

Modèles de blocs latents semi-supervisés par champ de Markov cachés

Journées AFIA SFC 2021

Paul Riverain^{1, 2}

Simon Fossier¹

Mohamed Nadif²

¹Thales

²Université de Paris

21/09/21

- 1 Contexte
- 2 Motivations
- 3 Modèle
- 4 Algorithmes
- 5 Expériences
- 6 Application : Classification de réseaux attribués

- 1 Contexte
- 2 Motivations
- 3 Modèle
- 4 Algorithmes
- 5 Expériences
- 6 Application : Classification de réseaux attribués

Contexte

Supervision réseau de transport



Figure: Un poste de contrôle centralisé pour une ligne de métro

Objectif général :

Résumer les flux de passagers sur le réseau avec des groupes homogènes de stations : classification non-supervisée.

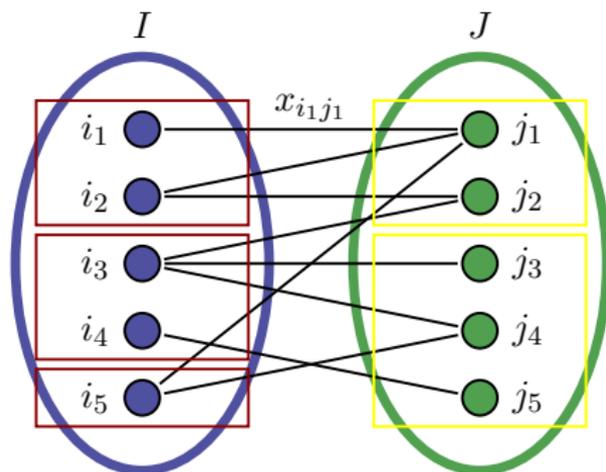
⇒ On ne présente à l'opérateur que les interactions entre groupes

Co-clustering

Clustering des nœuds d'un graphe biparti

Flux passagers dans un réseau sur une période donnée :

- $i \in I$ station de départ
- $j \in J$ station d'arrivée
- x_{ij} nombre de trajets en cours de i à j sur la période donnée
- \mathbf{X} décrit les flux passager (entrée/sortie) sur le réseau



Objectif technique

On distingue les stations d'entrée et de sortie : on cherche à faire du co-clustering sur le graphe des flux passager

- 1 Contexte
- 2 Motivations**
- 3 Modèle
- 4 Algorithmes
- 5 Expériences
- 6 Application : Classification de réseaux attribués

Motivations

Introduction de connaissance experte

Objectif de ces travaux

Permettre à l'opérateur d'exprimer de façon simple de la connaissance a priori sur les données et de l'introduire dans l'algorithme

Applications

- L'opérateur s'approprie l'algorithme de co-clustering en confrontant ses a priori sur les données avec les sorties de l'algorithme
- Permet de d'adapter marginalement un modèle de co-clustering pour qu'il offre des partitions plus souhaitables pour l'opérateur

- 1 Contexte
- 2 Motivations
- 3 Modèle**
- 4 Algorithmes
- 5 Expériences
- 6 Application : Classification de réseaux attribués

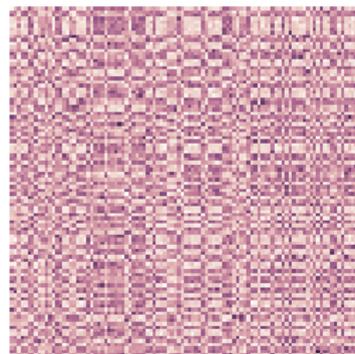
Co-clustering

Matrice de données $\mathbf{X} = (x_{ij})$ de taille $n \times d$, n individus, d attributs.

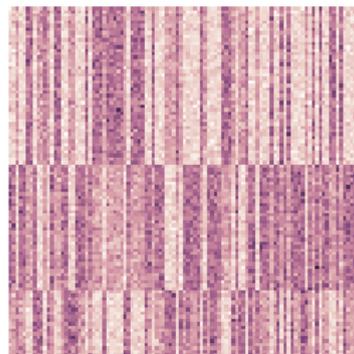
Co-clustering :

Classification simultanée des lignes et des colonnes de \mathbf{X}

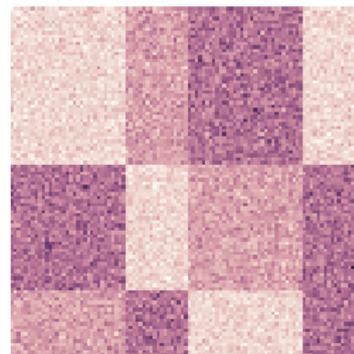
⇒ Classification et réduction de dimensionnalité *simultanément*.



(a) $\mathbf{X} = (x_{ij})$



(b) \mathbf{X} réorganisée par clusters d'individus



(c) \mathbf{X} réorganisée par clusters d'individus et d'attributs

Modèle de blocs latents

Un modèle probabiliste pour le Co-clustering

Latent Block Model (LBM) (Govaert and Nadif 2003)

- x_{ij} échantillonnés à partir d'une distribution paramétrique ϕ
- x_{ij} dépend du cluster de ligne z_i et du cluster de colonne w_j
 - g clusters en ligne et m clusters en colonne
 - $z_i = k \Leftrightarrow z_{ik} = 1$, idem $w_j = \ell \Leftrightarrow w_{j\ell} = 1$

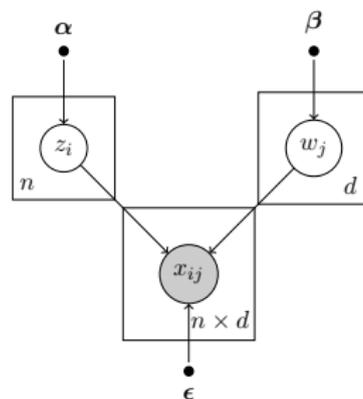
$$z_i \stackrel{iid}{\sim} \text{Categorielle}(\alpha_1, \dots, \alpha_g)$$

$$w_j \stackrel{iid}{\sim} \text{Categorielle}(\beta_1, \dots, \beta_m)$$

$$x_{ij} \stackrel{iid}{\sim} \phi(\cdot; \epsilon_{ij} z_i w_j)$$

Paramètres du modèle :

$$\Theta = \{\alpha, \beta, \epsilon\}$$



$$P(\mathbf{X}|\mathbf{Z}, \mathbf{W}; \Theta) = \prod_{ij} \phi(x_{ij}; \epsilon_{ijz_i w_j}) = \prod_{ijkl} \phi(x_{ij}; \epsilon_{ijkl})^{z_{ik} w_{jl}}.$$

Modèle de blocs latents

Inférence et apprentissage

$$\log P(\overbrace{\mathbf{X}}^{\text{observé}}, \overbrace{\mathbf{Z}, \mathbf{W}}^{\text{latent}}; \overbrace{\Theta}^{\text{à apprendre}}) = \sum_{ik} z_{ik} \log \alpha_k + \sum_{jl} w_{jl} \log \beta_l \\ + \sum_{ijkl} z_{ik} w_{jl} \log \phi(x_{ij}; \epsilon_{ijkl}).$$

Maximum de vraisemblance:

$$P(\mathbf{X}; \Theta) = \sum_{\mathbf{Z}, \mathbf{W} \in \mathcal{Z} \times \mathcal{W}} P(\mathbf{X}, \mathbf{Z}, \mathbf{W}; \Theta)$$

$$\text{Mais } |\mathcal{Z} \times \mathcal{W}| = g^n \times m^d$$

Modèle de blocs latents

Inférence et apprentissage

$$\log P(\overbrace{\mathbf{X}}^{\text{observé}}, \overbrace{\mathbf{Z}, \mathbf{W}}^{\text{latent}}; \overbrace{\Theta}^{\text{à apprendre}}) = \sum_{ik} z_{ik} \log \alpha_k + \sum_{j\ell} w_{j\ell} \log \beta_\ell + \sum_{ijkl} z_{ik} w_{j\ell} \log \phi(x_{ij}; \epsilon_{ijkl}).$$

EM classique

La fonction auxiliaire $Q(\Theta, \Theta^{(c)}) = \mathbb{E}(\log P(\mathbf{X}, \mathbf{Z}, \mathbf{W}; \Theta) | \mathbf{X}, \Theta^{(c)})$ requiert $P(z_{ik} w_{j\ell} = 1 | \mathbf{X}, \Theta^{(c)})$, qui n'est *pas tractable*, contrairement à un mélange de modèles classique.

⇒ **Variantes EM** : EM Variationnel, EM Classifiant, EM Stochastique

Objectif de ces travaux

Permettre à l'opérateur d'exprimer de façon simple de la connaissance a priori sur les données et de l'introduire dans l'algorithme

Approche

- Rendre les variables latentes (z_i) (resp. (w_j)) a priori dépendantes les unes des autres
- L'utilisateur détermine les dépendances entre les variables et rend certaines partitions (Z, W) plus ou moins probables.

⇒ Utilisation d'un champ de Markov caché (Hidden MRF : HMRF)

- Segmentation d'images non supervisée/classification avec contraintes spatiales avec mélanges de modèles
 - Neighborhood EM (Ambroise and Govaert 1998)
 - HMRF + Mean field (Celeux, Forbes, and Peyrard 2003)

- Segmentation d'images non supervisée/classification avec contraintes spatiales avec mélanges de modèles
 - Neighborhood EM (Ambroise and Govaert 1998)
 - HMRF + Mean field (Celeux, Forbes, and Peyrard 2003)
- Classification semi-supervisée ou sous contraintes
 - Constrained k -means (K. Wagstaff et al. 2001)
 - HMRF k -means (Basu, Bilenko, and Mooney 2004)

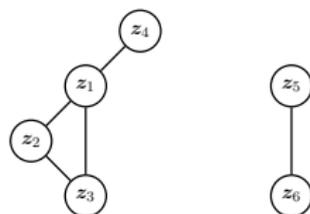
- Segmentation d'images non supervisée/classification avec contraintes spatiales avec mélanges de modèles
 - Neighborhood EM (Ambroise and Govaert 1998)
 - HMRF + Mean field (Celeux, Forbes, and Peyrard 2003)
- Classification semi-supervisée ou sous contraintes
 - Constrained k -means (K. Wagstaff et al. 2001)
 - HMRF k -means (Basu, Bilenko, and Mooney 2004)
- Régularisation Laplacienne
 - Modèles de mélange + régularisation Laplacienne (Zhu and Lafferty 2005)
 - Régularisation Laplacienne issue d'information auxiliaire (Salah and Nadif 2017)

Contraintes :

- Pas de connaissance sur tous les individus \implies supervision partielle
- Pas de vérité terrain \implies pas de supervision directe
- Interactions avec l'algorithme \implies certaines contraintes peuvent ne pas être respectées

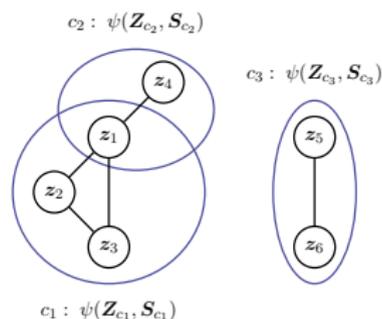
Champ de Markov : généralités

Un graphe \mathcal{G} non orienté décrit les dépendances d'un ensemble de variables aléatoires. Le nœud i dépend de ses voisins directs $i' \in \mathcal{N}_i$



Champ de Markov : généralités

Un graphe \mathcal{G} non orienté décrit les dépendances d'un ensemble de variables aléatoires. Le nœud i dépend de ses voisins directs $i' \in \mathcal{N}_i$



$$P(\mathbf{Z}; \mathbf{S}) = \Gamma(\mathbf{S})^{-1} \prod_{c \in \mathcal{C}} \psi(\mathbf{Z}_c; \mathbf{S}_c)$$

- $\Gamma(\mathbf{S})$ fonction de partition du MRF, assurant $\sum_{\mathbf{Z} \in \mathcal{Z}} P(\mathbf{Z}; \mathbf{S}) = 1$.
- \mathbf{Z} la matrice des attributs des nœuds du MRF
- \mathbf{Z}_c la matrice \mathbf{Z} restreinte aux nœuds de la clique c
- $\psi(\mathbf{Z}_c; \mathbf{S}_c) > 0$ le potentiel de la clique c

Champ de Markov : généralités

Exemple du modèle d'Ising : potentiels de noeuds et d'arêtes

Dipôles magnétiques à états $z_i \in \{-1, +1\}$

\mathcal{G} est un treillis 2D

Le MRF est défini par paires

Potentiel de l'arête :

$$\psi_{ij}(z_i, z_j) = \begin{cases} \exp(J) & \text{si } z_i = z_j \\ \exp(-J) & \text{si } z_i \neq z_j \end{cases}$$

J constante de couplage

Champ magnétique externe h_i qui s'applique au dipôle i

Potentiel du dipôle : $\psi_i(z_i) = \exp(h_i z_i)$

$h_i > 0$: le dipôle i veut s'aligner avec la direction $+1$.

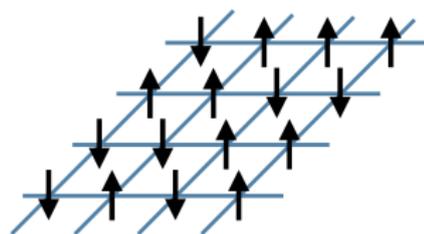


Figure: Modèle d'Ising

Champ de Markov : généralités

Exemple du modèle de Potts : le couplage entraîne des phénomènes complexes

Similaire à un modèle d'Ising, mais chaque z_i prend ses valeurs dans $\{1, \dots, K\}$

$$\psi_{ij}(z_i, z_j) = \begin{cases} \exp(J) & \text{si } z_i = z_j \\ 1 & \text{sinon} \end{cases}$$

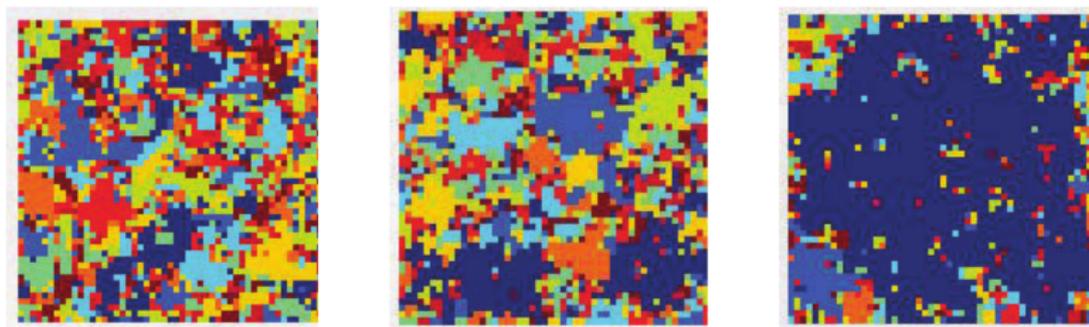
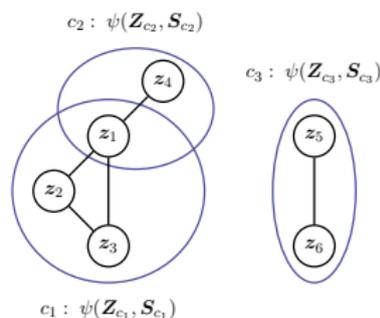


Figure: Échantillons d'un modèle de Potts à $K = 10$ états pour différentes valeurs de J . De gauche à droite : $J = 1.42$, $J = 1.44$, $J = 1.46$. (D'après Sudderth)

MRF cachés pour le LBM: HLBM

Un graphe \mathcal{G} décrit les dépendances d'un ensemble de variables aléatoires.
Le nœud i dépend de ses voisins directs $i' \in \mathcal{N}_i$



En général :

- le graphe des dépendances est donné
- les $(z_i)_i$ sont observées
- On cherche à estimer les paramètres \mathbf{S}

Dans notre cas :

- le graphe des dépendances \mathcal{G} est donné
- les $(z_i)_i$ sont inférées
- Les paramètres \mathbf{S} donnés

HMRFs pour le LBM: HLBM

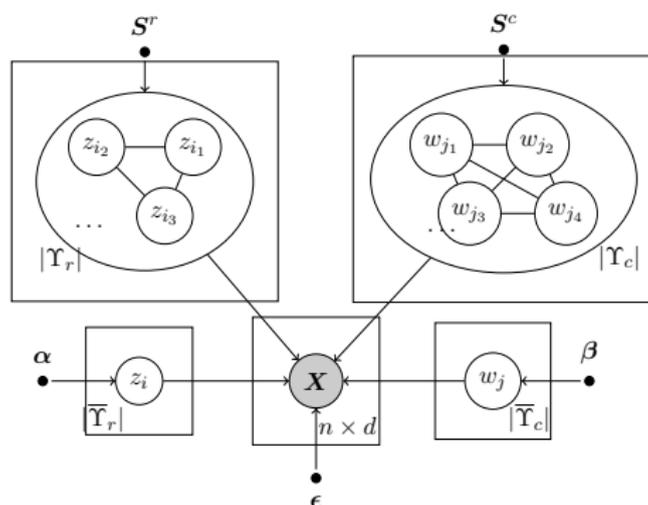


Figure: Modèle graphique du HLBM

$$P(\mathbf{Z}; \Theta) = \underbrace{\Gamma(\Xi^r)^{-1}}_{\text{Normalisation}} \times \underbrace{\prod_{i \in \bar{\mathcal{Y}}_r} \prod_k \alpha_k^{z_{ik}}}_{\text{Poids de mélange hors du MRF}} \times \underbrace{\prod_{\substack{i \sim i' \\ r}} \psi_{ii'}^r(\mathbf{z}_i, \mathbf{z}_{i'}; \Xi^r)}_{\text{Potentiels MRF}}$$

Définir les potentiels :

- Must Link (\mathcal{M})

- Cannot Link (\mathcal{C})

$$\psi_{ii'}^r(\mathbf{z}_i, \mathbf{z}_{i'}; \Xi^r) = \begin{cases} \exp(-\lambda_r \xi_{ii'}^r \mathbb{1}(z_{i'} \neq z_i)) & (i, i') \in \mathcal{M}^r \\ \exp(-\lambda_r \xi_{ii'}^r \mathbb{1}(z_{i'} = z_i)) & (i, i') \in \mathcal{C}^r. \end{cases}$$

On peut montrer que : $\Gamma = \Gamma(\Xi^r)$

Reparamétrisation :

$$s_{ii'}^r = \begin{cases} \xi_{ii'}^r & (i, i') \in \mathcal{M}^r \\ -\xi_{ii'}^r & (i, i') \in \mathcal{C}^r \\ 0 & \text{otherwise.} \end{cases}$$

Pour tout (i, i') ,

$$\psi_{ii'}^r(\mathbf{z}_i, \mathbf{z}_{i'}; \mathbf{S}^r) \propto \exp(\lambda_r s_{ii'}^r \sum_k z_{ik} z_{i'k})$$

$$\begin{aligned}\log P(\mathbf{X}, \mathbf{Z}, \mathbf{W}; \Theta) = & \\ & \sum_{i \in \bar{\mathcal{Y}}_r} \sum_k z_{ik} \log \alpha_k + \sum_{j \in \bar{\mathcal{Y}}_c} \sum_{\ell} w_{j\ell} \log \beta_{\ell} \\ & + \frac{\lambda_r}{2} \sum_{ii'k} s_{ii'}^r z_{ik} z_{i'k} + \frac{\lambda_c}{2} \sum_{jj'\ell} s_{jj'}^c w_{j\ell} w_{j'\ell} \\ & + \sum_{ijkl} z_{ik} w_{j\ell} \log \phi(x_{ij}; \epsilon_{ijkl}) + \text{cste}\end{aligned}$$

Difficulté

Nouvelles dépendances $z_{ik}z_{i'k}$ et $w_{j\ell}w_{j'\ell}$

- 1 Contexte
- 2 Motivations
- 3 Modèle
- 4 Algorithmes**
- 5 Expériences
- 6 Application : Classification de réseaux attribués

2 approches

- Variational EM (VEM) : Construire une approximation tractable de la distribution postérieure sur les partitions $P(\mathbf{Z}, \mathbf{W} | \mathbf{X}; \Theta)$.
- Classification EM : à chaque étape de EM, affectations fermes (\sim k -means). Convergence rapide et matrices creuses.

Par rapport au LBM

Les étapes M sont similaires et ne posent pas de problème

Seules les étape E sont distinctes (\leftarrow dépendances $z_{ik}z_{i'k}$ et $w_{j\ell}w_{j'\ell}$)

Algorithme CEM pour le HLBM

Étape CE

Iterated Conditional Modes (ICM), algorithme de montée de coordonnées pour maximiser $\log P(\mathbf{X}, \mathbf{Z}, \mathbf{W}; \Theta)$ par rapport à \mathbf{Z} ($z_i \in \{1, \dots, g\}$)

Pour i dans le MRF, étape CE :

$$z_i = \operatorname{argmax}_{z_i} \log P(\mathbf{X}, z_i, (z_{i'})_{i' \neq i}, \mathbf{W}; \Theta)$$

Finalement,

$$z_i = \operatorname{argmax}_k \begin{cases} \lambda_r \sum_{i'} s_{ii'}^r z_{i'k} + \sum_{j\ell} w_{j\ell} \log \phi(x_{ij}; \epsilon_{ijkl}) & i \in \Upsilon_r \\ \log \alpha_k + \sum_{j\ell} w_{j\ell} \log \phi(x_{ij}; \epsilon_{ijkl}) & i \in \bar{\Upsilon}_r. \end{cases}$$

Algorithme VEM pour le HLBM

Fonction objectif

Soit Q une distribution sur $\mathcal{Z} \times \mathcal{W}$. L'objectif en VEM est de minimiser la "distance" de Q à la vraie distribution postérieure

$$P^*(\mathbf{Z}, \mathbf{W}; \Theta) = P(\mathbf{Z}, \mathbf{W} | \mathbf{X}; \Theta) : \min D_{\text{KL}}(Q || P^*).$$

Comme $D_{\text{KL}}(Q || P^*) \geq 0$, on a une borne inférieure de la log-vraisemblance dont l'optimisation peut être rendue tractable.

$$F(\tilde{\mathbf{Z}}, \tilde{\mathbf{W}}, \Theta) = \mathbb{E}_Q(\log P(\mathbf{X}, \mathbf{Z}, \mathbf{W}; \Theta)) + \mathbb{H}(Q) \leq \log P(\mathbf{X}; \Theta)$$

Algorithme VEM pour le HLBM

Fonction objectif

Soit Q une distribution sur $\mathcal{Z} \times \mathcal{W}$. L'objectif en VEM est de minimiser la "distance" de Q à la vraie distribution postérieure

$$P^*(\mathbf{Z}, \mathbf{W}; \Theta) = P(\mathbf{Z}, \mathbf{W} | \mathbf{X}; \Theta) : \min D_{\text{KL}}(Q || P^*).$$

Comme $D_{\text{KL}}(Q || P^*) \geq 0$, on a une borne inférieure de la log-vraisemblance dont l'optimisation peut être rendue tractable.

$$F(\tilde{\mathbf{Z}}, \tilde{\mathbf{W}}, \Theta) = \mathbb{E}_Q(\log P(\mathbf{X}, \mathbf{Z}, \mathbf{W}; \Theta)) + \mathbb{H}(Q) \leq \log P(\mathbf{X}; \Theta)$$

Q paramétrée par $\tilde{\mathbf{Z}} = (\tilde{z}_{ik})$ and $\tilde{\mathbf{W}} = (\tilde{w}_{j\ell})$, telle que

$$Q(\mathbf{Z}, \mathbf{W}; \tilde{\mathbf{Z}}, \tilde{\mathbf{W}}) = Q(\mathbf{Z}; \tilde{\mathbf{Z}})Q(\mathbf{W}; \tilde{\mathbf{W}}) = \prod_{ik} \tilde{z}_{ik}^{z_{ik}} \prod_{j\ell} \tilde{w}_{j\ell}^{w_{j\ell}}$$

où $\sum_k \tilde{z}_{ik} = 1$ et $Q(z_{ik} = 1) = \tilde{z}_{ik}$

Algorithme VEM pour le HLBM

Etape VE

Dépendances entre les variables latentes voisines dans le HMRF : montée de coordonnées sur : $(\tilde{z}_1, \dots, \tilde{z}_n) \mapsto F(\tilde{\mathbf{Z}}, \tilde{\mathbf{W}}, \Theta)$

Finalement,

$$\tilde{z}_{ik} \propto \begin{cases} \exp(\lambda_r \sum_{i'} s_{ii'}^r \tilde{z}_{i'k}) \prod_{jl} \phi(x_{ij}; \epsilon_{ijkl})^{\tilde{w}_{jl}} & i \in \Upsilon_r \\ \alpha_k \prod_{jl} \phi(x_{ij}; \epsilon_{ijkl})^{\tilde{w}_{jl}} & i \in \bar{\Upsilon}_r. \end{cases}$$

\implies revient à introduire un terme de régularisation $\lambda_r \sum_{i'} s_{ii'}^r \tilde{z}_{i'k}$ à la place des proportions de mélange pour les nœuds dans le HMRF

Inclure les proportions de mélange dans le HMRF

On considère en plus un potentiel de nœud dans le MRF (\Leftrightarrow champ magnétique extérieur)

Etape VE :

$$\tilde{z}_{ik} \propto \alpha_k \exp \left(\lambda_r \sum_{i'} s_{ii'}^r \tilde{z}_{i'k} \right) \prod_{j\ell} \phi(x_{ij}; \epsilon_{ijkl})^{\tilde{w}_{j\ell}}$$

Etape CE :

$$z_i = \operatorname{argmax}_k \left(\log \alpha_k + \lambda_r \sum_{i'} s_{ii'}^r z_{i'k} + \sum_{j\ell} w_{j\ell} \log \phi(x_{ij}; \epsilon_{ijkl}) \right)$$

Difficulté

Dépendance fonction de partition $\Gamma = \Gamma(\alpha, \beta)$

Dépendance négligée pour l'étape M de α et β : pas de problème rencontré expérimentalement.

Risque des mises à jour parallèles

Oscillations

En VEM, on applique un *amortissement* $\tilde{z}_{ik}^{(c+\frac{1}{2})}$ probabilité variationnelle postérieure après une étape E

$$\eta \in (0, 1)$$

$$\tilde{z}_{ik}^{(c+1)} = (1 - \eta)\tilde{z}_{ik}^{(c+\frac{1}{2})} + \eta\tilde{z}_{ik}^{(c)}$$

Contrainte

Pas de version stochastique de EM (Celeux, Forbes, and Peyrard 2003)

Dans le LBM classique

- Dépendances étape E lignes (\mathbf{Z}) \Leftrightarrow étape E colonnes (\mathbf{W}) : on ne peut pas initialiser l'algorithme uniquement avec Θ (\neq modèles de mélange)
- $\mathbf{Z}^{(0)} = \text{clustering}(\mathbf{X})$, $\mathbf{W}^{(0)} = \text{clustering}(\mathbf{X}^T)$
- En fonction du modèle :
clustering = *k*-means, spherical *k*-means, ...

Dans HLBM

- Initialisation LBM sur des matrices transformées \mathbf{M}^r et \mathbf{M}^c
- On moyenne les attributs des voisins pour les relations ML uniquement : $m_{ij}^r = x_{ij} + \frac{1}{s_i^{r+}} \sum_{i'} s_{ii'}^{r+} x_{i'j}$
 \implies Données moins creuses

Potentiels du MRF :

$$\psi_{ii'}^r(\mathbf{z}_i, \mathbf{z}_{i'}; \mathbf{S}^r) \propto \exp(\lambda_r s_{ii'}^r \sum_k z_{ik} z_{i'k})$$

$$\psi_{jj'}^c(\mathbf{w}_j, \mathbf{w}_{j'}; \mathbf{S}^c) \propto \exp(\lambda_c s_{jj'}^c \sum_\ell w_{j\ell} w_{j'\ell})$$

Difficulté expérimentale : choix des facteurs d'échelle λ_r, λ_c

Compromis entre le poids du prior et de la vraisemblance.

Revient à déterminer la constante de couplage J dans un modèle de Potts

- 1 Contexte
- 2 Motivations
- 3 Modèle
- 4 Algorithmes
- 5 Expériences**
- 6 Application : Classification de réseaux attribués

Fixer le poids des contraintes

Sensibilité aux paramètres

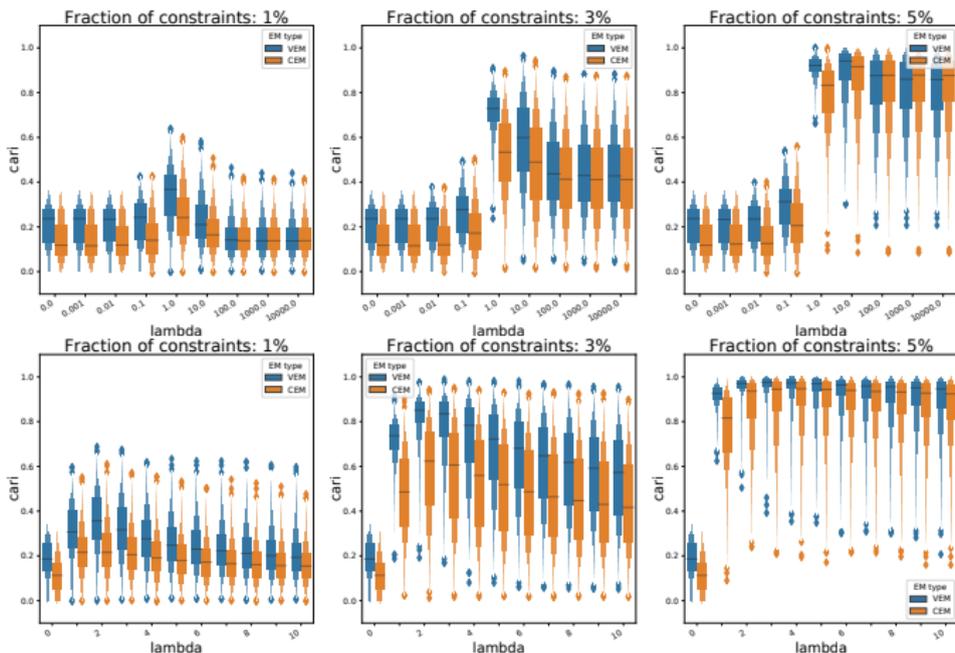


Figure: Distribution des valeurs de CARI

Fixer le poids des contraintes

En fonction du bruit

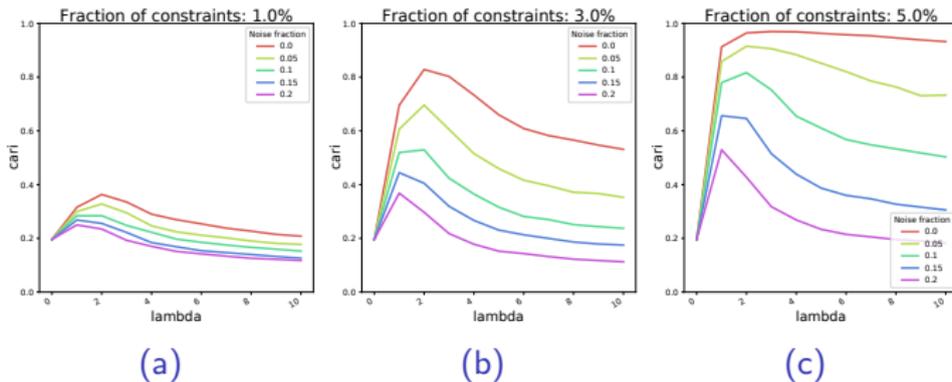


Figure: CARI median pour différentes valeurs de bruit dans les matrices de contraintes S^r and S^c

Appliquer une fermeture transitive?

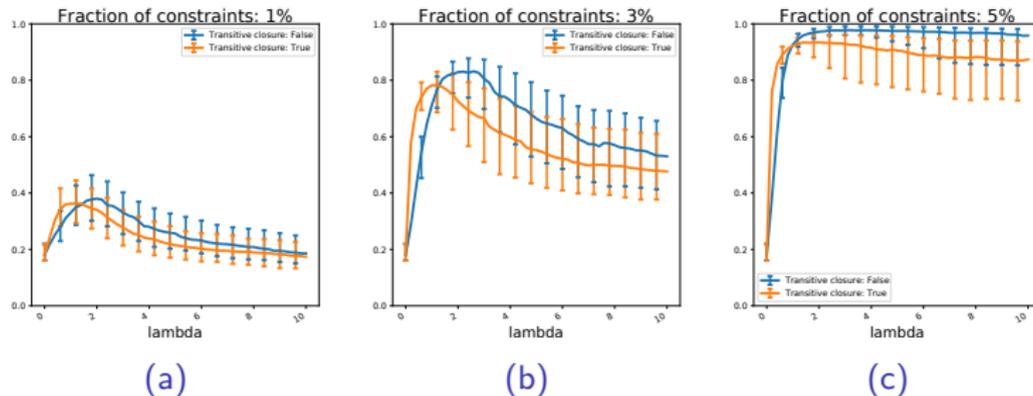


Figure: CARI avec et sans fermeture transitive

- 1 Contexte
- 2 Motivations
- 3 Modèle
- 4 Algorithmes
- 5 Expériences
- 6 Application : Classification de réseaux attribués

Caractéristiques des jeux de données

Réseaux attribués de la forme (\mathbf{A}, \mathbf{X})

$a_{ii'}$: nombre de fois où papier i cite papier i'

\mathbf{x}_i attributs bag-of-words issus de l'abstract du papier i

Classes : domaine de recherche du papier

Hypothèse

Des nœuds voisins dans le graphe ont plus de chances d'être dans le même cluster

$$\mathbf{A} \rightsquigarrow \mathbf{S}$$

Datasets	Normalization \mathbf{X}	Type \mathbf{A}	n	d	g	$n_{\text{NZX}}(\%)$	#Edges	Balance (%)
Cora	binary	binary	2708	1433	7	98.73	5294	22
Citeseer	binary	binary	3312	3703	6	99.14	4732	35
Wiki	tfidf	weighted	2405	4973	17	86.46	17981	2
Pubmed	tfidf	binary	19717	500	3	89.98	44338	52

Approche :

$$\epsilon_{ijkl} = \mu_i \nu_j \gamma_{kl}$$

ϕ : pmf de Poisson

Marges μ_i et ν_j pour absorber les degrés

$$x_{ij} \stackrel{iid}{\sim} \phi(\cdot; \mu_i \nu_j \gamma_{kl})$$

Résultats classification réseaux attribués

Métriques

	Cora		Citeseer		Wiki		Pubmed	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
GAE(*)	53.25	40.69	41.26	18.34	17.33	11.93	64.08	22.97
VGAE(*)	55.95	38.85	44.38	22.71	28.67	30.28	65.48	25.09
ARGE	64.00	44.90	57.3	35.0	41.40(*)	39.50(*)	59.12(*)	23.17(*)
ARVGE	63.80	45.00	54.4	26.10	41.55(*)	40.01(*)	58.22(*)	20.62(*)
MGAE	63.43	45.57	63.56	39.75	50.14	47.97	43.88(*)	8.60(*)
DAEGC	70.04	52.8	67.20	39.70	N/A	N/A	67.10	26.60
AGC	68.92	53.68	67.00	41.13	47.65	45.28	69.78	31.59
PHLBMCEM	68.6 ± 1.9	49.8 ± 0.9	66.2 ± 1.9	40.8 ± 1.4	53.3 ± 3.7	51.9 ± 1.3	67.4 ± 0.6	30.8 ± 0.9
PHLBMVEM	65.9 ± 2.6	49.7 ± 1.6	67.6 ± 1.2	42.1 ± 1.3	54.8 ± 2.4	52.2 ± 0.9	67.0 ± 1.6	30.9 ± 0.8

Contributions

- LBM semi-supervisé avec contraintes par paires
- Étude de la sensibilité aux poids des contraintes
- Application à la classification de réseaux attribués

Pistes pour de futurs travaux

- Approches stochastiques (SEM)
- MRF plus complexes (n-uplets)
- Heuristiques/méthodologie et théorie pour déterminer les facteurs d'échelle λ_r et λ_c des contraintes

Références bibliographiques I

- Ambroise and Govaert (Aug. 1998). “Convergence of an EM-Type Algorithm for Spatial Clustering”. In: *Pattern Recogn. Lett.* 19.10, pp. 919–927. ISSN: 0167-8655.
- Basu, Sugato, Mikhail Bilenko, and Raymond J. Mooney (2004). “A Probabilistic Framework for Semi-Supervised Clustering”. In: *SIGKDD*, pp. 59–68.
- Celeux, Gilles, Florence Forbes, and Nathalie Peyrard (2003). “EM procedures using mean field-like approximations for Markov model-based image segmentation”. In: *Pattern Recognition* 36.1, pp. 131–144. ISSN: 0031-3203.
- Davidson, Ian, Kiri L. Wagstaff, and Sugato Basu (2006). “Measuring Constraint-Set Utility for Partitional Clustering Algorithms”. In: *PKDD*. Ed. by Johannes Fürnkranz, Tobias Scheffer, and Myra Spiliopoulou, pp. 115–126.

Références bibliographiques II

- Govaert and Nadif (2003). “Clustering with block mixture models”. In: *Pattern Recognition* 36.2, pp. 463–473.
- Miele, Vincent, Franck Picard, and Stéphane Dray (2014). “Spatially constrained clustering of ecological networks”. In: *Methods in Ecology and Evolution* 5.8, pp. 771–779.
- Salah and Nadif (Sept. 2017). “Social Regularized von Mises—Fisher Mixture Model for Item Recommendation”. In: *Data Min. Knowl. Discov.* 31.5, pp. 1218–1241. ISSN: 1384-5810.
- Wagstaff, Kiri et al. (2001). “Constrained K-Means Clustering with Background Knowledge”. In: *ICML*, pp. 577–584.
- Zhu, Xiaojin and John Lafferty (2005). “Harmonic Mixtures: Combining Mixture Models and Graph-Based Methods for Inductive and Scalable Semi-Supervised Learning”. In: *ICML*, pp. 1052–1059.

Merci pour votre attention

Algorithm 1: PHLBMVEM

Input: Data matrix \mathbf{X} , constraints matrices \mathbf{S}^r and \mathbf{S}^c , number of row and column clusters g and m , damping factor η .

Output: Classification matrices \mathbf{Z}, \mathbf{W} , parameters α, β, γ

Initialization: Initialize $\tilde{\mathbf{Z}}, \tilde{\mathbf{W}}$ and set

$$\alpha_k = \frac{\tilde{z}_{.k}}{n}, \beta_\ell = \frac{\tilde{w}_{. \ell}}{d}, \gamma_{k\ell} = \frac{x_{k\ell}^{\tilde{\mathbf{Z}}\tilde{\mathbf{W}}}}{x_{k.}^{\tilde{\mathbf{Z}}} x_{. \ell}^{\tilde{\mathbf{W}}}}$$

while *Not converged* **do**

Compute $x_{i\ell}^{\tilde{\mathbf{W}}} = \sum_j \tilde{w}_{j\ell} x_{ij}$;

Row VE-step:

- In parallel, $\forall i, \tilde{z}_{ik} \propto \alpha_k \exp(\lambda_r \sum_{i'} s_{ii'}^r \tilde{z}_{i'k} + \sum_\ell x_{i\ell}^{\tilde{\mathbf{W}}} \log \gamma_{k\ell})$;
- Normalize $\tilde{\mathbf{Z}}$ and apply damping;

Row M-step: $\alpha_k = \frac{\tilde{z}_{.k}}{n}, \gamma_{k\ell} = \frac{\sum_i \tilde{z}_{ik} x_{i\ell}^{\tilde{\mathbf{W}}}}{x_{. \ell}^{\tilde{\mathbf{W}}} \sum_i \tilde{z}_{ik} x_{i.}}$;

Compute $x_{kj}^{\tilde{\mathbf{Z}}} = \sum_i \tilde{z}_{ik} x_{ij}$;

Column VE-step:

- In parallel, $\forall j, \tilde{w}_{j\ell} \propto \beta_\ell \exp(\lambda_c \sum_{j'} s_{jj'}^c \tilde{w}_{j'\ell} + \sum_k x_{kj}^{\tilde{\mathbf{Z}}} \log \gamma_{k\ell})$;
- Normalize $\tilde{\mathbf{W}}$ and apply damping;

Column M-step: $\beta_\ell = \frac{\tilde{w}_{. \ell}}{d}, \gamma_{k\ell} = \frac{\sum_j \tilde{w}_{j\ell} x_{kj}^{\tilde{\mathbf{Z}}}}{x_{k.}^{\tilde{\mathbf{Z}}} \sum_j \tilde{w}_{j\ell} x_{.j}}$;

end

Mesurer la concordance d'une partition avec les contraintes données

Pour une partition \mathcal{P} des lignes ou des colonnes, on définit

$$\mathcal{R}(\mathbf{S}, \mathcal{P}) = \frac{\sum_{ii'} |s_{ii'}| \text{unsat}(\mathcal{P}, s_{ii'})}{\sum_{ii'} |s_{ii'}|},$$

où,

$$\text{unsat}(\mathcal{P}, s_{ii'}) = \begin{cases} 1 & \text{si la contrainte } s_{ii'} \text{ est non vérifiée dans } \mathcal{P} \\ 0 & \text{si la contrainte } s_{ii'} \text{ est vérifiée dans } \mathcal{P} \text{ ou si } s_{ii'} = 0 \end{cases}$$

Mesurer la concordance d'une partition avec les contraintes données

Pour une partition \mathcal{P} des lignes ou des colonnes, on définit

$$\mathcal{R}(\mathbf{S}, \mathcal{P}) = \frac{\sum_{ii'} |s_{ii'}| \text{unsat}(\mathcal{P}, s_{ii'})}{\sum_{ii'} |s_{ii'}|},$$

Plusieurs applications selon la nature de \mathcal{P} :

- \mathcal{P} partition renvoyée par l'algorithme sans régularisation ($\lambda_r = \lambda_c = 0$) : proportion pondérée de contraintes qui ne sont pas déjà dans les données (Davidson, K. L. Wagstaff, and Basu 2006)
- \mathcal{P} vraie partition : mesure du bruit dans la matrice de contraintes \mathbf{S} (Miele, Picard, and Dray 2014).
- \mathcal{P} partition retournée par l'algorithme avec régularisation ($\lambda_r \lambda_c \neq 0$), proportion de contraintes qui n'ont pas été respectées par l'algorithme après convergence.

Résultats classification réseaux attribués

Effets des contraintes

