that is associated with each $\mathcal{L}(\mathcal{A}_k)$, for $k = 1, \ldots, K$, non-parametrically without dimensionality reduction. To reduce dimension while maintaining flexibility, we model each $\mathcal{L}(\mathcal{A}_k)$ via a cluster-dependent mixture of latent eigenmodels. This leads to the following probability for the co-subscription network A_i in agency i given its membership to cluster k:

$$\mathrm{pr}\{\mathcal{L}(\mathcal{A}_k) = \mathcal{L}(A_i)\} = \sum_{h=1}^{H} \nu_{hk} \prod_{l=1}^{V(V-1)/2} (\pi_l^{(h)})^{\mathcal{L}(A_i)_l} (1 - \pi_l^{(h)})^{1 - \mathcal{L}(A_i)_l}, \tag{2}$$

with each component-specific edge probability vector $\boldsymbol{\pi}^{(h)} = (\pi_1^{(h)}, \ldots, \pi_{V(V-1)/2}^{(h)})^{\mathrm{T}} \in (0, 1)^{V(V-1)/2}$ defined as a function of a shared similarity vector $\mathbf{Z} \in \Re^{V(V-1)/2}$ and a component-specific vector $\mathbf{D}^{(h)} \in \Re^{V(V-1)/2}$ characterized via matrix factorization representations. In particular, we let

$$\boldsymbol{\pi}^{(h)} = \{1 + \exp(-\mathbf{Z} - \mathbf{D}^{(h)})\}^{-1}, \qquad \mathbf{D}^{(h)} = \mathcal{L}(X^{(h)} \Lambda^{(h)} X^{(h)^{\mathrm{T}}}), \quad h = 1, \ldots, H, \tag{3}$$

with the logistic mapping in expression (3) applied elementwise. Equations (2)–(3) carefully incorporate cluster dependence via cluster-specific mixing probabilities $\boldsymbol{\nu}_k = (\nu_{1k}, \ldots, \nu_{Hk})$, $k = 1, \ldots, K$, as well as network information by considering a latent eigenmodel for each mixture component.

Focusing on the mixture component h, the latent eigenmodel (Hoff, 2008) defines the undirected edges as realizations from conditionally independent Bernoulli random variables given their corresponding edge probabilities $\pi_l^{(h)} \in (0, 1)$, $l = 1, \ldots, V(V-1)/2$, and then borrows network information across these edge probabilities via lower dimensional representations. In particular, according to expression (3)—and letting l correspond to the pair of products v and u, $v > u$—each $\pi_l^{(h)}$ is constructed as a function of the pairwise similarity between products v and

u in a latent space, with this similarity arising from the dot product of the products' latent co-ordinate vectors $\mathbf{X}_v^{(h)} = (X_{v1}^{(h)}, \ldots, X_{vR}^{(h)})^{\mathrm{T}} \in \Re^R$, $v = 1, \ldots, V$, with $\mathbf{X}_v^{(h)^{\mathrm{T}}}$ the vth row of $X^{(h)}$. Hence, products having co-ordinates in the same direction are more likely to be co-subscribed than products with co-ordinates in opposite directions, with the $R \times R$ matrix $\Lambda^{(h)} = \mathrm{diag}(\boldsymbol{\lambda}^{(h)}) = \mathrm{diag}(\lambda_1^{(h)}, \ldots, \lambda_R^{(h)})$ weighting the similarity in each dimension r by a non-negative parameter $\lambda_r^{(h)}$. Note that, consistent with our notation, the vector $\mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)^{\mathrm{T}}})$ in expression (3) is equal to $(\mathbf{X}_2^{(h)^{\mathrm{T}}}\Lambda^{(h)}\mathbf{X}_1^{(h)}, \mathbf{X}_3^{(h)^{\mathrm{T}}}\Lambda^{(h)}\mathbf{X}_1^{(h)}, \ldots, \mathbf{X}_V^{(h)^{\mathrm{T}}}\Lambda^{(h)}\mathbf{X}_1^{(h)}, \mathbf{X}_3^{(h)^{\mathrm{T}}}\Lambda^{(h)}\mathbf{X}_2^{(h)}, \ldots, \mathbf{X}_V^{(h)^{\mathrm{T}}}\Lambda^{(h)}\mathbf{X}_2^{(h)}, \ldots, \mathbf{X}_V^{(h)^{\mathrm{T}}}\Lambda^{(h)}\mathbf{X}_{V-1}^{(h)})^{\mathrm{T}}$.

The latent eigenmodel provides an appealing choice in reducing the dimensionality from $V(V-1)/2$ edge probabilities to $V \times R$ latent co-ordinates and R weights—typically $R \ll V$—and has been shown to provide a more flexible characterization of the connectivity patterns and network structures than stochastic block models (Nowicki and Snijders, 2001), latent distance models (Hoff *et al*., 2002) and mixed membership stochastic block models (Airoldi *et al*., 2008). However, as discussed in Durante *et al*. (2015), a single latent eigenmodel fails in flexibly characterizing the probabilistic generative mechanism underlying a network-valued random variable.

To improve flexibility and to maintain computational tractability, equations (2)–(3) mix together H latent eigenmodels, while adding a common similarity vector $\mathbf{Z} \in \Re^{V(V-1)/2}$ shared between all the co-subscription networks and centring the mixture components to improve computational and clustering performance. According to Durante and Dunson (2015), this characterization guarantees full flexibility in approximating the cluster-specific probability mass functions for the co-subscription networks. Our goal is to exploit representation (1)–(3) to develop shared cross-sell strategies.

Fig. 2 provides an example of the output from our model for decision making in business intelligence when there are $n = 8$ agencies and $K = 3$ clusters. According to Fig. 2, agencies 1, 4 and 5 share common profiles of monoproduct customer choices and co-subscription networks as $C_1 = C_4 = C_5 = 1$. In Fig. 2, monoproduct choices in cluster 1 are characterized by the vector $\mathbf{p}_1$, whereas the co-subscription behaviour is summarized by the expectation of the network-valued random variable in cluster 1, $E\{\mathcal{L}(\mathcal{A}_1)\} = \bar{\boldsymbol{\pi}}_1 = \Sigma_{\mathcal{L}(A)\in\mathbb{A}_V}\, \mathcal{L}(A)\, \mathrm{pr}\{\mathcal{L}(\mathcal{A}_1) = \mathcal{L}(A)\}$, with $\mathbb{A}_V$ denoting the sample space of all the possible network configurations among V products. As discussed in Durante and Dunson (2015), under representation (2), $\bar{\boldsymbol{\pi}}_1 = \Sigma_{h=1}^{H} \nu_{h1}\boldsymbol{\pi}^{(h)}$ and coincides with the vector of co-subscription probabilities for pairs of products in cluster 1. Hence, $\bar{\boldsymbol{\pi}}_1$ can be used to define the set of cross-sell strategies for agencies 1, 4 and 5. The same description holds for clusters 2 and 3.

These quantities are estimated from our model and considered when defining the cluster-specific cross-sell strategies $u_{k1}, \ldots, u_{kV}$ and computing their performance indicators $e_{k1}, \ldots, e_{kV}$, for each $k = 1, \ldots, K$. Adapting the initial discussion in Section 1 to the output of our model, cross-sell strategy u_{kv} offers to monoproduct customers who subscribed to v in agencies within cluster k the additional product $u_{kv} = \arg\max_u \{\mathrm{pr}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$. The probability of a co-subscription between products v and u for agencies in cluster k, $\mathrm{pr}(\mathcal{A}_{k[vu]} = 1)$, is easily available from our model as $\bar{\pi}_{kl}$, with l the index denoting the pair v and u in the vectorized representation of the adjacency matrix. The performance indicator that is associated with u_{kv} is instead $e_{kv} = p_{kv} \max\{\mathrm{pr}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$.

As motivated above, the main purpose of our analysis is to cluster agencies having customers with similar monoproduct and multiproduct purchasing behaviour, while providing accurate estimates of the performance indicators for the various strategies. To develop procedures for tractable estimation and inference, it is useful to rely on a hierarchical specification that is equivalent to equations (2)–(3), which introduces an additional class index $G_i \in \{1, \ldots, H\}$ for each agency $i = 1, \ldots, n$, as follows:

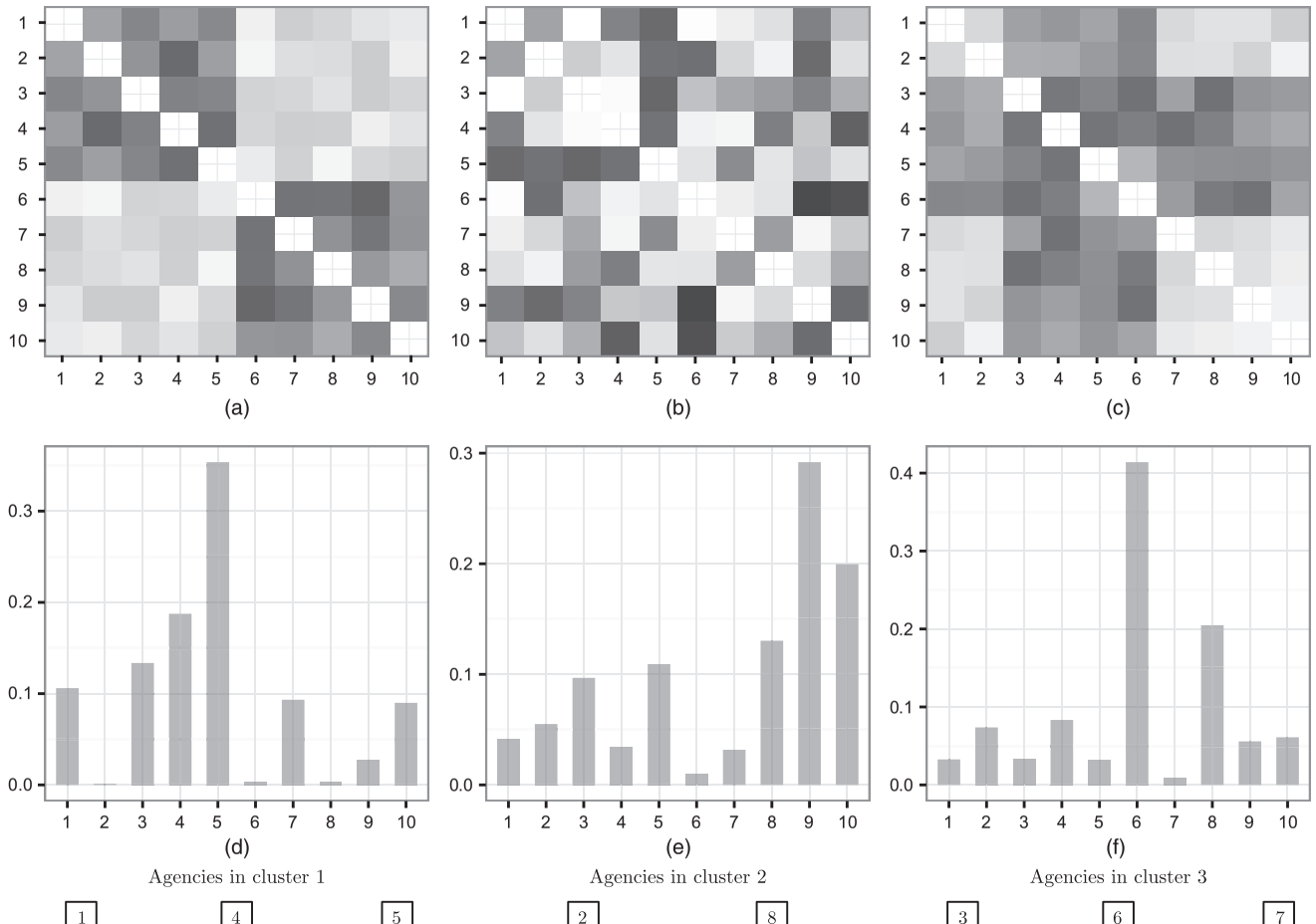

Fig. 2. Example of a possible output from our model for cross-selling (quantities in (a)–(c) are required to define the cross-sell strategies in each cluster; quantities in (d)–(f) provide the additional information to define the associated performance indicators): cluster-specific edge probability vectors (a) $\bar{\pi}_1$, (b) $\bar{\pi}_2$ and (c) $\bar{\pi}_3$—rearranged in adjacency matrix form—representing co-subscription profiles in clusters 1, 2 and 3 respectively (the colour goes from white to black as the probability goes from 0 to 1); cluster-specific probability vectors (d) $\mathbf{p}_1$, (e) $\mathbf{p}_2$ and (f) $\mathbf{p}_3$ measuring monoproduct customer choices in clusters 1, 2 and 3 respectively

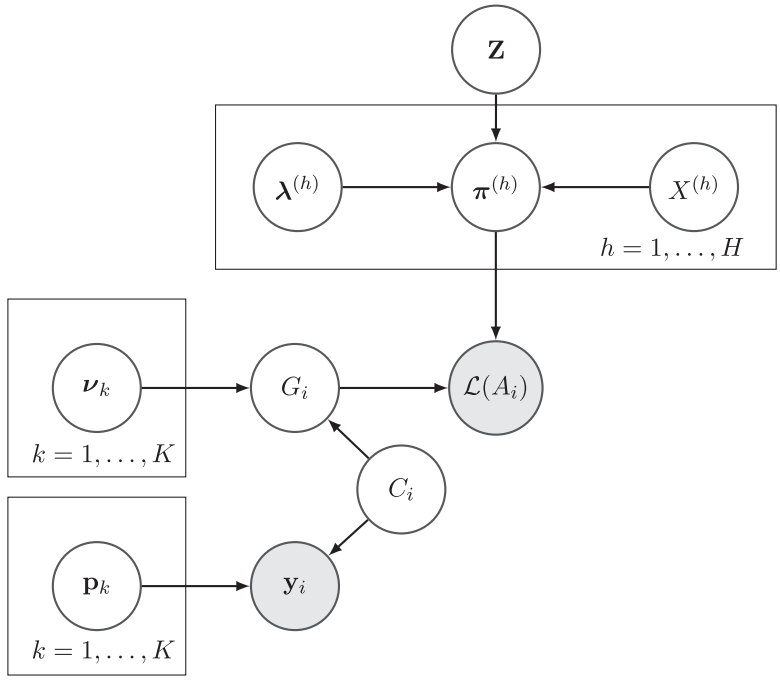

Fig. 3. Graphical representation of the mechanism to generate $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$, under representation (1) and (4)

$$\left.\begin{array}{cc} \mathcal{L}(\mathcal{A}_i)_l|\pi_{il} \overset{\text{indep}}{\sim} \text{Bern}(\pi_{il}), & l = 1, \ldots, V(V-1)/2, \\ (\boldsymbol{\pi}_i|G_i = h, \boldsymbol{\pi}^{(h)}) = \boldsymbol{\pi}^{(h)}, & \boldsymbol{\pi}^{(h)} = [1 + \exp\{-\mathbf{Z} - \mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)^{\mathrm{T}}})\}]^{-1} \\ \text{pr}(G_i = h|C_i = k) = \nu_{hk}, & h = 1, \ldots, H, \end{array}\right\} \quad (4)$$

independently for $i = 1, \ldots, n$, with $\boldsymbol{\pi}_i \in (0,1)^{V(V-1)/2}$ denoting the edge probability vector of agency i. Recalling that C_i is the cluster index for agency i, representation (4) shows that agencies in the same cluster have a common set of probability weights over the components in the mixture model for the co-subscription network. Moreover, although agencies in different clusters have different probabilities of being allocated in each mixture component, the model does not force the class indices to be necessarily different for agencies belonging to different clusters. In fact, under representation (4), two agencies i and j in different clusters $C_i \neq C_j$ are allowed to share the same component $G_i = G_j$ in the mixture model for the co-subscription networks. This facilitates efficient borrowing of information across clusters in modelling the mixture components characterizing the cluster-dependent mixture of latent eigenmodels. Refer to Fig. 3 for a graphical representation of our hierarchical model.

The next section develops a Bayesian approach for inference under the model proposed. Although model (2)–(3) is reminiscent of recent proposals from Durante *et al.* (2015) and Durante and Dunson (2015), the procedure that is developed in this contribution has key differences. Durante *et al.* (2015) developed a non-parametric mixture of latent eigenmodels to characterize the probability mass function of a network-valued random variable. Durante and Dunson (2015) tested for changes in this probability mass function across the values of a categorical predictor. Our current focus is instead on using the flexible and parsimonious joint model that is presented in Fig. 3 to develop shared cross-sell strategies. In our business intelligence problem, the con-

ditional distribution of the co-subscription networks varies as a function of a latent clustering variable, which is endogenously determined by monoproduct choices and co-subscription patterns shared across subsets of agencies. These cluster assignments are key unknown quantities, requiring careful priors and adapted computational methods.

3. Prior specification and posterior computation

3.1. Prior specification

The assignment vector $\mathbf{C} = (C_1, \ldots, C_n)$ of the agencies and the total number of clusters K are key unknowns in our analysis. There is a rich literature on prior probability models for $\mathbf{C}$ and K. A widely used approach is the CRP (Aldous, 1985), in which each cluster attracts new units in proportion to its size. In particular, letting $\mathbf{C} = (C_1, \ldots, C_n) \sim \mathrm{CRP}(\alpha_c)$, the prior distribution over clusters for the ith agency, conditioned on the membership of the others, $C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n$, is

$$\mathrm{pr}(C_i = k | C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n) = \begin{cases} \dfrac{n_{k,-i}}{n-1+\alpha_c} & \text{for } k = 1, \ldots, K_{-i}, \\ \dfrac{\alpha_c}{n-1+\alpha_c} & \text{for } k = K_{-i}+1, \end{cases} \tag{5}$$

where K_{-i} is the total number of non-empty clusters after removing C_i, $n_{k,-i}$ is the total number of agencies allocated to cluster k, excluding the ith, and $\alpha_c > 0$ is a concentration parameter controlling the expected number of occupied clusters $E(K) = O\{\alpha_c \log(n)\}$. High values of α_c favour more clusters *a priori*, with $E(K)$ growing as the sample size n increases.

Each cluster characterizes a specific joint profile of monoproduct and multiproduct customer behaviours, with these profiles potentially arising from specific preferences of customers in different regions or economic areas where each agency operates. Differences in profiles may be also associated with particular business administration choices of each agency. Hence, we expect a growing K as the number of agencies in the sample increases, with the sublinear relationship between $E(K)$ and n favouring parsimony in the number of cross-sell strategies to be defined.

Equation (5) defines the conditional distribution of C_i given the other assignments, after rearranging indices so that clusters $1, \ldots, K_{-i}$ are non-empty after removing C_i. These operations are possible since the cluster labels are arbitrary and observations are exchangeable on the basis of the CRP prior; refer to Griffiths and Ghahramani (2011) and Gershman and Blei (2012) for an introductory overview. Although we focus on the CRP, our model can easily accommodate other commonly used priors for random partitions, such as the Pitman–Yor process or the Kingman paintbox—among others. Refer to Hjort *et al.* (2010) for a general overview.

Accurate clustering of agencies also relies on careful modelling of the sequence of cluster-specific monoproduct choices defined in equation (1) and the collection of cluster-specific probabilistic generative mechanisms associated with the co-subscription networks via equations (2)–(3). Efficient estimation of these quantities is fundamental to develop cross-sell strategies $u_{k1}, \ldots, u_{kV}$ in each cluster $k = 1, \ldots, K$ and to quantify their performance via $e_{k1}, \ldots, e_{kV}$. Hence, we define large support priors, which do not rule out *a priori* any generative mechanism while maintaining tractable computations. As $\mathbf{p}_k$ represents the probability mass function for the categorical random variable $\mathcal{Y}_k$, we let

$$\mathbf{p}_k = (p_{k1}, \ldots, p_{kV}) \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_V), \qquad k = 1, \ldots, K. \tag{6}$$

The prior for the probabilistic generative mechanism of the co-subscription networks in each cluster is defined by choosing independent priors for the quantities in equations (2)–(3). To

maintain computational tractability, we consider independent Gaussian priors for the common similarities $Z_l \sim N(\mu_l, \sigma_l^2)$, $l = 1, \ldots, V(V-1)/2$, and standard Gaussian priors for the latent coordinates $X_{vr}^{(h)} \sim N(0, 1)$, for each $v = 1, \ldots, V$, $r = 1, \ldots, R$ and $h = 1, \ldots, H$. Adapting Rousseau and Mengersen (2011), we choose independent Dirichlet priors with small hyperparameters for each cluster-specific mixing probability vector $\boldsymbol{\nu}_k = (\nu_{1k}, \ldots, \nu_{Hk}) \sim \text{Dirichlet}(1/H, \ldots, 1/H)$, $k = 1, \ldots, K$, to favour deletion of redundant mixture components that are not required to characterize the co-subscription networks. Finally, as the dimension of the latent spaces in the eigenmodels is unknown, we consider a shrinkage prior on the weights vector $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$, which adaptively deletes redundant latent space dimensions that are not required to characterize the co-subscription probabilities. This is accomplished by choosing multiplicative inverse gamma $\text{MIG}(a_1, a_2)$ priors (Bhattacharya and Dunson, 2011) for vectors $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$, which favour shrinkage effects by inducing priors on elements $\lambda_r^{(h)}$ for $r = 1, \ldots, R$, that are increasingly concentrated close to zero as r increases, for appropriate a_2. Hence

$$\lambda_r^{(h)} = \prod_{m=1}^{r} \frac{1}{\vartheta_m^{(h)}}, \qquad \vartheta_1^{(h)} \sim \text{Ga}(a_1, 1), \quad \vartheta_{m \geqslant 2}^{(h)} \sim \text{Ga}(a_2, 1), \quad r = 1, \ldots, R, \quad h = 1, \ldots, H. \tag{7}$$

Beside providing simple algorithms for posterior inference, adapting results in Durante and Dunson (2015), this specification induces a prior on the probabilistic generative mechanism for the co-subscription networks in each cluster having full support. Full prior support is a key property to ensure good performance in defining accurate cross-sell strategies, because, without prior support about the true data-generating process, the posterior cannot possibly concentrate around the truth.

3.2. Posterior computation

Posterior computation is available via a simple Gibbs sampler which exploits results in Neal (2000) to allocate agencies to clusters under the CRP prior and steps in Durante *et al.* (2015) to update the quantities in equations (2)–(3). The Markov chain Monte Carlo (MCMC) routine alternates between two main parts summarized below. Complete pseudocode with step-by-step guidelines for implementation is provided in Appendix A.

The first part of the Gibbs sampler updates the cluster-specific probabilistic representation of the monoproduct customer choices and the cluster-specific probabilistic generative mechanism underlying co-subscription networks, given the cluster assignments $C_1, \ldots, C_n$ and the data $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$, $i = 1, \ldots, n$. In particular we have the following steps.

Step 1: given the monoproduct choice data $\mathbf{y}_1, \ldots, \mathbf{y}_n$ and the cluster assignments $C_1, \ldots, C_n$, the monoproduct choice probabilities $\mathbf{p}_k = (p_{k1}, \ldots, p_{kV})$ in each cluster $k = 1, \ldots, K$ are updated by exploiting the Dirichlet–multinomial conjugacy.

Step 2: in updating the cluster-specific probabilistic generative mechanism underlying the co-subscription data in equations (2)–(3), each network is first allocated to one of the H mixture components. This is accomplished by sampling each G_i from a discrete distribution conditioned on C_i, $\mathcal{L}(A_i)$ and the quantities in representation (4), for every $i = 1, \ldots, n$. This data augmentation step exploits standard derivations for posterior inference in mixture models.

Step 3: given $G_1, \ldots, G_n$ and $C_1, \ldots, C_n$, the vectors of mixing probabilities $\boldsymbol{\nu}_k = (\nu_{1k}, \ldots, \nu_{Hk})$ in each cluster $k = 1, \ldots, K$ are updated by exploiting again the Dirichlet–multinomial conjugacy.

Step 4: given $G_1, \ldots, G_n$ and $\mathcal{L}(A_1), \ldots, \mathcal{L}(A_n)$, the quantities $\mathbf{Z}$, $X^{(h)}$ and $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$—characterizing the mixture components in equations (2)–(3)—are updated by exploiting the

steps in Durante *et al.* (2015), which recast the problem in terms of Bayesian logistic regression and leverage a recent Pólya–gamma data augmentation scheme (Polson *et al.*, 2013) to maintain conjugacy.

Step 5: given $\mathbf{Z}$, $X^{(h)}$ and $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$, the samples for the component-specific edge probability vectors $\boldsymbol{\pi}^{(h)}$ are obtained by exploiting equation (3), for each $h = 1, \ldots, H$.

In performing these steps, the number of mixture components H and the dimension of the latent spaces R are set at conservative upper bounds, allowing the shrinkage priors for these quantities to empty redundant components that are not required to characterize the observed data. If all the mixture components are occupied or the posteriors for some weights $\lambda_r^{(h)}$ are not concentrated close to zero for any h, this suggests that H or R should be increased respectively.

Steps 1–5 are straightforward to compute exploiting the model representation in Fig. 3. This representation provides key computational benefits also in the second part of the MCMC routine in which the cluster indicators $C_1, \ldots, C_n$ are sampled from their full conditional. Adapting results in Neal (2000), each C_i is updated via a sequential reseating procedure given $\mathbf{C}_{-i} = (C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n)$, data $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$ and quantities in representations (1) and (4). The key step underlying this second part is the following step.

Step 6: update the cluster of i from the full conditional categorical variable with cluster probabilities

$$\text{pr}(C_i = k|-) \propto \begin{cases} \dfrac{n_{k,-i}}{n-1+\alpha_c} \text{pr}\{\mathcal{L}(A_i), \mathbf{y}_i, G_i | C_i = k, \boldsymbol{\pi}^{(G_i)}, \mathbf{p}_k, \boldsymbol{\nu}_k\} & \text{for } k = 1, \ldots, K_{-i}, \\ \dfrac{\alpha_c}{n-1+\alpha_c} \text{pr}\{\mathcal{L}(A_i), \mathbf{y}_i, G_i | C_i = K_{-i}+1, \boldsymbol{\pi}^{(G_i)}\} & \text{for } k = K_{-i}+1, \end{cases} \tag{8}$$

for each $i = 1, \ldots, n$.

To perform step 6 we need to compute the conditional probabilities in equation (8) at each MCMC iteration. Although this is apparently a cumbersome task, our model formulation (1)–(3) along with its hierarchical representation in equation (4) allows key simplifications, substantially improving computational tractability. Specifically, under our model, the conditional probability $\text{pr}\{\mathcal{L}(A_i), \mathbf{y}_i, G_i | C_i = k, \boldsymbol{\pi}^{(G_i)}, \mathbf{p}_k, \boldsymbol{\nu}_k\}$ for augmented data—including G_i—can be factorized as

$$\text{pr}\{\mathcal{L}(A_i)|G_i, \boldsymbol{\pi}^{(G_i)}\} \text{pr}(\mathbf{y}_i|C_i = k, \mathbf{p}_k) \text{pr}(G_i|C_i = k, \boldsymbol{\nu}_k) = \text{pr}\{\mathcal{L}(A_i)|G_i, \boldsymbol{\pi}^{(G_i)}\} \prod_{v=1}^{V} p_{kv}^{n_{iv}} \prod_{h=1}^{H} \nu_{hk}^{\mathbf{1}_{(h)}(G_i)} \tag{9}$$

with $\mathbf{1}_{(h)}(G_i) = 1$ if $G_i = h$ and $\mathbf{1}_{(h)}(G_i) = 0$ otherwise. According to equation (9) inducing cluster-dependence through the mixing probabilities $\boldsymbol{\nu}_k$, while considering cluster-independent mixture components in equations (2)–(3), has the key benefit of maintaining $\text{pr}\{\mathcal{L}(A_i)|G_i, \boldsymbol{\pi}^{(G_i)}\}$ constant across the cluster assignments. In fact, recalling formulation (4), the probability of observing $\mathcal{L}(A_i)$, given that $G_i = h$, is

$$\text{pr}\{\mathcal{L}(A_i)|G_i = h, \boldsymbol{\pi}^{(h)}\} = \prod_{l=1}^{V(V-1)/2} (\pi_l^{(h)})^{\mathcal{L}(A_i)_l} (1 - \pi_l^{(h)})^{1-\mathcal{L}(A_i)_l}$$

and does not depend on the cluster assignment C_i. As a result $\text{pr}\{\mathcal{L}(A_i)|G_i, \boldsymbol{\pi}^{(G_i)}\}$ acts as a multiplicative constant in expression (8), allowing expression (8) to be further simplified as

$$\text{pr}(C_i = k|-) \propto \begin{cases} \dfrac{n_{k,-i}}{n-1+\alpha_c} \displaystyle\prod_{v=1}^{V} p_{kv}^{n_{iv}} \prod_{h=1}^{H} \nu_{hk}^{\mathbf{1}_{(h)}(G_i)} & \text{for } k = 1, \ldots, K_{-i}, \\ \dfrac{\alpha_c}{n-1+\alpha_c} \text{pr}(\mathbf{y}_i|C_i = K_{-i}+1)\, \text{pr}(G_i|C_i = K_{-i}+1) & \text{for } k = K_{-i}+1, \end{cases}$$

where the two marginal probabilities corresponding to a newly occupied cluster are easily available by exploiting the Dirichlet–multinomial conjugacy. In particular, it is easy to show that

$$\text{pr}(\mathbf{y}_i|C_i = K_{-i}+1) = \int \prod_{v=1}^{V} p_{K_{-i}+1,v}^{n_{iv}} \text{d}\Pi(\mathbf{p}_{K_{-i}+1}) = \frac{\Gamma\left(\displaystyle\sum_{v=1}^{V} \alpha_v\right)}{\displaystyle\prod_{v=1}^{V} \Gamma(\alpha_v)} \frac{\displaystyle\prod_{v=1}^{V} \Gamma(\alpha_v + n_{iv})}{\Gamma\left\{\displaystyle\sum_{v=1}^{V} (\alpha_v + n_{iv})\right\}}, \tag{10}$$

for the monoproduct choices, and

$$\text{pr}(G_i|C_i = K_{-i}+1) = \int \prod_{h=1}^{H} \nu_{h,K_{-i}+1}^{\mathbf{1}_{(h)}(G_i)} \text{d}\Pi(\boldsymbol{\nu}_{K_{-i}+1}) = \frac{\Gamma\left(\displaystyle\sum_{h=1}^{H} 1/H\right)}{\displaystyle\prod_{h=1}^{H} \Gamma(1/H)} \frac{\displaystyle\prod_{h=1}^{H} \Gamma\{1/H + \mathbf{1}_{(h)}(G_i)\}}{\Gamma\left[\displaystyle\sum_{h=1}^{H} \{1/H + \mathbf{1}_{(h)}(G_i)\}\right]}, \tag{11}$$

for the class indicator variable in the cluster-dependent mixture of latent eigenmodels.

Hence, considering only cluster dependence in the mixing probability vectors $\boldsymbol{\nu}_k$, $k = 1, \ldots, K$, and exploiting the augmented data G_i, $i = 1, \ldots, n$, in the mixture representation for the generative mechanism underlying co-subscription networks allow a massive gain in computational tractability for step 6 in the second part of the MCMC routine. In fact, whereas equations (10) and (11) can be easily derived in closed form, the marginal probability of the co-subscription networks with respect to the edge probability vectors arising from the construction (4) is not analytically available.

4. Simulation studies

We consider a simulation study to evaluate the performance of our method in accurately recovering clusters of agencies and in efficiently estimating the key quantities that are required to define the set of cross-sell strategies for each cluster and their associated performance indicators. In simulating data, we define a scenario mimicking the structure of our application.

We focus on $n = 200$ agencies equally divided in $K = 4$ latent clusters and consider a total number of $V = 15$ products. Graphical analyses of our data—highlighted in Fig. 1—show that monoproduct customers typically concentrate on a small subset of the available products with high probability, while choosing the remaining set with very low frequency. We maintain this behaviour in constructing $\mathbf{p}_k^0$, $k = 1, \ldots, 4$, while defining a challenging scenario with small changes in $\mathbf{p}_k^0$ across clusters. In particular, we set p_{1v}^0 and p_{2v}^0 to be equal for all products v except for permuting 1 and 9, so $\text{pr}(\mathcal{Y}_1^0 = 1) = \text{pr}(\mathcal{Y}_2^0 = 9)$ and $\text{pr}(\mathcal{Y}_1^0 = 9) = \text{pr}(\mathcal{Y}_2^0 = 1)$. We adopt a similar strategy for clusters 3 and 4 by considering p_{3v}^0 and p_{4v}^0 equal for all products v except 3 and 7, where we let $\text{pr}(\mathcal{Y}_3^0 = 3) = \text{pr}(\mathcal{Y}_4^0 = 7)$ and $\text{pr}(\mathcal{Y}_3^0 = 7) = \text{pr}(\mathcal{Y}_4^0 = 3)$. We simulate monoproduct subscription data y_{is}, $i = 1, \ldots, 200$ and $s = 1, \ldots, 500$, from the categorical random variable $\mathcal{Y}_k^0$ with probability mass function $\mathbf{p}_k^0$, where $k = 1$ for agencies $i = 1, \ldots, 50$, $k = 2$ for $i = 51, \ldots, 100$, $k = 3$ for $i = 101, \ldots, 150$ and $k = 4$ for $i = 151, \ldots, 200$. Although agencies in our application have at least about 1000 monoproduct customers, we consider a smaller number $n_i = 500$ for each agency $i = 1, \ldots, 200$ to evaluate the performance when there are less data.

Co-subscription networks are simulated from our constructive representation of the mixture model in equation (2). We consider $H = 3$ mixture components with each edge probability vector $\boldsymbol{\pi}^{0(h)}$ generated to mimic a possible co-subscription scenario. Specifically, $\boldsymbol{\pi}^{0(1)}$ is characterized by one dense community among 10 possibly highly related products, while assigning low probability to the remaining pairs of products. Vector $\boldsymbol{\pi}^{0(2)}$ represents the case of four hub products, which occur with high probability in multiproduct customer choices, while fixing the remaining co-subscription probabilities at low values. To reduce separation between the mixture components and to provide a more challenging scenario, the edge probability vector $\boldsymbol{\pi}^{0(3)}$ is very similar to $\boldsymbol{\pi}^{0(2)}$ with the exception of product $v = 4$ which is held out from the hub products. In avoiding the eigenmodel construction (3) in the definition of $\boldsymbol{\pi}^{0(h)}$, $h = 1, \ldots, 3$, we additionally aim to evaluate the performance of representation (3) in accurately characterizing the co-subscription probabilities for each component h.

In simulating data $\mathcal{L}(A_i), i = 1, \ldots, 200$, we consider cluster-specific mixing probability vectors $\boldsymbol{\nu}_1^0 = \boldsymbol{\nu}_2^0 = (0.9, 0.05, 0.05)$, $\boldsymbol{\nu}_3^0 = (0.05, 0.9, 0.05)$ and $\boldsymbol{\nu}_4^0 = (0.05, 0.05, 0.9)$. This choice allows the first co-subscription scenario defined by $\boldsymbol{\pi}^{0(1)}$ to be very likely in agencies belonging to clusters 1 and 2. Scenarios characterized by $\boldsymbol{\pi}^{0(2)}$ and $\boldsymbol{\pi}^{0(3)}$ are instead more likely in clusters 3 and 4 respectively. Letting $\boldsymbol{\nu}_1^0 = \boldsymbol{\nu}_2^0$ further reduces separation between clusters 1 and 2. These two clusters have very similar monoproduct choice probabilities and equal generative processes for the co-subscription networks, providing a challenging scenario to evaluate clustering performance.

We analyse the simulated data under our model (1)–(3). As in Durante *et al.* (2015), we set $a_1 = 2.5$, $a_2 = 3.5$ and $\sigma_l^2 = 10$, $l = 1, \ldots, V(V-1)/2$. Quantities μ_l, $l = 1, \ldots, V(V-1)/2$, are defined as $\mu_l = \text{logit}\{\Sigma_{i=1}^n \mathcal{L}(A_i)_l/n\}$, to centre the mixture representation for the co-subscription networks around a structure shared by all the agencies in the company. We adopt a similar strategy for the hyperparameters of the Dirichlet prior (6) by setting $\alpha_v = \Sigma_{i=1}^n n_{iv}/n$ for each $v = 1, \ldots, V$, to centre prior (6) near the averaged preferences of monoproduct customers in the entire company. This empirical choice improves performance relative to symmetric Dirichlet priors, which can be sensitive to the concentration parameter, overly penalizing the addition of new clusters when concentration is small. Finally, we set the concentration parameter $\alpha_c = 1$ in the CRP prior for the cluster assignments, according to standard practice. Our approach can be easily generalized to learn α_c from the data as in Escobar and West (1995). However, we obtain the same results in terms of clustering, cross-sell strategies and performance indicators when using $\alpha_c = 0.5$, $\alpha_c = 5$, $\alpha_c = 10$, $\alpha_c = 15$ and $\alpha_c = 20$.

We collect 5000 Gibbs iterations and set $H = 15$ and $R = 10$ as upper bounds for the number of mixture components and the dimension of the latent spaces respectively. These upper bounds provide a good choice, with the sparse Dirichlet prior for the mixing probabilities in $\boldsymbol{\nu}_k$, $k = 1, \ldots, K$, and the multiplicative inverse gamma prior for the weights in $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$, adaptively removing redundant components. Trace plots suggest that convergence is reached after a burn-in of 1000 and mixing for the quantities of interest for inference is good. As inference focuses on cluster-specific structures, it is important first to check for label switching issues, and to relabel the clusters at each MCMC iteration by using, for example, Stephens (2000) as needed. Trace plots suggest that label switching is not an issue in our simulation.

We initialize our MCMC algorithm by assigning all agencies to a single cluster. Then, using samples after burn-in, we estimate $\hat{\mathbf{C}}$ as corresponding to the partition of agencies observed in the maximum proportion of draws, providing an estimate of the maximum *a posteriori*. We found that this simple approach had good performance and stability in our simulations but, as the number of agencies increases, using more refined procedures, such as Medvedovic *et al.* (2004) and Lau and Green (2007), may be preferable. We estimate $\hat{K} = 4$, correctly grouping

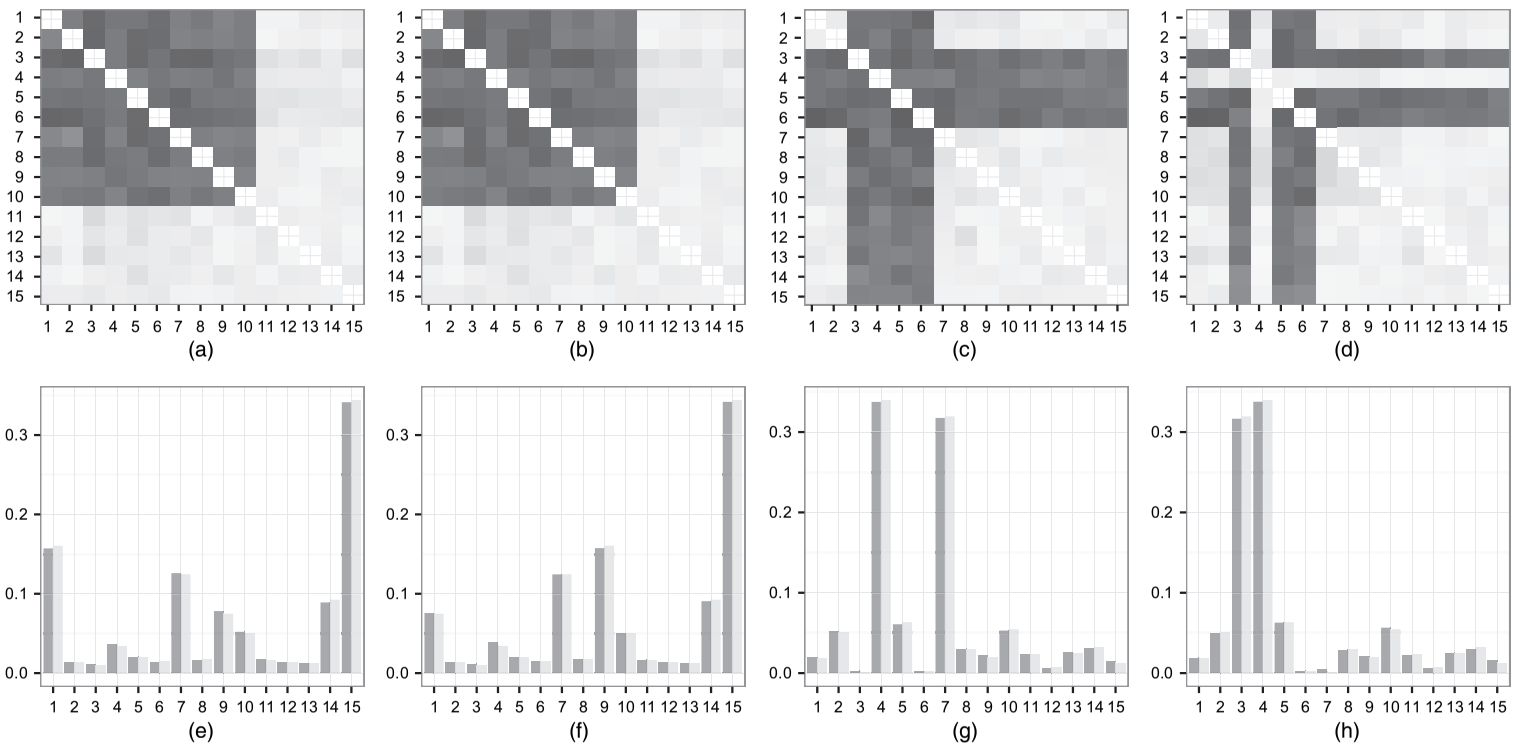

Fig. 4. For the four non-empty clusters, posterior mean $\hat{\bar{\pi}}_k$ of the co-subscription probabilities among pairs of products in each cluster k (lower triangle) and true $\bar{\pi}_k^0$ (upper triangle) for (a) $k=1$, (b) $k=2$, (c) $k=3$ and (d) $k=4$ (the colour goes from white to black as the probability goes from 0 to 1) and posterior mean $\hat{\mathbf{p}}_k$ (■) of the monoproduct choices in each cluster k and true $\mathbf{p}_k^0$ (■), for (e) $k=1$, (f) $k=2$, (g) $k=3$ and (h) $k=4$

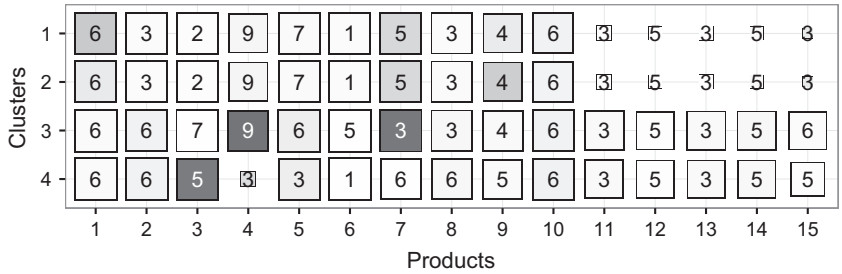

Fig. 5. Summary of the estimated cluster-specific cross-sell strategies $\hat{u}_{k1}, \ldots, \hat{u}_{kV}$ along with their performance indicators $\hat{e}_{k1}, \ldots, \hat{e}_{kV}$: each cell $[k, v]$ of the matrix defines the cross-sell strategy for monoproduct customers subscribed to product v in agencies belonging to cluster k, along with its performance; the number in cell $[k, v]$ corresponds to the best offer $\hat{u}_{kv} = \arg\max_u \{\widehat{\text{pr}}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$; the dimension of each square is proportional to $\max\{\widehat{\text{pr}}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$, whereas the greyscale of the colour is proportional to the corresponding estimated performance $\hat{e}_{kv} = \hat{p}_{kv} \max\{\widehat{\text{pr}}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$; the quantity $\widehat{\text{pr}}(\mathcal{A}_{k[vu]} = 1)$ is easily available from the posterior mean $\hat{\bar{\pi}}_{kl}$ of $\bar{\pi}_{kl}$ for each $k = 1, \ldots, K$ and $l = 1, \ldots, V(V-1)/2$

all the simulated agencies, including those in clusters 1 and 2, which are characterized by very subtle differences in their generating process.

Accurate clustering further allows efficient estimation of the cluster-specific components that are required for cross-selling. According to Fig. 4, we correctly estimate the true probability mass functions $\mathbf{p}_k^0$ characterizing the monoproduct choices in each cluster, with a similar performance in recovering the true vectors of co-subscription probabilities $\bar{\boldsymbol{\pi}}_k^0$. These estimates are a key to define the cluster-specific cross-sell strategies $u_{k1}, \ldots, u_{kV}$ and related performance indicators $e_{k1}, \ldots, e_{kV}$, as shown in Fig. 5. Consistent with results in Fig. 4, cross-sell strategies are the same in clusters 1 and 2 as $\hat{\bar{\boldsymbol{\pi}}}_1 \approx \hat{\bar{\boldsymbol{\pi}}}_2$, whereas performance indicators differ only for strategies targeting monoproduct customers subscribed to $v = 1$ or $v = 9$. The first segment is more profitable in cluster 1 whereas the second is in cluster 2. This is consistent with our estimates in Fig. 4 highlighting $\widehat{\text{pr}}(\mathcal{Y}_1 = 1) > \widehat{\text{pr}}(\mathcal{Y}_1 = 9)$ and $\widehat{\text{pr}}(\mathcal{Y}_2 = 9) > \widehat{\text{pr}}(\mathcal{Y}_2 = 1)$. Monoproduct customers subscribed to $v = 4$ and $v = 7$ are highly profitable in cluster 3 in being highly represented and having high co-subscription probability with at least one further product. Although $v = 4$ is highly populated in cluster 4, according to $\hat{\bar{\boldsymbol{\pi}}}_4$ it is not possible to find another product u having high co-subscription probability with $v = 4$, so such customers are not profitable in terms of targeting of cross-sell campaigns.

Conditionally on $\hat{C}_i = k$, agency i is associated with the cross-sell strategies suggested by $\hat{\bar{\boldsymbol{\pi}}}_k$. Hence, in ensuring accurate targeting, it is important to assess to what extent the estimated cluster-specific co-subscription probabilities adequately characterize the observed networks. Fig. 6 displays the boxplot of the areas under the receiver operating characteristic curve, AUC, with each AUC_i, $i = 1, \ldots, n$, obtained by using the simulated $\mathcal{L}(A_i)$ and the estimated co-subscription probability vector $\hat{\bar{\boldsymbol{\pi}}}_k$ of its corresponding cluster. Most of the co-subscription networks are adequately characterized by the co-subscription probability vectors that are specific to their clusters, with almost all the AUCs greater than 0.75. Decreased predictive performance is obtained for a small set of agencies that are highlighted by dots. For such agencies, the company may devise *ad hoc* cross-sell strategies based on the posterior mean of their agency-specific edge probability vector $\hat{\bar{\boldsymbol{\pi}}}_i$ rather than considering the cross-selling that is suggested to the clusters that they belong to. To evaluate the fit with respect to the monoproduct data, we consider the standardized L_1-distance between observed and estimated product frequencies $\epsilon_i = \Sigma_{v=1}^{V} |n_{iv}/n_i - \hat{p}_{kv}|/V$ with k the cluster to which agency i is allocated. In our simulation the

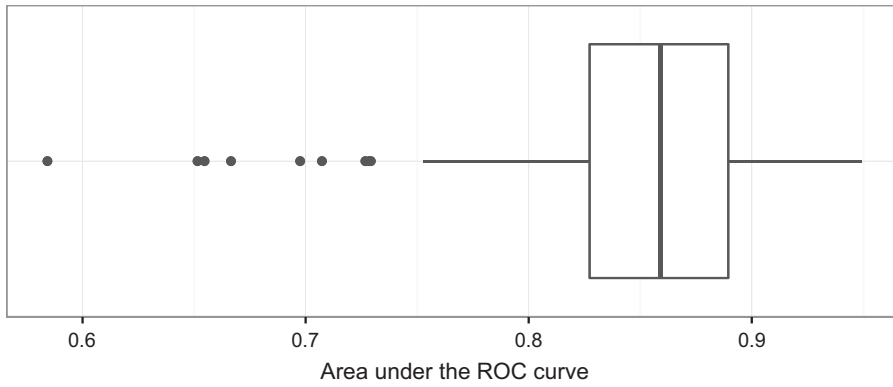

Fig. 6. Boxplot of the areas under the receiver operating characteristic curve $AUC_1, \ldots, AUC_{200}$: each AUC_i is obtained by using the simulated $\mathcal{L}(A_i)$ and the estimated co-subscription probability vector $\hat{\pi}_k$ of its corresponding cluster

maximum of these quantities is $\max(\epsilon_1, \ldots, \epsilon_n) = 0.013$ meaning that monoproduct choices in each cluster k are adequately characterized by $\hat{\mathbf{p}}_k$.

5. Cross-selling in the Italian insurance market

We apply the model that was outlined in Section 2 to our motivating business intelligence data set described in Section 1, which comprises monoproduct choice data and co-subscription networks for $n = 130$ agencies—within the same company—selling $V = 15$ different insurance products. Posterior computation uses the same settings as in the simulation study. Also in this case we obtain convergence, good mixing and no issues of label switching. Similarly to the simulation study, clustering of agencies and the associated cross-sell strategies along with their performance indicators do not substantially change when considering $\alpha_c = 0.5$, $\alpha_c = 5$, $\alpha_c = 10$, $\alpha_c = 15$ and $\alpha_c = 20$ instead of $\alpha_c = 1$.

The posterior distribution for the cluster assignments suggests a total of 20 clusters in our data. This is an appealing reduction of dimensionality in allowing the company to define 20 sets of shared cross-sell strategies, instead of $n = 130$ different campaigns. Fig. 7 provides a summarized overview of our estimated cluster-specific cross-sell strategies $\hat{u}_{k1}, \ldots, \hat{u}_{kV}$ along with their performance indicators $\hat{e}_{k1}, \ldots, \hat{e}_{kV}$, $k = 1, \ldots, 20$. As we can clearly see, the different clusters are typically characterized by similar cross-sell strategies, highlighting minor differences in monoproduct and multiproduct customer profiles across different clusters of agencies. This is a reasonable insight provided that the focus of our study is on agencies operating in similar markets within the same insurance company.

Accurate modelling of the cluster-specific probabilistic generative mechanisms for the co-subscription networks allows the definition of efficient sets of cluster-specific cross-sell strategies, offering to monoproduct customers policies that are complementary to their current choice. According to the results in Fig. 7, family- or individual-type policies, such as insurance for houses, cars, injuries, savings, investments and retirement plans, are typically associated with cross-sell offers from the same basket. The same holds for business-type policies covering business liability insurance, business activity insurance, compensatory damages and payment protection. Deepening the analysis of Fig. 7, popular best offers comprise insurance on investments, insurance on injuries and compensatory damages. Insurance on investments reasonably attracts monoprod-

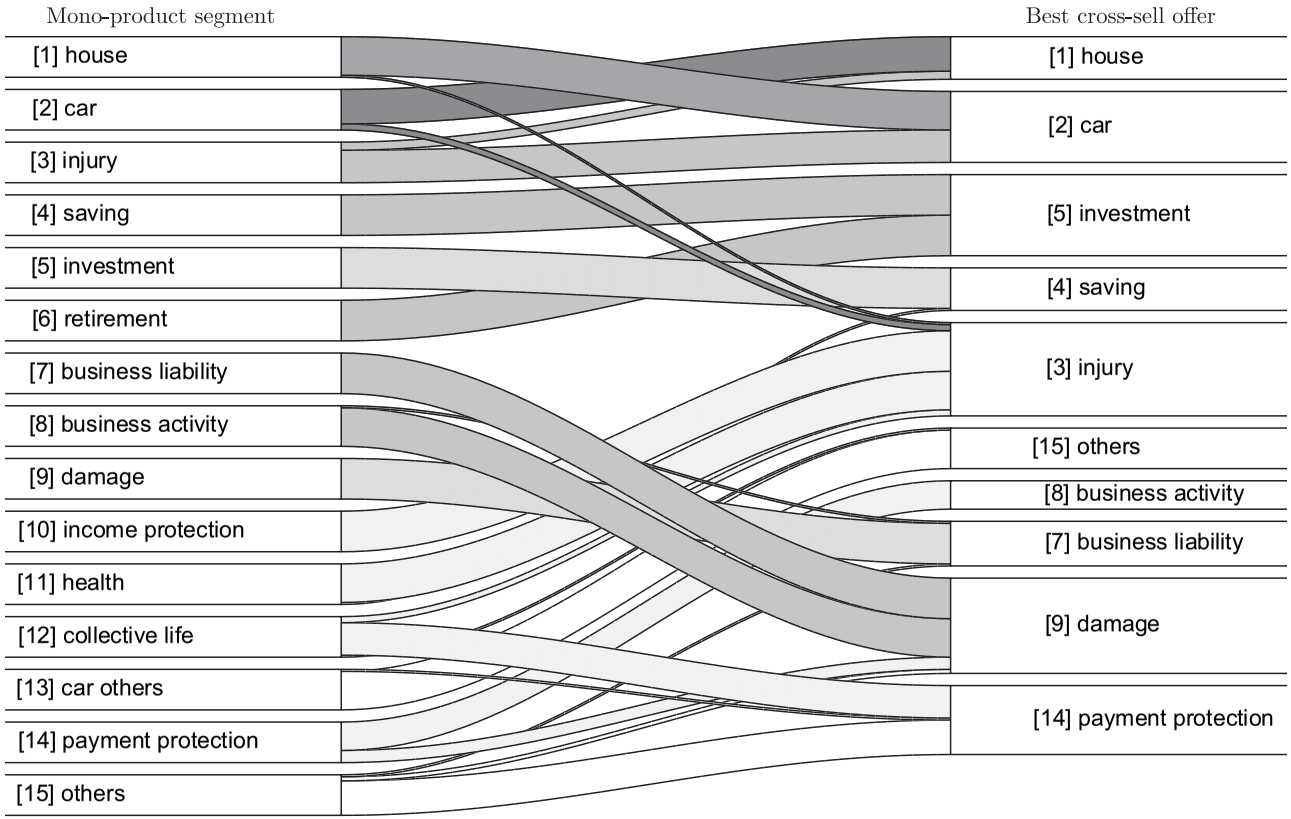

Fig. 7. Flow chart summarizing the estimated cluster-specific cross-sell strategies $\hat{u}_{k1},\ldots,\hat{u}_{kV}$ along with their performance indicators $\hat{e}_{k1},\ldots,\hat{e}_{kV}$ for the $k = 1,\ldots,20$ clusters in our application: the dimension of each edge connecting every product v—subscribed to by current monoproduct customers—to a best offer is proportional to the total number of clusters for which that best offer coincides; the greyscale of the colour for the edge starting from v is proportional to its performance indicators averaged across clusters $\sum_{k=1}^{20}\hat{e}_{kv}/20$, $v = 1,\ldots,15$; the dimension of each right-hand box is proportional to the total number of different cross-sell strategies suggesting the corresponding product as the best offer

uct customers who are currently subscribed to policies on savings and retirement plans, whereas compensatory damages are a convenient offer for customers who are currently subscribed to business-type insurance products. Finally insurance on injuries is a profitable offer for mono-product customers associated with health-type policies, such as medical health coverage and income protection insurance. Health-type policies guarantee benefits to policy holders who are unable to work because of illness.

Our results provide consistent findings highlighting the accuracy of our procedure for profiling and cross-selling. This is further confirmed by the analysis of the performance indicators that are associated with each strategy. According to Fig. 7 house and car insurance are in general more effective in creating new multiproduct customers. Beside being characterized by high co-subscription probabilities with other polices, these products are also highly populated by monoproduct customers as house insurance represents a common policy for families and car insurance—covering third-party liability policies—is compulsory in Italy.

Fig. 8 provides further clarifications of our findings by focusing on posterior summaries for the key quantities $\bar{\pi}_4$ and $\mathbf{p}_4$ that are associated with the most populated cluster $k = 4$, comprising 40 agencies. According to Fig. 8, our procedure provides a good joint representation of the different monoproduct and multiproduct sources of variability, while confirming previous insights on monoproduct and multiproduct customer choices with respect to insurance products. Focusing on the posterior mean of $\bar{\pi}_4$, we note two communities of products characterized by high co-subscription probabilities between polices in the same community and comparatively lower co-subscription probabilities between policies in different communities. Consistent with results in Fig. 7, the first group comprises the family- or individual-type policies that were previously discussed, whereas the second contains the business-related policies. The remaining policies are instead less common in customer multiple choices and therefore are consistently characterized by low co-subscription probabilities, suggesting ineffective cross-sell strategies. In addition the family- or individual-type policies are further characterized by two subcommunities. The first comprises house insurance, car insurance and insurance on injuries, whereas the second contains insurance on savings, insurance on investments and retirement plans. This is a consistent finding, as the two subcommunities properly characterize non-life and life insurances policies respectively.

The posterior mean of $\mathbf{p}_4$ in Fig. 8 confirms the previous comments on the performance indicators. Consistent with Fig. 7, house and car insurance policies are—in fact—the two products with the highest monoproduct choice probabilities. Therefore, these policies are highly populated by monoproduct customers. As a result, cross-sell strategies targeting these two segments have a high ceiling on effectiveness. Consistent with these results, the cross-sell strategies for cluster 4, along with their performance indicators, are mostly in line with those discussed in Fig. 7.

Posterior quartiles in Fig. 8 additionally highlight how accounting for network structure and borrowing information within and between clusters provide posterior distributions efficiently concentrated around our estimates. Posterior concentration is more evident for the monoproduct choice probabilities. This result is not surprising provided that data on monoproduct customers for agencies in cluster 4 jointly contribute to the posterior of parameters in $\mathbf{p}_4$, with the average number of monoproduct customers in each agency being about 5500.

Fig. 9 summarizes the estimated $\hat{\bar{\pi}}_k$ and $\hat{\mathbf{p}}_k$, for two additional representative clusters. As we can see, although monoproduct and multiproduct customers of agencies in different clusters are characterized by a very similar behaviour, our flexible procedure can capture subtle differences in the cluster-specific co-subscription probabilities and monoproduct customer choices. The estimated co-subscription probabilities in cluster 3 confirm the two-community structure that was previously discussed for cluster 4, while further highlighting the non-life and life sub-

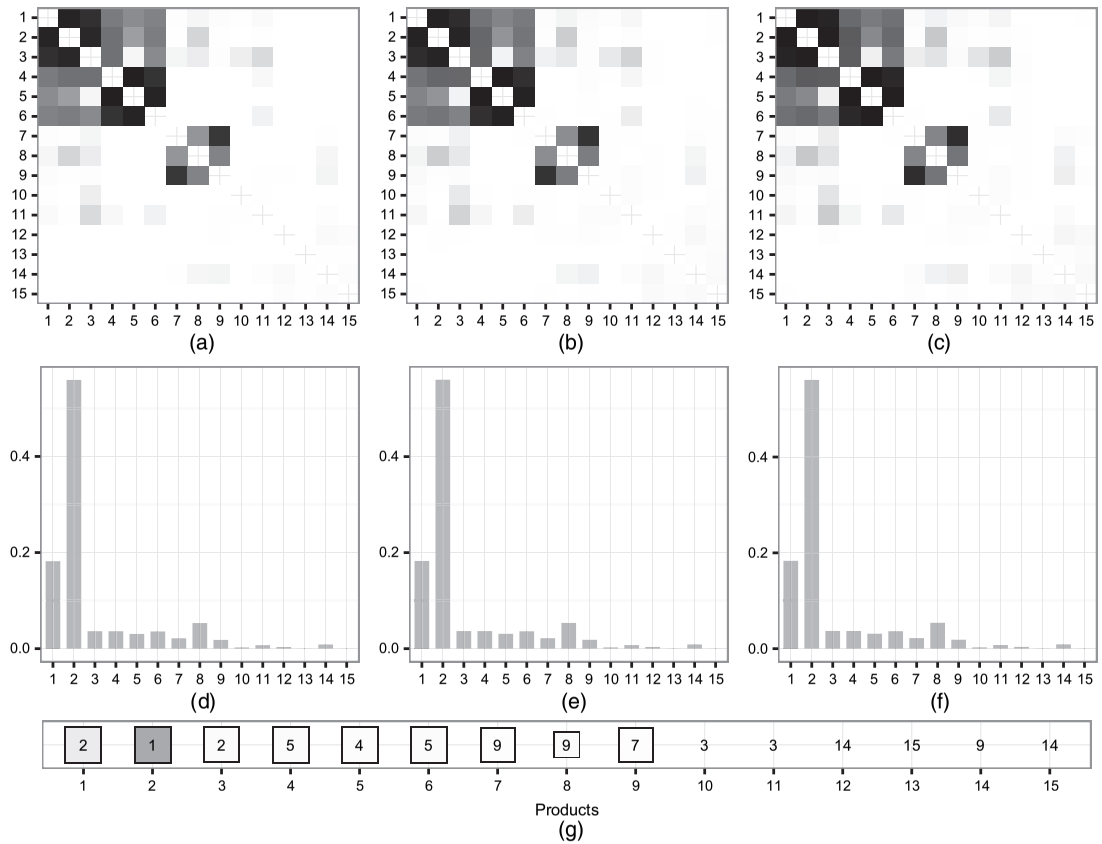

Fig. 8. For the most populated cluster $k = 4$, summary of the posterior distribution of the co-subscription probabilities $\bar{\pi}_{4l}$, $l = 1, \ldots, V(V-1)/2$, rearranged in matrix form (the colour goes from white to black as the probability goes from 0 to 1): (a) first quartile, (b) posterior mean and (c) third quartile, summary of the posterior distribution of the monoproduct choice probabilities p_{4v}, $v = 1, \ldots, 15$, for (d) first quartile, (e) posterior mean and (f) third quartile; (g) summary of the estimated cluster-specific cross-sell strategies $\hat{u}_{41}, \ldots, \hat{u}_{4V}$ along with their performance indicators $\hat{e}_{41}, \ldots, \hat{e}_{4V}$ (the number in cell v corresponds to the best offer $\hat{u}_{4v}$; the dimension of each square v is proportional to $\max\{\widehat{\text{pr}}(\mathcal{A}_{4[vu]} = 1) : u \neq v\}$, whereas the greyscale of the colour is proportional to the corresponding estimated performance $\hat{e}_{4v}$; to facilitate graphical analysis only squares associated with strategies with $\max\{\widehat{\text{pr}}(\mathcal{A}_{4[vu]} = 1) : u \neq v\} > 0.5$ are represented)

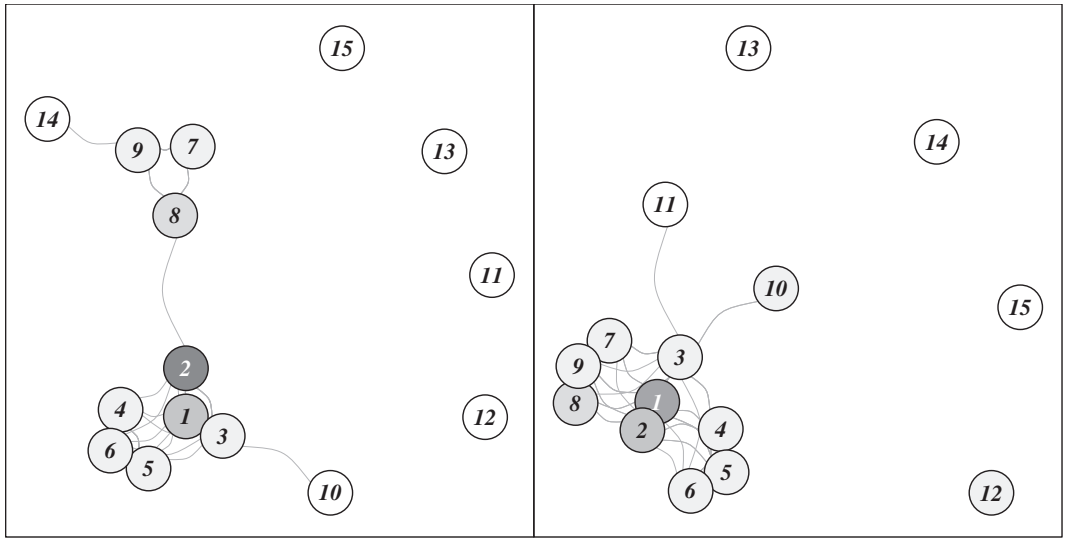

Fig. 9. For two representative clusters (a) cluster 3 and (b) cluster 14, weighted network visualization with weights given by the posterior mean $\hat{\bar{\pi}}_k$ of their associated co-subscription probability vectors $\bar{\pi}_k$: nodes positions are obtained by applying the Fruchterman and Reingold (1991) force-directed placement algorithm, whereas the greyscale of the colour of each node v is proportional to $\hat{\rho}_{kv}$, $v = 1, \ldots, 15$

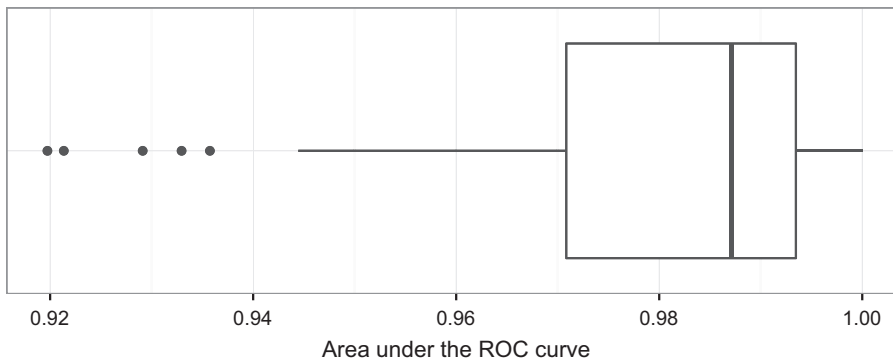

Fig. 10. Boxplot of the areas under the receiver operating characteristic curve $\text{AUC}_1, \ldots, \text{AUC}_{130}$: each AUC_i is obtained by using the observed $\mathcal{L}(A_i)$ and the estimated co-subscription probability vector $\hat{\bar{\pi}}_k$ of its corresponding cluster

communities within the family- or individual-type policies. Similarly to cluster 4, monoproduct customers in cluster 3 have relatively high preferences for house insurance and car insurance. However, we found that car insurance was slightly less populated in cluster 3 in favour of house insurance policies, modifying the corresponding performance indicators. Correctly identifying differences in $\mathbf{p}_k$ is a key to evaluate and rank the cross-sell strategies $\hat{u}_{k1}, \ldots, \hat{u}_{kV}$ according to their performance indicators $\hat{e}_{k1}, \ldots, \hat{e}_{kV}$.

The co-subscription network in cluster 14 displays instead more evident differences, with less separation between the two communities characterizing family- or individual-type policies and business-type policies. Hence, this cluster may refer to agencies dealing with customers who are motivated to co-subscribe to policies both for their business activity and for their private

Table 1. Part I of the Gibbs sampler for joint modelling of mixed domain data

Conditionally on $C_1, \ldots, C_n$ and the data $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$, $i = 1, \ldots, n$, update the priors for quantities in equations (1), (2) and (3) as follows

Step 1: update the cluster-specific monoproduct choice probabilities in $\mathbf{p}_k$, $k = 1, \ldots, K$:
for $k = 1, \ldots, K$ update $\mathbf{p}_k$ from

$$(p_{k1}, \ldots, p_{kV})|- \sim \text{Dirichlet}\left(\alpha_1 + \sum_{i:C_i=k} n_{i1}, \ldots, \alpha_V + \sum_{i:C_i=k} n_{iV}\right)$$

end for
Step 2: allocate each co-subscription network to one of the mixture components:
for $i = 1, \ldots, n$ sample the class indicator G_i from the discrete distribution with probabilities

$$\text{pr}(G_i = h|-) = \frac{\nu_{hC_i} \prod_{l=1}^{V(V-1)/2} (\pi_l^{(h)})^{\mathcal{L}(A_i)_l} (1-\pi_l^{(h)})^{1-\mathcal{L}(A_i)_l}}{\sum_{q=1}^{H} \nu_{qC_i} \prod_{l=1}^{V(V-1)/2} (\pi_l^{(q)})^{\mathcal{L}(A_i)_l} (1-\pi_l^{(q)})^{1-\mathcal{L}(A_i)_l}},$$

for each $h = 1, \ldots, H$, with $\boldsymbol{\pi}^{(h)}$ defined in equation (3)
end for
Step 3: update the cluster-specific vectors of mixing probabilities $\boldsymbol{\nu}_k$, $k = 1, \ldots, K$:
for $k = 1, \ldots, K$ update each $\boldsymbol{\nu}_k$ from

$$(\nu_{1k}, \ldots, \nu_{Hk})|- \sim \text{Dirichlet}(1/H + n_{1k}, \ldots, 1/H + n_{Hk})$$

with n_{hk} denoting the number of agencies in cluster k, currently associated with mixture component h
end for
Step 4: update quantities $\mathbf{Z}$, $X^{(h)}$ and $\boldsymbol{\lambda}^{(h)}$, $h = 1, \ldots, H$, via Pólya–gamma data augmentation (Polson *et al.*, 2013) by exploiting derivations in Durante *et al.* (2015):
for $h = 1, \ldots, H$

(a) for each $l = 1, \ldots, V(V-1)/2$ update augmented data $\omega_l^{(h)}$ from the Pólya–gamma data augmentation

$$\omega_l^{(h)}|- \sim \text{PG}\{n_h, Z_l + \mathcal{L}(X^{(h)} \Lambda^{(h)} X^{(h)\text{T}})_l\},$$

with $\text{PG}(b, c)$ denoting the Pólya–gamma distribution with parameters $b > 0$ and $c \in \Re$, and n_h the total number of agencies in mixture component h
end for
(b) update the shared similarity vector $\mathbf{Z}$ from

$$\mathbf{Z}|- \sim N_{V(V-1)/2}\{\boldsymbol{\mu}_Z, \text{diag}(\sigma^2_{Z_1}, \ldots, \sigma^2_{Z_{V(V-1)/2}})\},$$

with $\boldsymbol{\mu}_Z$ having elements

$$\mu_{Z_l} = \sigma^2_{Z_l}\left[\sigma_l^{-2}\mu_l + \sum_{h=1}^{H} \{\mathcal{L}(A^{(h)})_l - n_h/2 - \omega_l^{(h)} \mathcal{L}(X^{(h)} \Lambda^{(h)} X^{(h)\text{T}})_l\}\right],$$

where $\sigma^2_{Z_l} = 1/(\sigma_l^{-2} + \Sigma_{h=1}^{H} \omega_l^{(h)})$, for each $l = 1, \ldots, V(V-1)/2$ and $\mathcal{L}(A^{(h)})_l = \Sigma_{i:G_i=h} \mathcal{L}(A_i)_l$
for $h = 1, \ldots, H$

(c) as $\mathbf{D}^{(h)} = \mathcal{L}(X(h)\Lambda^{(h)} X^{(h)\text{T}}) = \mathcal{L}(\bar{X}^{(h)} \bar{X}^{(h)\text{T}})$, with $\bar{X}^{(h)} = X^{(h)} \Lambda^{(h)1/2}$, update elements in $\bar{X}^{(h)}$, knowing that, under our prior specification $\bar{X}^{(h)}_{vr}|\lambda_r^{(h)} \sim N(0, \lambda_r^{(h)})$; therefore
for $v = 1, \ldots, V$ block-sample the vth row of $\bar{X}^{(h)}$; define $\bar{\mathbf{X}}_v^{(h)} = (\bar{X}^{(h)}_{v1}, \ldots, \bar{X}^{(h)}_{vR})^{\text{T}}$ and let $\bar{X}^{(h)}_{(-v)}$ denote the $(V-1) \times R$ matrix obtained by removing the vth row in $\bar{X}^{(h)}$; consider the logistic regression

$$\begin{aligned} \mathcal{L}(A^{(h)})_{(v)} &\sim \text{Binom}(n_h, \boldsymbol{\pi}^{(h)}_{(v)}), \\ \text{logit}(\boldsymbol{\pi}^{(h)}_{(v)}) &= \mathbf{Z}_{(v)} + \bar{X}^{(h)}_{(-v)} \bar{\mathbf{X}}^{(h)}_v, \end{aligned}$$

with $\mathcal{L}(A^{(h)})_{(v)}$ and $\mathbf{Z}_{(v)}$ obtained by stacking elements $\mathcal{L}(A^{(h)})_l$ and Z_l respectively for all l corresponding to pairs having v as one of the two nodes, and ordered consistently with the linear

(*continued*)

Table 1 (*continued*)

predictor; exploiting previous logistic regression, and letting $\Omega_{(v)}^{(h)}$ be the diagonal matrix with the corresponding Pólya–gamma augmented data, the full conditional is

$$\bar{\mathbf{X}}_v^{(h)}|- \sim N_R\{(\bar{X}_{(-v)}^{(h)\mathrm{T}}\Omega_{(v)}^{(h)}\bar{X}_{(-v)}^{(h)} + \Lambda^{(h)^{-1}})^{-1}\boldsymbol{\eta}_v^{(h)}, (\bar{X}_{(-v)}^{(h)\mathrm{T}}\Omega_{(v)}^{(h)}\bar{X}_{(-v)}^{(h)} + \Lambda^{(h)^{-1}})^{-1}\},$$

with $\boldsymbol{\eta}_v^{(h)} = \bar{X}_{(-v)}^{(h)\mathrm{T}}\{\mathcal{L}(A^{(h)})_{(v)} - \mathbf{1}_{V-1}n_h/2 - \Omega_{(v)}^{(h)}\mathbf{Z}_{(v)}\}$

end for

(d) sample the gamma quantities defining the shrinkage weights $\lambda_1^{(h)}, \ldots, \lambda_R^{(h)}$:

$$\vartheta_1^{(h)}|- \sim \mathrm{Ga}\left\{a_1 + \frac{VR}{2}, 1 + \frac{1}{2}\sum_{m=1}^{R}\theta_m^{(-1)}\sum_{v=1}^{V}(\bar{X}_{vm}^{(h)})^2\right\},$$

$$\vartheta_{r\geqslant 2}^{(h)}|- \sim \mathrm{Ga}\left\{a_2 + \frac{V(R-r+1)}{2}, 1 + \frac{1}{2}\sum_{m=r}^{R}\theta_m^{(-r)}\sum_{v=1}^{V}(\bar{X}_{vm}^{(h)})^2\right\},$$

where $\theta_m^{(-r)} = \Pi_{t=1, t\neq r}^{m}\vartheta_t^{(h)}$ for $r = 1, \ldots, R$

end for

Step 5: update the edge probability vectors $\boldsymbol{\pi}^{(h)}$, in each mixture component $h = 1, \ldots, H$:

for $h = 1, \ldots, H$ compute

$$\boldsymbol{\pi}^{(h)} \text{ as } \boldsymbol{\pi}^{(h)} = [1 + \exp\{-\mathbf{Z} - \mathcal{L}(\bar{X}^{(h)}\bar{X}^{(h)\mathrm{T}})\}]^{-1}$$

end for

coverage. Moreover—differently from clusters 3 and 4—the monoproduct preferences for house insurance are higher than car insurance in cluster 14. Therefore, targeting the monoproduct customers who are currently subscribed to house insurance—instead of car insurance—may be more effective at increasing the number of multiproduct customers in agencies within cluster 14.

According to the boxplot of the AUCs in Fig. 10, the model performs well in characterizing the observed co-subscription networks $\mathcal{L}(A_i)$, $i = 1, \ldots, 130$, considering the co-subscription probability vectors $\hat{\bar{\boldsymbol{\pi}}}_k$, $k = 1, \ldots, 20$, specific to their clusters. All the AUCs are greater than 0.9, with few agencies displaying a lower performance. For such agencies, the company may devise cross-sell strategies based on their specific $\hat{\boldsymbol{\pi}}_i$. Our estimates provide also a good fit with respect to the monoproduct choice data with the maximum of the standardized L_1-distances between observed and estimated product frequencies for each agency being $\max(\epsilon_1, \ldots, \epsilon_{130}) = 0.055$.

6. Discussion

Motivated by a complex business intelligence problem for targeted advertising of cross-sell strategies in different agencies, we developed a flexible joint model for mixed domain data including monoproduct customer choices and co-subscription networks measuring multiple-purchasing behaviour. Our procedure defines shared sets of cross-sell strategies by effectively clustering agencies characterized by comparable customer behaviours. Each segment is carefully profiled by modelling monoproduct customer choices and co-subscription networks via a cluster-dependent mixture of latent eigenmodels. Exploiting such estimates, we construct cluster-specific sets of cross-sell strategies informing for each product v which additional product $u \neq v$ should be offered to obtain the highest probability of a co-subscription by a monoproduct customer subscribed to v. We evaluate the effectiveness of each strategy via performance indicators accounting also for monoproduct customer choice data. We provide straightforward algorithms

for posterior computation. The application to multiple agencies from an Italian insurance company highlights the key benefits of our model in providing interesting insights and effective strategies.

Although we focus on cross-sell strategies offering a single additional product, our procedure can be easily generalized to multioffer cross-sell strategies, advertising to monoproduct customers currently subscribed to v in agencies within cluster k the additional set of products $\mathbf{u}_{kv} = (u_{kv1}, \ldots, u_{kvM}) = \arg\max_{u_1,\ldots,u_M} \{\Pr(\mathcal{A}_{k[vu_1]} = 1, \ldots, \mathcal{A}_{k[vu_M]} = 1) : u_1 \neq \ldots \neq u_M \neq v\}$. As our representation (2)–(3) induces a probability mass function on the entire co-subscription network, the quantity $\Pr(\mathcal{A}_{k[vu_1]} = 1, \ldots, \mathcal{A}_{k[vu_M]} = 1)$ can be easily derived by marginalizing out in equation (2) all the pairs of products that did not enter the previous joint probability. It is also worth considering further research including additional information, such as costs of the strategies and product prices. Such quantities impact on the definition and evaluation of the strategies along with monoproduct and multiproduct choices. Our model still provides a valuable building block in estimating such quantities.

Although we focus on cross-selling in business intelligence, our model has a broad range of applications. For example, our procedure can provide key information on efficient allocation of resources in public services based on monoservice and multiservice citizen data from different cities or states.

Acknowledgements

We thank the Joint Editor, the Guest Associate Editor and the referees for their valuable comments. This work was partially funded by grant CPDA154381/15 from the University of Padova, Italy, and by grant N00141410245 of the US Office of Naval Research.

Appendix A: Pseudocode for posterior computation

This appendix provides detailed steps of the Gibbs sampler for posterior computation under our model

Table 2. Part II of the Gibbs sampler for joint modelling of mixed domain data

Conditionally on quantities in equation (1) and (2)–(3) and data $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$, update the cluster indices in $\mathbf{C}$

Step 6: sample the cluster assignments $C_1, \ldots, C_n$ via a sequential reseating procedure:

for $i = 1, \ldots, n$ update C_i conditionally on $\mathbf{C}_{-i} = (C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n)$

(a) remove agency i since we are going to sample its cluster membership C_i,
(b) if no other agencies are in the same cluster of i, this cluster becomes empty and is removed along with its associated monoproduct choice probabilities and co-subscription network probability mass function,
(c) reorder cluster indices so that clusters $1, \ldots, K_{-i}$ are non-empty,
(d) update the cluster of i from the full conditional categorical variable with cluster probabilities

$$\Pr(C_i = k|-) \propto \begin{cases} \dfrac{n_{k,-i}}{n-1+\alpha_c} \displaystyle\prod_{v=1}^{V} p_{kv}^{n_{iv}} \prod_{h=1}^{H} \nu_{hk}^{\mathbf{1}_{(h)}(G_i)} & \text{for } k = 1, \ldots, K_{-i}, \\ \dfrac{\alpha_c}{n-1+\alpha_c} \Pr(\mathbf{y}_i|C_i = K_{-i}+1) \Pr(G_i|C_i = K_{-i}+1) & \text{for } k = K_{-i}+1, \end{cases}$$

where $\Pr(\mathbf{y}_i|C_i = K_{-i}+1)$ and $\Pr(G_i|C_i = K_{-i}+1)$ are analytically available in equations (10) and (11) respectively,
(e) if i is assigned a new cluster $K_{-i}+1$, add this new cluster and sample a new $\mathbf{p}_{K_{-i}+1}$ and $\boldsymbol{\nu}_{K_{-i}+1}$ conditionally on $\mathbf{y}_i$ and G_i according to steps 1 and 3 respectively.

end for

formulation. The first part of the MCMC routine—outlined in Table 1—updates the quantities in expressions (1) and (2)–(3), given the cluster assignments $C_1, \ldots, C_n$ and the data $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$, $i = 1, \ldots, n$.

Given the quantities in equation (1) and (2)–(3) and the data $\{\mathbf{y}_i, \mathcal{L}(A_i)\}$, $i = 1, \ldots, n$, the second part of the MCMC routine proceeds by updating the cluster assignments $C_1, \ldots, C_n$ via the sequential reseating procedure that is outlined in Table 2.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2008) Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, **9**, 1981–2014.

Aldous, D. (1985) Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII 1983* (ed. P. Hennequin), pp. 1–198. Berlin: Springer.

Azzalini, A. and Scarpa, B. (2012) *Data Analysis and Data Mining: an Introduction*. Oxford: Oxford University Press.

Banerjee, A., Murray, J. and Dunson, D. B. (2013) Bayesian learning of joint distributions of objects. *J. Mach. Learn. Wrkshps Proc.*, **31**, 1–9.

Bhattacharya, A. and Dunson, D. B. (2011) Sparse Bayesian infinite factor models. *Biometrika*, **98**, 291–306.

Bigelow, J. L. and Dunson, D. B. (2009) Bayesian semiparametric joint models for functional predictors. *J. Am. Statist. Ass.*, **104**, 26–36.

Dunson, D. B., Herring, A. H. and Siega-Riz, A. M. (2008) Bayesian inference on changes in response densities over predictor clusters. *J. Am. Statist. Ass.*, **103**, 1508–1517.

Durante, D. and Dunson, D. B. (2015) Bayesian inference and testing of group differences in brain networks. *Preprint arXiv:1411.6506.* Department of Statistical Sciences, University of Padova, Padova.

Durante, D., Dunson, D. B. and Vogelstein, J. T. (2015) Nonparametric Bayes modeling of populations of networks. *Preprint arXiv:1406.7851*. Department of Statistical Sciences, University of Padova, Padova.

Escobar, M. D. and West, M. (1995) Bayesian density estimation and inference using mixtures. *J. Am. Statist. Ass.*, **90**, 577–588.

Fruchterman, T. M. and Reingold, E. M. (1991) Graph drawing by force-directed placement. *Softwr. Pract. Exper.*, **21**, 1129–1164.

Gershman, S. J. and Blei, D. M. (2012) A tutorial on Bayesian nonparametric models. *J. Math. Psychol.*, **56**, 1–12.

Griffiths, T. L. and Ghahramani, Z. (2011) The indian buffet process: an introduction and review. *J. Mach. Learn. Res.*, **12**, 1185–1224.

Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. (2010) *Bayesian Nonparametrics*. Cambridge: Cambridge University Press.

Hoff, P. D. (2008) Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems*, vol. 20 (eds J. Platt, D. Koller, Y. Singer and S. Roweis), pp. 657–664. Cambridge: MIT Press.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.

Kaishev, V. K., Nielsen, J. P. and Thuring, F. (2013) Optimal customer selection for cross-selling of financial services products. *Exprt Syst. Appl.*, **40**, 1748–1757.

Kamakura, W. A., Ramaswami, S. N. and Srivastava, R. K. (1991) Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *Int. J. Res. Marktng*, **8**, 329–349.

Kamakura, W. A., Wedel, M., de Rosa, F. and Mazzon, J. A. (2003) Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction. *Int. J. Res. Marktng*, **20**, 45–65.

Lau, J. W. and Green, P. J. (2007) Bayesian model-based clustering procedures. *J. Computnl Graph. Statist.*, **16**, 526–558.

Matiş, C. and Ilieş, L. (2014) Customer relationship management in the insurance industry. *Proc. Econ. Finan.*, **15**, 1138–1145.

Medvedovic, M., Yeung, K. Y. and Bumgarner, R. E. (2004) Bayesian mixture model based clustering of replicated microarray data. *Bioinformatics*, **20**, 1222–1232.

Neal, R. M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *J. Computnl Graph. Statist.*, **9**, 249–265.

Nowicki, K. and Snijders, T. A. B. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.

Polson, N. G., Scott, J. G. and Windle, J. (2013) Bayesian inference for logistic models using Pólya–gamma latent variables. *J. Am. Statist. Ass.*, **108**, 1339–1349.

Rousseau, J. and Mengersen, K. (2011) Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Statist. Soc.* B, **73**, 689–710.

Stephens, M. (2000) Dealing with label switching in mixture models. *J. R. Statist. Soc.* B, **62**, 795–809.

Thuring, F. (2012) A credibility method for profitable cross-selling of insurance products. *Ann. Act. Sci.*, **6**, 65–75.

Thuring, F., Nielsen, J., Guillén, M. and Bolancé, C. (2012) Selecting prospects for cross-selling financial products using multivariate credibility. *Exprt Syst. Appl.*, **39**, 8809–8816.
Verhoef, P. C. and Donkers, B. (2001) Predicting customer potential value: an application in the insurance industry. *Decsn Supprt Syst.*, **32**, 189–199.