Bayesian modeling of networks in complex business intelligence problems

Sally Paganin* Daniele Durante* Bruno Scarpa* David B. Dunson†
sallypaganin@gmail.com
University of Padua* Italy. Duke University†, Durham, USA.
Università degli Studi di Padova

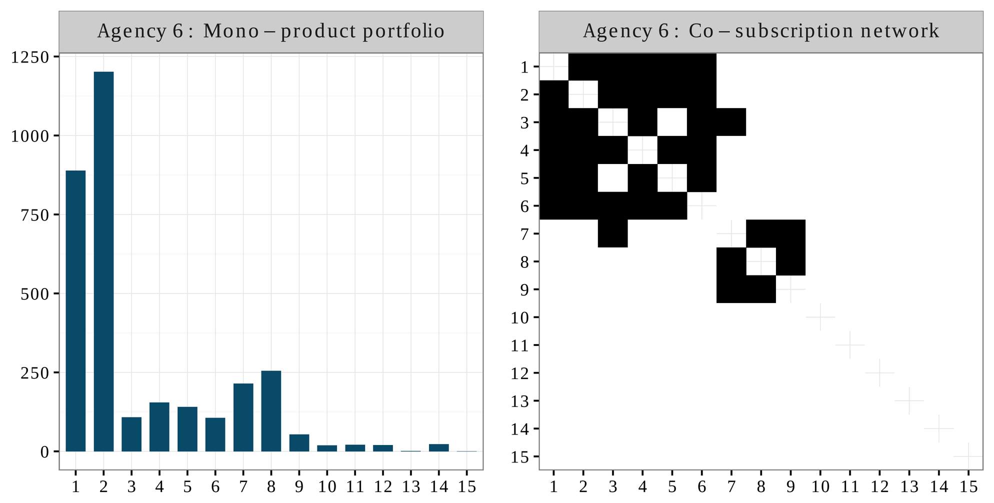
[1] Introduction

- **Marketing problem:** define effective cross-selling campaigns on the basis of customers behavior.
- **Motivation:** targeting the existing customer base with effective strategies leads to the growth of the company.

Data

An insurance company administrates $n = 130$ agencies selling the same $V = 15$ products. For each agency i is available:

- Number of mono-product customers n_{iv} subscribed to each product $v = 1, \dots, V$.
- Co-subscription network A_i among products, relative to multi-product customers.



Aim: Define cross-selling campaigns for mono-product customers exploiting co-subscription networks.

[2] Cross-sell strategy

A **strategy** is the best product u to offer to a mono-product customer with a product $v \neq u$.

Shared strategies among similar agencies are better!

- Agencies in the same company may exhibit clusters.
- Using the same strategy reduces administrative overhead.

$\Rightarrow$ We associate agencies to a vector of cluster assignments

$$\boldsymbol{C} = (C_1, \dots, C_n)$$

with $C_i \in \{1, \dots, K\}$ the cluster membership of agency i.

Same cluster $\Rightarrow$ same strategies

Cluster-specific cross-selling

- **The best offer to a mono-product customer** with product v in agencies belonging to cluster k is defined as

$$u_{kv} = \text{argmax}_u \{\Pr(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$$

  - $\mathcal{A}_{k[vu]}$: random variable measuring presence or absence of a co-subscription among products v and u in cluster k.
- **Evaluate the strategies** using a performance indicator

$$e_{kv} = p_{kv} \max\{\Pr(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$$

  - p_{kv}: probability that a mono-product customer is subscribed to v in agencies belonging to cluster k.

Strategies with high e_{kv} are likely to work better!

[3] Joint model for mixed domain data

Cluster dependent probabilistic representation

Number of mono-product customers $\Rightarrow$ discrete distribution

$$(n_{i1}, \dots, n_{iV}) \sim \text{Multinom}\{n_i; p_{k1}, \dots, p_{kV}\}$$

Co-subscription networks $\Rightarrow$ mixture of latent eigenmodels

Introduce an index $G_i \in \{1, \dots, H\}$ for the mixture component.

- Edges between pairs of products $l = (u, v)$ are realizations from

$$\mathcal{L}(\mathcal{A}_i)_l | \pi_{il} \overset{\text{indep}}{\sim} \text{Bern}(\pi_{il}), \quad l = 1, \dots, V(V-1)/2$$

- Probability of connection changes across the mixture components

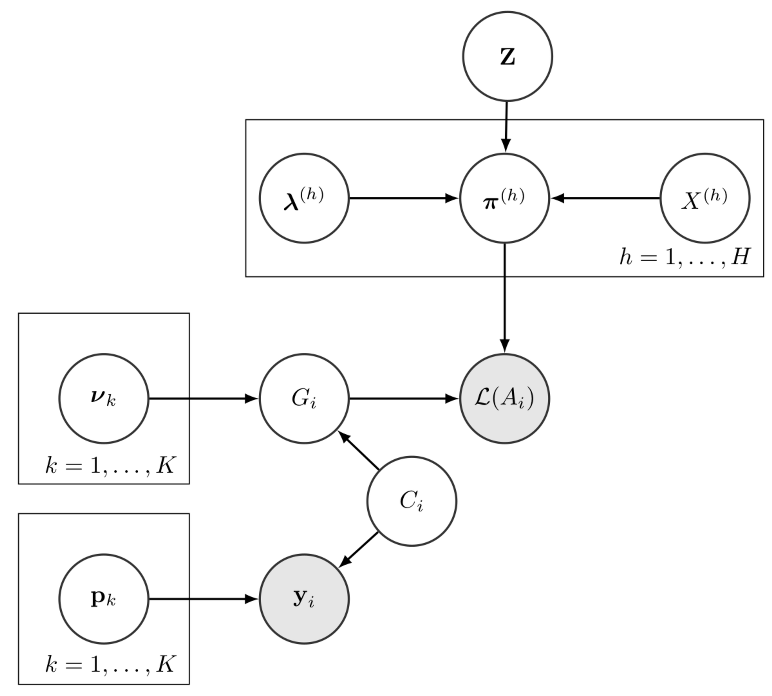
$$\boldsymbol{\pi}_i | G_i = h, \boldsymbol{\pi}^{(h)} = \boldsymbol{\pi}^{(h)}$$

- Mixing probabilities depends on the cluster allocation

$$\Pr(G_i = h | C_i = k) = \nu_{hk} \quad h = 1, \dots, H, \quad k = 1, \dots, K$$

allowing **sharing of mixture components across the clusters.**

Graphical model



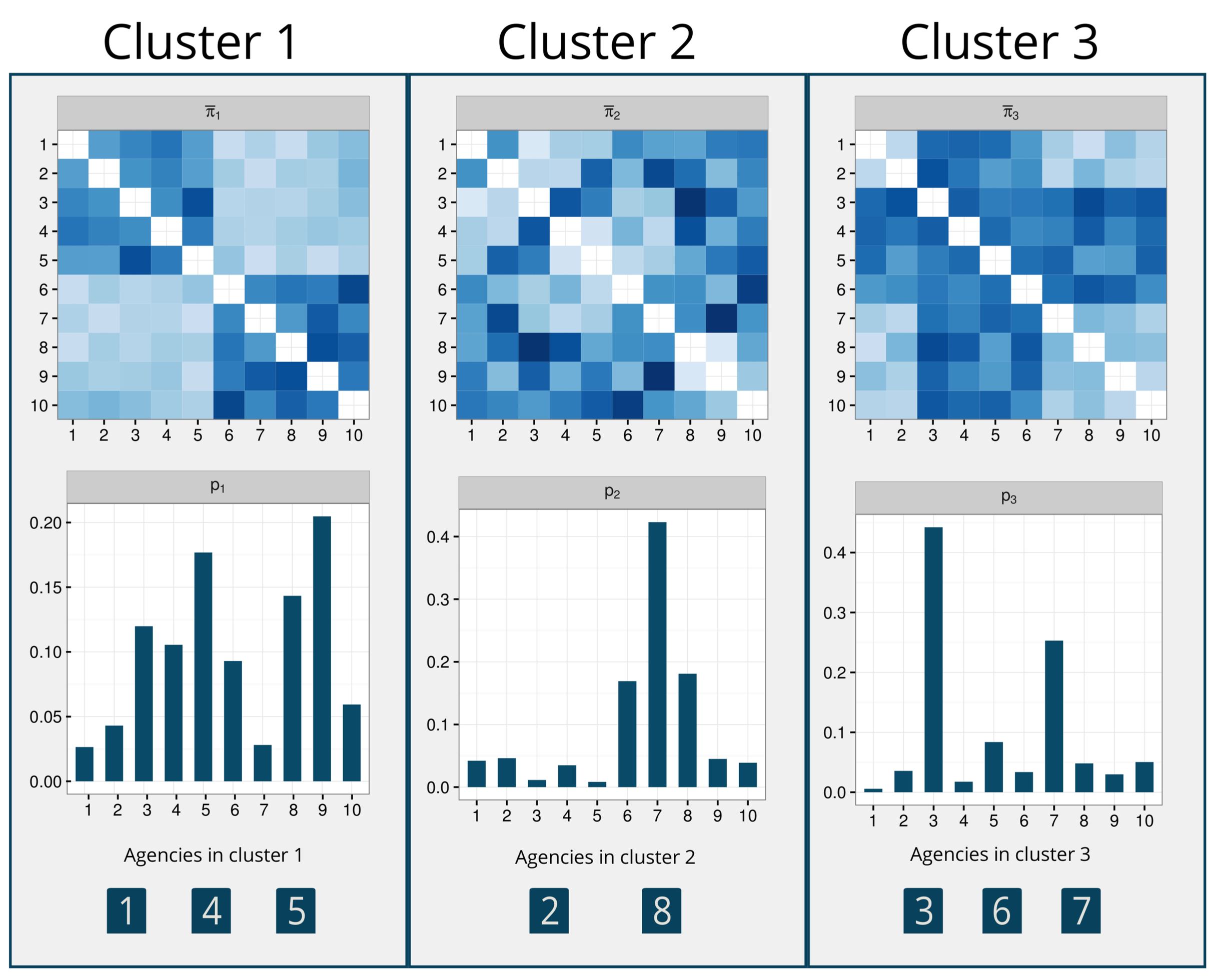
Component-specific **edge probability vector** $\Rightarrow$ characterized via **matrix factorization representation**

$$\boldsymbol{\pi}^{(h)} = [1 + \exp\{-\boldsymbol{Z} - \boldsymbol{D}\}]^{-1} \quad \boldsymbol{D} = \mathcal{L}(\boldsymbol{X}^{(h)} \boldsymbol{\Lambda}^{(h)} \boldsymbol{X}^{(h)T})$$

- $\boldsymbol{Z}$ **common similarity** vector shared among all the co-subscription networks $\Rightarrow$ centers the mixture components
- $\boldsymbol{D}$ **component-specific** vector which accounts for pairwise similarity among products in a latent space ($R \ll V$)
  - $\boldsymbol{X}^{(h)}$ products' latent coordinates $\Rightarrow$ Products with coordinates in the same direction are more likely to be co-subscribed than products with coordinates in the opposite direction.
  - $\boldsymbol{\Lambda}^{(h)} = \text{diag}(\boldsymbol{\lambda}^{(h)})$ weights the similarity in each dimension r by a non-negative parameter $\lambda_r^{(h)}$.

Output from the model

Example of output from our model for decision making in marketing when there are $n = 8$ agencies and $K = 3$ clusters.



Interpretation

- Agencies 1, 4 and 5 share common profiles of mono-product customer choices and co-subscription networks ($C_1 = C_4 = C_5 = 1$).
- **Co-subscription behavior** is summarized by the expectation of the network-valued variable associated to cluster 1; results in [2] show that, under the hiererchical representation, it corresponds to

$$\text{E}\{\mathcal{L}(\mathcal{A}_1)\} = \bar{\boldsymbol{\pi}}_1 = \sum_{h=1}^{H} \nu_{h1} \boldsymbol{\pi}^{(h)}$$

which coincides with the vector of co-subscription probabilities for pairs of products in cluster 1.

$\Rightarrow$ $\bar{\boldsymbol{\pi}}_1$ **defines the set of cross-selling strategies.**

- **Mono-product customer choices** in cluster 1 are characterized by the vector $\boldsymbol{p}_1$.

$\Rightarrow$ $\boldsymbol{p}_1$ provides **additional information** to obtain the **performance indicators** for agencies in cluster 1.

The **performance indicator** is estimated from the model as $\hat{e}_{kv} = \hat{p}_{kv} \max\{\hat{\Pr}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$ for each product v and cluster k, with $\{\hat{\Pr}(\mathcal{A}_{k[vu]} = 1) : u \neq v\}$ available from $\hat{\bar{\pi}}_{kl}$, where l denotes the pair of products (v, u).

$\Rightarrow$ The **conditional distribution of the co-subscription network** varies as a function of a **latent clustering variable**, which is determined **endogenously** by mono-product customer choices and co-subscription patterns shared across subset of agencies.

[4] Bayesian inference

Prior specification

- **Number of clusters** K **unknown** $\Rightarrow$ Cluster assignment vector $\boldsymbol{C} \sim \text{CRP}(\alpha_c)$
- Mono-product customer choices $\Rightarrow$ $\boldsymbol{p}_k \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_V)$
- Priors for co-subscription networks are chosen to **maintain computational tractability** and **favor deletion of redundant dimensions**; in particular
  - Cluster-specific **mixing probability vector** $\Rightarrow$ $\boldsymbol{\nu}_k \sim \text{Dirichlet}(1/H, \dots, 1/H)$ for $k = 1, \dots, K$, favor deletion of unnecessary mixture components.
  - Shrinkage prior [1] on **latent weights** $\Rightarrow$ $\boldsymbol{\lambda}^{(h)} = (\lambda_1^{(h)}, \dots, \lambda_R^{(h)}) \sim \text{MIG}(a_1, a_2)$ which induces priors on the elements $\lambda_r^{(h)}$ for $r = 1, \dots, R$ that are increasingly concentrated around 0 as r increases

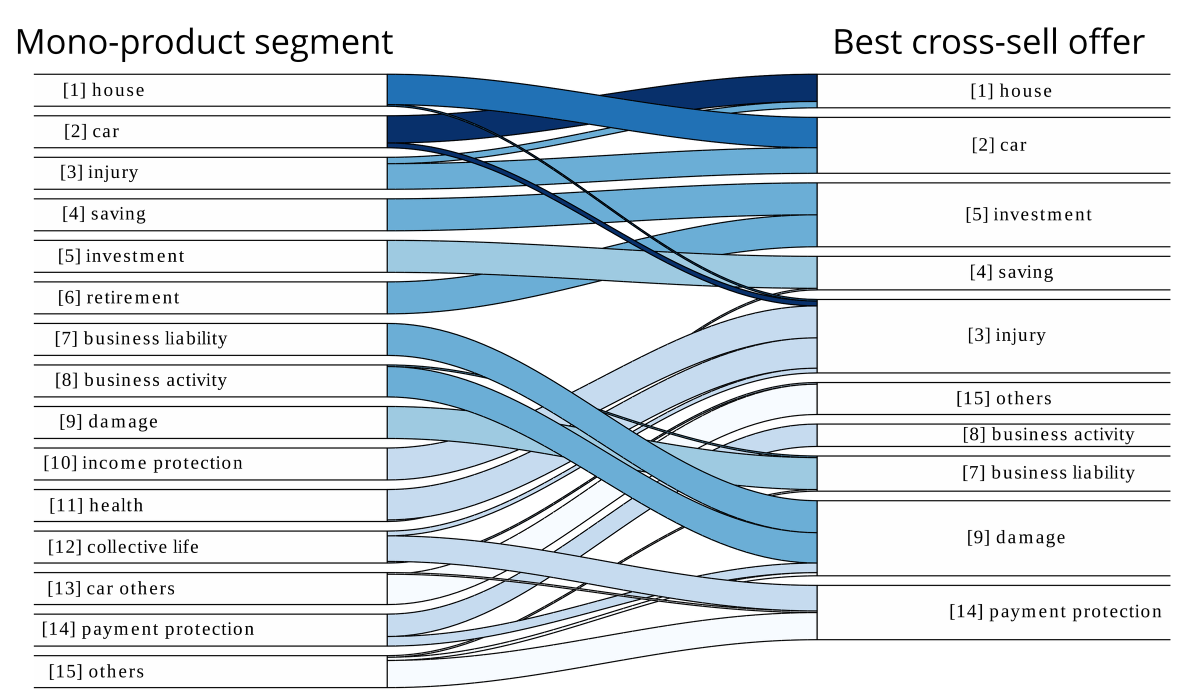
Posterior computation

Posterior distributions are obtained via **Gibbs sampling**, with key steps

- Agencies are allocated to clusters via a sequential re-seating procedure [3] $\Rightarrow$ **cluster assignment vector estimated as MAP**.
- Updating of the parameters associated to the co-subscription networks follow steps in [2] which exploits **Pólya-Gamma data augmentation for Bayesian logistic regression**.

[5] Cross-selling in Italian insurance company

Cluster-specific cross-selling strategies



20 clusters, with **similar behavior** summarized in the riverplot

Edge dimension is proportional to the number of clusters for which the best offer coincides with the indicated one.

Edge color gradient represents strategy performance indicator averaged across clusters.

Right box dimension is proportional to the number of strategies suggesting that product, as the best offer.

References

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

[2] Durante, D., Dunson, D. B. and Vogelstein, J. T. (2015) Nonparametric Bayes modeling of populations of networks. *ArXiv:1406.7851 e-prints*.

[3] Neal, R. M. (2000) Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Participation in this conference was supported by Google "Google-ISBA junior travel support grant"