

## Classification automatique de réseaux dynamiques avec sous-graphes : étude du scandale Enron

**Title:** Clustering of dynamic networks through subgraphs : a study of the Enron scandal

Rawya Zreik<sup>1</sup>, Pierre Latouche<sup>2</sup> et Charles Bouveyron<sup>3</sup>

**Résumé :** Ces dernières années, de nombreux modèles de graphes aléatoires ont été proposés pour extraire des informations à partir de réseaux dans des domaines variés. Parmi ces modèles, nous considérons les modèles de clustering qui consistent à chercher des groupes de nœuds ayant des profils de connexion homogènes. La majorité de ces modèles est limitée à des réseaux statiques ayant des arêtes binaires ou discrètes et ne prennent donc pas en compte une éventuelle dimension temporelle. Ce travail est motivé par la volonté d'analyser un réseau dynamique décrivant les communications électroniques (emails) entre les employés de l'entreprise Enron, bien connue pour son scandale financier, où nous le verrons, les positions sociales jouent un rôle important. Nous proposons dans cet article une extension au cadre dynamique du modèle de graphe aléatoire RSM (Random Subgraph Model) qui a été récemment proposé pour modéliser à l'aide de groupes latents des réseaux statiques pour lesquels une partition en sous-graphes est connue. Notre approche est basée sur l'utilisation d'un modèle à espace d'état pour modéliser l'évolution au cours du temps des proportions des groupes latents. Le modèle ainsi obtenu est appelé modèle de sous-graphes aléatoires dynamiques (dRSM) et un algorithme de type VEM (Variational Expectation Maximization) est proposé pour effectuer l'inférence. Nous montrons que les approximations variationnelles conduisent à un nouveau modèle à espace d'état à partir duquel les paramètres ainsi que les états cachés peuvent être estimés en utilisant le filtre de Kalman et le lisseur de Rauch-Tung-Striebel (RTS). Des données simulées sont considérées pour évaluer l'efficacité de notre approche. La méthodologie est finalement appliquée au jeu des données emails de l'entreprise Enron et permet de mettre en évidence une réaction anticipée des cadres par rapport aux autres employés concernant le scandale à venir.

**Abstract:** In recent years, many clustering methods have been proposed to extract information from networks. The principle is to look for groups of vertices with homogenous connection profiles. Most of the models for clustering are suitable for static networks, that is to say, not taking into account the temporal dimension, but can handle different types of edges, whether binary or discrete. This work is motivated by the will of analysing an evolving network describing email communications between employees of the Enron company where social positions play an important role. Therefore, in this paper, we consider the random subgraph model (RSM) which was proposed recently to model a network through latent clusters built within a known partition into subgraphs. Using a state space model to characterize the cluster proportions, RSM is then extended in order to deal with dynamic networks. We call the latter the dynamic random subgraph model (dRSM). A variational expectation maximisation (VEM) algorithm is proposed to perform inference. We show that the variational approximations lead to a new state space model from which the parameters along with hidden states can be estimated using the standard Kalman filter and Rauch-Tung-Striebel (RTS) smoother. Simulated data sets are considered to assess the proposed approach. The methodology is finally applied to the Enron email data set and allows to discover an early reaction of the partners and directors compared to the other employees regarding the coming scandal.

<sup>1</sup> Laboratoire SAMM, EA 4543, Université Paris 1 Panthéon-Sorbonne.

Laboratoire MAP5, UMR CNRS 8145, Université Paris Descartes.

E-mail : [rawyazreik@gmail.com](mailto:rawyazreik@gmail.com)

<sup>2</sup> Laboratoire SAMM, EA 4543, Université Paris 1 Panthéon-Sorbonne.

E-mail : [pierre.latouche@univ-paris1.fr](mailto:pierre.latouche@univ-paris1.fr)

<sup>3</sup> Laboratoire MAP5, UMR CNRS 8145, Université Paris Descartes.

E-mail : [charles.bouveyron@parisdescartes.fr](mailto:charles.bouveyron@parisdescartes.fr)

**Mots-clés :** Réseau dynamique, sous-graphes, random subgraph model (RSM), modèle à espace d'état, classification automatique, algorithme VEM, données Enron.

**Keywords:** Dynamic networks , subgraphs, random subgraph model, state space model, Automatic classification, variational inference, variational expectation maximization, Enron data.

**Classification AMS 2000 :** 35L05, 35L70

## 1. Introduction

Depuis les travaux précurseurs de [Moreno \(1934\)](#), l'analyse des réseaux est devenue une discipline forte, qui ne se limite plus à la sociologie et qui est à présent appliquée à des domaines très variés tels que la biologie, la géographie ou l'histoire. L'intérêt croissant pour l'analyse des réseaux s'explique d'une part par la forte présence de ce type de données dans le monde numérique d'aujourd'hui et, d'autre part, par les progrès récents dans la modélisation et le traitement de ces données. En effet, informaticiens et statisticiens ont porté leurs efforts depuis plus d'une dizaine d'années sur ces données de type réseau et ont proposé de nombreuses techniques permettant leur analyse. Les méthodes de clustering permettent en particulier de découvrir une structure en groupes cachés dans le réseau. Parmi ces méthodes, nous pouvons citer les travaux de [Hofman and Wiggins \(2008\)](#) qui cherchent une partition des sommets où les groupes présentent une propriété de transitivité. Le modèle de [Handcock et al. \(2007\)](#) suppose quant à lui que les liens entre les sommets dépendent des positions des sommets dans un espace latent. Une approche populaire également, bien qu'asymptotiquement biaisée ([Bickel and Chen, 2009](#)), est celle proposée par [Girvan and Newman \(2002\)](#) reposant sur la notion de modularité.

Outre l'approche de [Handcock et al.](#), les méthodes statistiques récentes sont généralement basées sur le modèle à blocs stochastiques SBM (Stochastic Block Model) [Daudin et al. \(2008\)](#); [Nowicki and Snijders \(2001\)](#); [Wang and Wong \(1987\)](#), qui est une généralisation probabiliste de la méthode appliquée par [White et al. \(1976\)](#) sur les données de Sampson ([Fienberg and Wasserman, 1981](#)). Le modèle SBM suppose que chaque sommet appartient à un groupe latent et que la probabilité de connexion entre une paire de sommets dépend exclusivement de leur groupe. Parmi les nombreuses extensions récentes du modèle SBM, nous pouvons citer d'une part les modèles autorisant un nœud du réseau à appartenir à un ou plusieurs groupes ([Airoldi et al., 2008](#); [Latouche et al., 2011](#)). D'autre part, certains auteurs se sont intéressés à la modélisation de réseaux avec des arêtes valuées ([Mariadassou et al., 2010](#)) et à la prise en compte d'une information a priori ([Matias and Robin, 2014](#)). En particulier, le modèle RSM (Random Subgraph Model) [Jernite et al. \(2014\)](#) considère des arêtes catégorielles et nécessite la connaissance d'une partition du réseau en sous-graphes. Ce modèle suppose que chaque sous-graphe possède son propre mélange de groupes, ces derniers pouvant être présents dans tous les sous-graphes. Les sommets sont alors connectés entre eux avec une probabilité dépendante seulement des sous-graphes alors que le type d'arête est supposé être échantillonné conditionnellement aux groupes latents. Ce modèle a été appliqué avec succès à l'analyse d'un réseau historique dans la Gaule mérovingienne ([Jernite et al., 2014](#)). Toutefois, la dynamique temporelle du réseau, à l'origine présente dans les données, avait dû être écartée à cause de l'incapacité du modèle à gérer cet aspect.

Plusieurs approches ont été proposées récemment afin d'étendre les modèles de graphes aléatoires existants au cadre dynamique. L'idée principale consiste à introduire un processus temporel afin de caractériser l'évolution temporelle des nœuds et/ou des arêtes à travers le temps.

Ainsi, [Yang et al. \(2011\)](#) ont suggéré une version dynamique du modèle SBM autorisant chaque nœud à changer de groupe au temps  $t + 1$  en fonction de son état au temps  $t$ . Les probabilités de changement d'un groupe donné à un autre sont toutes décrites à l'aide d'une matrice de transition. L'approche alternative de [Xu and Hero III \(2013\)](#) pour le modèle SBM cherche quant à elle à modéliser les changements temporels à travers un modèle à espace d'état. Les auteurs ont recours à un filtre de Kalman pour l'inférence. Contrairement à [Yang et al. \(2011\)](#), [Xu and Hero III \(2013\)](#) traitent les probabilités d'apparition d'arêtes comme des paramètres temporels. En parallèle, le modèle MMSBM (Mixed Membership SBM) de [Airoldi et al. \(2008\)](#), capable de caractériser des groupes chevauchants, a été adapté pour gérer les réseaux dynamiques par [Xing et al. \(2010\)](#), en s'appuyant à nouveau sur un modèle à espace d'état. De plus, [Sarkar and Moore \(2005\)](#) ont proposé une version dynamique du modèle de [Handcock et al. \(2007\)](#), conservant ainsi une propriété de transitivité par laquelle deux nœuds proches dans un espace social caché ont une probabilité plus forte de connexion. Finalement, nous souhaitons mettre également en avant les travaux de [Dubois et al. \(2013\)](#) ainsi que ceux de [Heaukulani and Ghahramani \(2013\)](#). Dans [Dubois et al. \(2013\)](#), un processus de Poisson non homogène est considéré. Ainsi, contrairement à la plupart des modèles temporels pour le clustering de nœuds, un temps continu est pris en compte et un seul évènement, *i.e.* la création ou la suppression d'une arête, a lieu à chaque temps. Alors que les modèles cherchent généralement à décrire la dynamique des réseaux à travers l'évolution de structures latentes, Heaukulani et Ghahramani se sont plutôt intéressés à expliquer comment des interactions sociales connues pouvaient affecter des structures latentes non observées.

Dans cet article, en nous appuyant sur une version légèrement différente du modèle RSM et sur un modèle à espace d'état, nous proposons un nouveau modèle de graphes aléatoires pour les réseaux dynamiques que nous baptisons *dynamic random subgraph model* (dRSM). Le modèle relie les probabilités de connexion des nœuds à chaque instant  $t$  à des paramètres d'état dans un espace latent [Ahmed and Xing \(2007\)](#); [Blei and Lafferty \(2007a\)](#). L'inférence de ce modèle sera faite grâce à un algorithme de type EM variationnel (VEM). Des données simulées seront considérées pour évaluer l'approche proposée. Finalement, le modèle dRSM sera appliqué au célèbre jeu de données Enron qui présente l'évolution des communications électroniques entre les employés pendant les deux ans (2000, 2001) avant et durant la faillite de l'entreprise en décembre 2001. A titre d'illustration, la figure 1 présente l'évolution du réseau de communications emails dans les quatre mois clés (août – novembre 2001) de la crise Enron.

## 2. Le modèle de sous-graphes aléatoires dynamiques

Ce paragraphe pose le contexte de notre travail et introduit le modèle dRSM. L'inférence de ce modèle est ensuite traitée et un algorithme EM variationnel est proposé pour ce faire.

### 2.1. Contexte et notations

Nous considérons un ensemble de  $T$  réseaux  $\{\mathcal{G}^{(t)}\}_{t=1}^T$ , où  $\mathcal{G}^{(t)}$  est un graphe dirigé observé au temps  $t$  et qui est représenté par une matrice d'adjacence  $X^{(t)}$  de taille  $N \times N$ ,  $N$  désignant le nombre de nœuds. Chaque arête  $X_{ij}^{(t)}$ , décrivant la relation entre les nœuds  $i$  et  $j$ , est supposée prendre ses valeurs dans  $\{0, \dots, C\}$  tel que  $X_{ij}^{(t)} = c$  signifie que les nœuds  $i$  et  $j$  sont liés par une

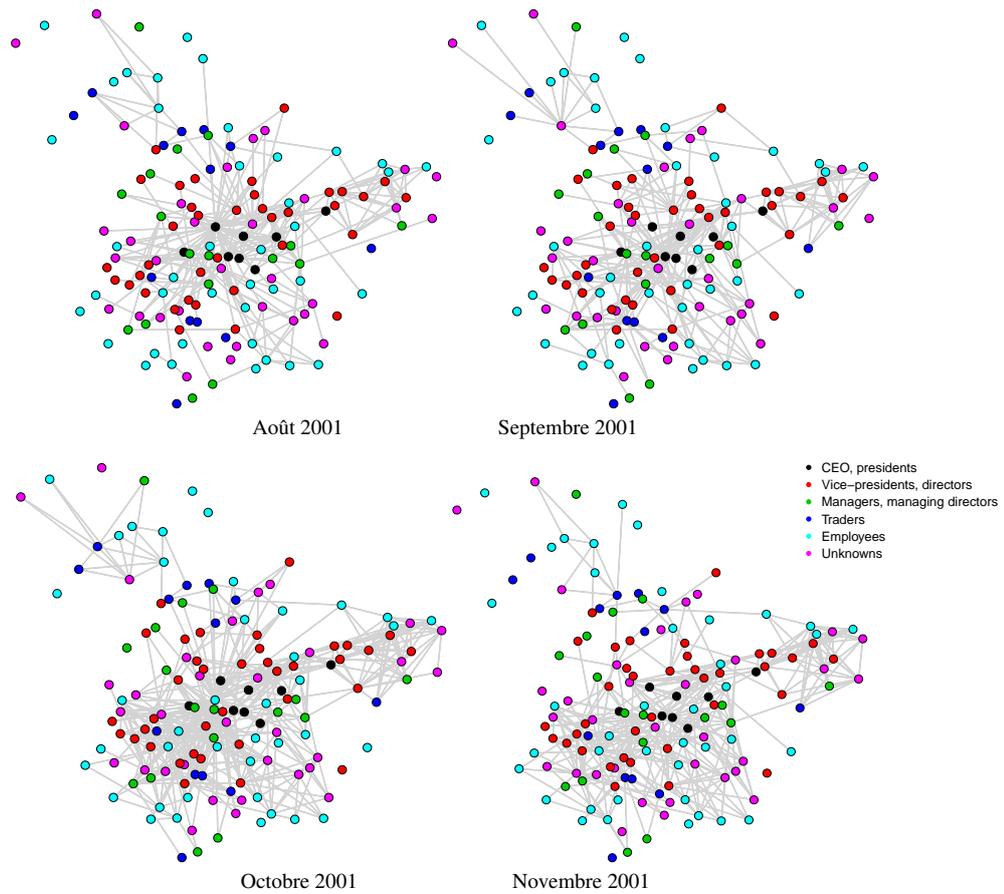


FIGURE 1. Réseau des communications électroniques entre 148 employés d'Enron durant les 4 mois (août - novembre 2001) précédant la faillite de l'entreprise.

relation de type  $c$  au temps  $t$  et  $X_{ij}^{(t)} = 0$  indique en particulier l'absence de relation entre les deux nœuds à cet instant. Notons que nous ne considérons pas les boucles, c'est à dire les connexions d'un nœud sur lui même, et donc  $X_{ii}^{(t)} = 0, \forall i, t$ .

Nous supposons en outre qu'une partition  $\mathcal{P}$  du réseau en  $S$  classes est également connue. Notons que la partition observée induit une décomposition du graphe en sous-graphes où chaque classe de nœuds correspond à un sous-graphe. Cette partition pourra par exemple être issue de covariables sur les nœuds, telles que le statut de l'employé dans l'entreprise. Pour ce faire, nous introduisons la variable  $s$  qui prend ses valeurs dans  $\{1, \dots, S\}$ , de manière à ce que  $s_i^{(t)}$  indique à quel sous-graphe le nœud  $i$  appartient au temps  $t$ . Dans certains cas, et afin de clarifier les équations, nous considérerons également une variable indicatrice  $y_{is}$ , telle que  $y_{is} = 1$  si le nœud  $i$  appartient au sous-graphe  $s$ , 0 sinon. Enfin, puisqu'un nœud  $i$  ne peut appartenir qu'à un seul sous-graphe, nous avons  $\sum_{s=1}^S y_{is} = 1, \forall s$ .

Notre objectif est donc de regrouper, à chaque temps  $t$ , les  $N$  nœuds en  $K$  groupes latents de profils de connexions homogènes, *i.e.* trouver une estimation de la matrice binaire  $Z$  telle que  $Z_{ik}^{(t)} = 1$  si, à l'instant  $t$ , le nœud  $i$  appartient à la classe  $k$  et 0 sinon. Notons que  $N, C, \mathcal{P}, S$  et  $K$  sont tous supposés constants au cours du temps.

## 2.2. Le modèle à un instant de temps

Le réseau, représenté par sa matrice d'adjacence  $X^{(t)}$ , est supposé être généré à chaque instant  $t$  comme suit. Tout d'abord, chaque nœud  $i$  est associé à une classe latente  $k = 1, \dots, K$ , avec une probabilité dépendante du sous-graphe auquel il appartient. Nous supposons donc que, pour un nombre  $K$  de groupes latents donné, la variable  $Z_i^{(t)}$  est distribuée selon une loi multinomiale de paramètre  $\alpha_{s_i}^{(t)}$  :

$$Z_i^{(t)} \sim \mathcal{M}(1, \alpha_{s_i}^{(t)}),$$

où  $\alpha_s^{(t)} = (\alpha_{s1}^{(t)}, \dots, \alpha_{sK}^{(t)})$  est le vecteur des probabilités *a priori* des  $K$  groupes latents dans le sous-graphe  $s$  à l'instant  $t$  et est tel que :

$$\forall s \in 1, \dots, S, \sum_{k=1}^K \alpha_{sk}^{(t)} = 1.$$

Le modèle autorise donc chaque sous-graphe à avoir des proportions  $\alpha_s^{(t)}$  des groupes latents différentes et cela pour chaque instant de temps. De plus, notons qu'il n'y a pas de dépendance directe entre  $Z_i^{(t)}$  et  $Z_i^{(t-1)}$  à ce niveau de modélisation. La dépendance temporelle sera gérée par un processus temporel sur les proportions des groupes latents (voir paragraphe suivant).

Nous supposons d'autre part que le type de lien entre les nœuds  $i$  et  $j$  est, conditionnellement aux variables  $Z_i^{(t)}$  et  $Z_j^{(t)}$ , distribué à nouveau selon une loi multinomiale :

$$X_{i,j}^{(t)} | Z_{ik}^{(t)} Z_{jl}^{(t)} = 1 \sim \mathcal{M}(1, \Pi_{kl}),$$

avec  $\Pi_{kl} \in [0, 1]^{C+1}$  et  $\sum_{c=0}^C \Pi_{kl}^c = 1, \forall k, l$ . Remarquons que, par souci de parcimonie, les paramètres  $\Pi_{kl}^c$  sont supposés constants au cours du temps.

TABLEAU 1. *Résumé des principales notations utilisées.*

Notations	Descriptions
$X$	Matrice d'adjacence. $X_{i,j}^{(t)} \in \{0, \dots, C\}$ à chaque instant $t$ .
$Z$	Matrice binaire d'appartenance aux groupes. $Z_{i,k}^{(t)} = 1$ si $i \in$ groupe $k$ au temps $t$ .
$N$	Nombre de sommets dans le réseau.
$K$	Nombre des groupes latents.
$S$	Nombre de sous-graphes.
$C$	Nombre de types d'arêtes.
$\Pi$	$\Pi_{kl}^c$ est la probabilité d'une arête de type $c$ entre les groupes $k$ et $\ell$ .
$\alpha$	$\alpha_{sk}^{(t)} = f_k(\gamma_s^{(t)})$ est la proportion du groupe $k$ dans le sous-graphe $s$ au temps $t$ .

Comme pour le modèle RSM, tous les vecteurs  $Z_i^{(t)}$  sont supposés indépendants et, conditionnellement à tous ces vecteurs, les arêtes sont tirées de manière indépendante. Ainsi, contrairement au modèle original RSM, les arêtes dépendent ici des groupes latents exclusivement, et il n'y a pas de dépendance directe vis à vis des sous graphes. Chaque arête entre une paire  $(i, j)$  de nœuds dépend des sous graphes  $s_i$  et  $s_j$ , mais seulement à travers le fait que l'arête dépend des groupes latents qui eux mêmes dépendent des sous graphes. La dépendance est ici indirecte alors que dans le modèle RSM d'origine, les sous graphes et les groupes interviennent tous dans la création d'une arête et ont des rôles différents. En effet, la présence d'une arête entre  $i$  et  $j$  est d'abord tirée à partir d'une loi de Bernoulli dépendant de  $s_i$  et  $s_j$ . Si une arête est présente, son type est alors généré en fonction des groupes latents. La distinction entre les rôles des sous graphes et des groupes dans la création d'une arête était à l'origine motivée par des considérations liées au réseau historique analysé. Nous ne faisons pas de telles hypothèses dans ce papier. Les groupes latents expliquent à la fois la création d'une arête et son type.

### 2.3. Modélisation de l'évolution des sous-graphes aléatoires

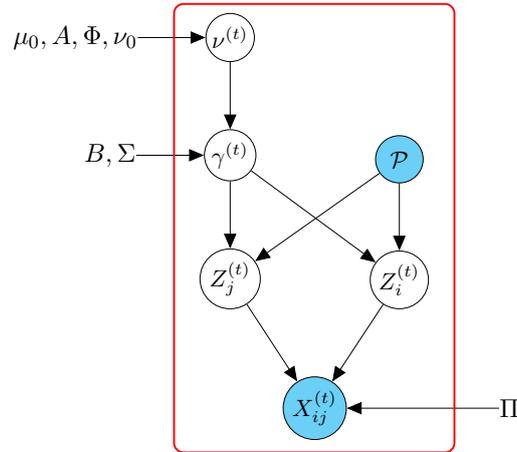
Nous ajoutons à présent un état caché, sous la forme d'un modèle à espace d'état [Xing et al. \(2010\)](#), pour modéliser l'évolution des proportions des groupes dans les sous-graphes au cours du temps. Nous introduisons donc une nouvelle variable latente,  $\gamma_s^{(t)}$ , permettant de faire le lien entre les  $\alpha_s^{(t)}$  aux différents temps et cela grâce à une transformation  $f(\cdot)$  de type logistique :

$$\alpha_s^{(t)} = f(\gamma_s^{(t)}),$$

telle que

$$\alpha_{sk}^{(t)} = \exp(\gamma_{sk}^{(t)} - C(\gamma_s^{(t)})), \forall s, k, t,$$

où  $\gamma_{sK}^{(t)} = 0$  et  $C(\gamma_s^{(t)}) = \log(\sum_{l=1}^K \exp(\gamma_{sl}^{(t)}))$ . Le choix de fixer la dernière composante du vecteur  $\gamma_s^{(t)}$  arbitrairement à 0 est largement utilisé dans la littérature [Blei and Lafferty \(2007b,a\)](#); [Lafferty and Blei \(2006\)](#); [Xing et al. \(2010\)](#). Il permet de régler les contraintes de bijectivité de la fonction  $f(\cdot)$  qui nécessite que  $\gamma_s^{(t)}$  vive dans un espace vectoriel de dimension  $K - 1$  puisque  $\alpha_s^{(t)}$  a  $K - 1$  degrés de liberté. Cela induit  $\gamma_{sk}^{(t)} = \log(\alpha_{sk}^{(t)} / \alpha_{sK}^{(t)})$ ,  $\forall s, k, t$ . En outre, les  $K - 1$  premières

FIGURE 2. *Modèle graphique associé à dRSM.*

composantes du vecteur latent  $\gamma_s^{(t)}$  sont supposées être distribuées selon une loi normale centrée en  $B\mathbf{v}^{(t)}$  et de matrice de covariance  $\Sigma$  :

$$\gamma_{s \setminus K}^{(t)} \sim \mathcal{N}(B\mathbf{v}^{(t)}, \Sigma), \quad (1)$$

où  $\gamma_{s \setminus K}^{(t)}$  est le vecteur  $\gamma_s^{(t)}$  privé de sa dernière composante. Les matrices  $B$  et  $\Sigma$  sont de taille  $(K-1) \times (K-1)$  alors que  $\mathbf{v}^{(t)}$  est un vecteur de taille  $K-1$ .

Le reste de la modélisation fait maintenant intervenir un modèle à espace d'état classique pour systèmes dynamiques linéaires. Le modèle est comme suit :

$$\begin{cases} \mathbf{v}^{(t)} = A\mathbf{v}^{(t-1)} + \boldsymbol{\omega} \\ \mathbf{v}^{(1)} = \boldsymbol{\mu}_0 + u. \end{cases}$$

Les termes de bruit  $\boldsymbol{\omega}$  et  $u$  sont d'autre part supposés indépendants et de distributions gaussiennes :

$$\begin{cases} \boldsymbol{\omega} \sim \mathcal{N}(0, \Phi) \\ u \sim \mathcal{N}(0, V_0). \end{cases}$$

A nouveau, les matrices  $A$ ,  $\Phi$ , et  $V_0$  sont de taille  $(K-1) \times (K-1)$  alors que  $\boldsymbol{\mu}_0$  est un vecteur de taille  $K-1$ . Ainsi, les proportions  $\alpha_s^{(t)}$  du mélange dans chaque sous-graphe aux différents temps  $t = 1, \dots, T$  sont liées entre elles au travers du modèle à espace d'état. Remarquons que le modèle ne suppose pas une dépendance directe entre  $Z_i^{(t)}$  et  $Z_i^{(t-1)}$ . La dépendance temporelle n'intervient qu'au travers du processus gérant les proportions.

Notons que le modèle à espace d'état pour systèmes dynamiques linéaires souffre de problèmes d'identifiabilité et par conséquent des contraintes doivent être introduites (voir par exemple Harvey [Harvey \(1989\)](#)) afin de garantir l'identifiabilité du modèle. En pratique, dans toutes les expériences que nous avons effectuées, les matrices  $A$ ,  $B$ , et  $V_0$  ont toujours été supposées égales à la matrice identité  $I_{K-1}$ . De plus,  $\Phi$  a été contrainte à être diagonale, de la forme  $\Phi = \sigma^2 I_{K-1}$ . Enfin, toutes

les composantes de  $\mu_0$  ont été supposées égales à zéro. Par la suite, la description du modèle et de son inférence restera générale car d'autres contraintes peuvent être utilisées.

Le modèle ainsi décrit possède trois variables latentes  $\mathbf{v} = (\mathbf{v}^{(t)})_t$ ,  $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_s^{(t)})_{st}$ ,  $\mathbf{Z} = (\mathbf{Z}_{ik}^{(t)})_{ikt}$  et est paramétré par  $\boldsymbol{\theta} = (\mu_0, A, B, \Phi, V_0, \Sigma, \Pi)$ . Notons que l'ensemble des paramètres dans  $\boldsymbol{\theta}$  ne dépendent ni du temps, ni des sous-graphes. Ce modèle est appelé *dynamic random subgraph model* (dRSM) dans la suite du document. La table 1 résume les principales notations utilisées et la figure 2 présente le modèle graphique associé à dRSM.

#### 2.4. Distribution jointe

Le modèle dRSM est donc défini par la distribution jointe suivante :

$$\begin{aligned} p(X, Z, \boldsymbol{\gamma}, \mathbf{v} | \boldsymbol{\theta}) &= p(X, \boldsymbol{\gamma}, \mathbf{v} | Z, \boldsymbol{\theta}) p(Z | f(\boldsymbol{\gamma})) \\ &= p(X | Z, \Pi) p(Z | f(\boldsymbol{\gamma})) p(\boldsymbol{\gamma}_{\setminus K} | B, \mathbf{v}, \Sigma) p(\mathbf{v} | \mu_0, A, \Phi, V_0), \end{aligned}$$

où  $\boldsymbol{\gamma}_{\setminus K} = (\boldsymbol{\gamma}_{s \setminus K}^{(t)})_{st}$ , et

$$p(X | Z, \Pi) = \prod_{t=1}^T \prod_{k,l}^K \prod_{c=0}^C (\Pi_{kl}^c)^{\sum_{i \neq j} \delta(X_{i,j}^{(t)}=c) Z_{ik}^{(t)} Z_{jl}^{(t)}}.$$

De plus,

$$\begin{aligned} p(Z | f(\boldsymbol{\gamma})) &= \prod_{t=1}^T \prod_{k=1}^K \prod_{i=1}^N f_k(\boldsymbol{\gamma}_{s_i}^{(t)})^{Z_{ik}^{(t)}} \\ &= \prod_{t=1}^T \prod_{k=1}^K \prod_{s=1}^S f_k(\boldsymbol{\gamma}_s^{(t)})^{\sum_{i=1}^N y_{is} Z_{ik}^{(t)}}. \end{aligned}$$

Notons que

$$p(\boldsymbol{\gamma}_{\setminus K} | B, \mathbf{v}, \Sigma) = \prod_{t=1}^T \prod_{s=1}^S \mathcal{N}(\boldsymbol{\gamma}_{s \setminus K}^{(t)}; B\mathbf{v}^{(t)}, \Sigma),$$

où  $\mathcal{N}(\boldsymbol{\gamma}_{s \setminus K}^{(t)}; B\mathbf{v}^{(t)}, \Sigma)$  désigne la distribution gaussienne multivariée, de vecteur moyenne  $B\mathbf{v}^{(t)}$  et de matrice de covariance  $\Sigma$ , évaluée en  $\boldsymbol{\gamma}_{s \setminus K}^{(t)}$ . Finalement,

$$p(\mathbf{v} | \mu_0, A, \Phi, V_0) = p(\mathbf{v}^{(1)} | \mu_0, V_0) \prod_{t=2}^T \log p(\mathbf{v}^{(t)} | \mathbf{v}^{(t-1)}, A, \Phi).$$

#### 2.5. Un cadre variationnel pour l'inférence du modèle dRSM

Nous considérons à présent l'inférence du modèle dRSM introduit dans les paragraphes précédents. Nous cherchons à maximiser la log-vraisemblance  $\log p(X | \boldsymbol{\theta})$ . Pour réaliser cette optimisation, il est d'usage de recourir à l'utilisation d'un algorithme EM [Dempster et al. \(1977\)](#); [Krishnan and McLachlan \(1997\)](#). Malheureusement, ce dernier ne peut pas être dérivé dans notre cas car la loi

conditionnelle des variables latentes, sachant les données et les paramètres, n'est pas calculable. Pour palier à ce problème, nous proposons ici de recourir à un algorithme de type EM variationnel (VEM) [Csisz et al. \(1984\)](#); [Hathaway \(1986\)](#) qui optimise localement les paramètres du modèle par rapport à une borne inférieure de la log-vraisemblance. Cet algorithme permet d'approcher les distributions variationnelles a posteriori pour toutes les variables latentes.

Étant donnée une distribution variationnelle  $q$  de  $(Z, \gamma, \nu)$ , la log-vraisemblance peut s'écrire :

$$\log p(X|\theta) = \mathcal{L}(q, \theta) + KL(q(\cdot) \parallel p(\cdot|X, \theta)),$$

où  $\mathcal{L}$  est définie comme suit :

$$\mathcal{L}(q, \theta) = \sum_z \int_{\gamma} \int_{\nu} q(Z, \gamma, \nu) \log \frac{p(X, Z, \gamma, \nu|\theta)}{q(Z, \gamma, \nu)} d\gamma d\nu, \quad (2)$$

et la divergence de Kullback-Leibler est donnée par :

$$KL(q(\cdot) \parallel p(\cdot|X, \theta)) = - \sum_z \int_{\gamma} \int_{\nu} q(Z, \gamma, \nu) \log \frac{p(Z, \gamma, \nu|X, \theta)}{q(Z, \gamma, \nu)} d\gamma d\nu.$$

Trouver la meilleure approximation de la distribution a posteriori  $p(Z, \gamma, \nu|X, \theta)$  au sens de la divergence  $KL$  est équivalent à trouver  $q(\cdot)$  qui maximise la borne inférieure  $\mathcal{L}$  de la log-vraisemblance. Or,  $p(X, Z, \gamma, \nu|\theta)$ , fait intervenir la quantité  $p(Z|f(\gamma))$  dont le calcul de l'espérance est rendu difficile du fait de la constante de normalisation  $C(\gamma_s^{(t)}) = \log(\sum_{l=1}^K \exp(\gamma_{sl}^{(t)}))$ .

Pour résoudre ce problème, nous proposons une borne supérieure, comme dans [Jordan et al. \(1999\)](#), pour  $C(\gamma_s^{(t)})$ . Ainsi, pour n'importe quel  $\xi_s^{(t)}$  dans  $\mathbb{R}^{*+}$ , comme  $C(\cdot)$  est une fonction concave de  $\sum_{l=1}^K \exp(\gamma_{sl}^{(t)})$ , la relation suivante s'obtient naturellement :

$$\log\left(\sum_{l=1}^K \exp(\gamma_{sl}^{(t)})\right) \leq \xi_s^{-1(t)} \left(\sum_{l=1}^K \exp(\gamma_{sl}^{(t)})\right) - 1 + \log(\xi_s^{(t)}).$$

Cette expression permet de construire une borne inférieure de  $\log p(Z|f(\gamma))$  :

$$\begin{aligned} \log p(Z|f(\gamma)) &= \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N Z_{ik}^{(t)} \log(f_k(\gamma_{si}^{(t)})) \\ &= \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N \sum_{s=1}^S y_{is} Z_{ik}^{(t)} \left( \gamma_{sk}^{(t)} - \log\left(\sum_{l=1}^K \exp(\gamma_{sl}^{(t)})\right) \right) \\ &\geq \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N \sum_{s=1}^S y_{is} Z_{ik}^{(t)} \left( \gamma_{sk}^{(t)} - \left( \xi_s^{-1(t)} \sum_{l=1}^K \exp(\gamma_{sl}^{(t)}) - 1 + \log(\xi_s^{(t)}) \right) \right). \end{aligned}$$

Soit  $h$  une fonction des variables latentes dans  $(Z, \gamma, \xi)$  de sorte que :

$$\log h(Z, \gamma, \xi) = \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N \sum_{s=1}^S y_{is} Z_{ik}^{(t)} \left( \gamma_{sk}^{(t)} - \left( \xi_s^{-1(t)} \sum_{l=1}^K \exp(\gamma_{sl}^{(t)}) - 1 + \log(\xi_s^{(t)}) \right) \right).$$

Alors,  $\log p(Z|f(\gamma)) \geq \log h(Z, \gamma, \xi)$  où  $\xi = (\xi_s^{(t)})_{st}$  est un ensemble de paramètres variationnels. Notons qu'un paramètre variationnel  $\gamma_s^{(t)}$  est considéré pour chaque sous-graphe  $s$  et chaque temps  $t$ , pour plus de flexibilité et afin d'améliorer la qualité des approximations. Comme nous le verrons par la suite, les paramètres  $\gamma_s^{(t)}$  peuvent être optimisés de manière à améliorer les bornes ainsi obtenues.

---

**Algorithme 1:** Algorithme VEM pour le modèle dRSM à  $K$  groupes latents.
 

---

Initialisation de  $\theta^0 = (\mu_0, A, \Phi, V_0, \Pi, B, \Sigma)$   
 Initialisation de la matrice  $\tau$  à chaque instant  $t$   
 Déterminer  $\hat{v}, \hat{V}, (\hat{\gamma}_{sk}^{(t)}, \hat{\sigma}_{sk}^{2(t)})_{sk}, \hat{\theta}, \hat{\xi}$   
 Calculer  $\tilde{\mathcal{L}}(q, \hat{\theta}, \hat{\xi})$   
 Tant que  $|\tilde{\mathcal{L}}^{new} - \tilde{\mathcal{L}}^{old}| > \varepsilon$   
     Mise à jour de  $\tau, \hat{v}, \hat{V}, (\hat{\gamma}_{sk}^{(t)}, \hat{\sigma}_{sk}^{2(t)})_{skt}$  (Étape E)  
     Mise à jour de  $\hat{\theta}$  (Étape M)  
     Mise à jour de  $\hat{\xi}$   
     Calcul de  $\mathcal{L}(q, \hat{\theta}, \hat{\xi})$   
 fin boucle

---

En remplaçant  $\log p(Z|f(\gamma))$  par  $\log h(Z, \gamma, \xi)$  dans l'équation (2), une nouvelle borne inférieure est obtenue  $\tilde{\mathcal{L}}(q, \theta, \xi)$  pour  $\log p(X|\theta)$  :

$$\log p(X|\theta) \geq \mathcal{L}(q, \theta) \geq \tilde{\mathcal{L}}(q, \theta, \xi),$$

où :

$$\tilde{\mathcal{L}}(q, \theta, \xi) = \sum_z \int_{\gamma} \int_{\mathcal{V}} q(Z, \gamma, \mathbf{v}) \log \frac{p(X|Z, \Pi) h(Z, \gamma, \xi) p(\gamma|\mathbf{v}, \Sigma) p(\mathbf{v}|\mu_0, A, \Phi, V_0)}{q(Z, \gamma, \mathbf{v})} d\gamma d\mathbf{v}.$$

Pour permettre la maximisation de  $\tilde{\mathcal{L}}(q, \theta, \xi)$ , nous supposons en outre que  $q(Z, \gamma, \mathbf{v})$  a la forme variationnelle suivante :

$$\begin{aligned} q(Z, \gamma, \mathbf{v}) &= q(Z)q(\gamma)q(\mathbf{v}) \\ &= \left( \prod_{t=1}^T \prod_{i=1}^N q(Z_i^{(t)}) \right) q(\gamma)q(\mathbf{v}). \end{aligned}$$

De plus,  $q(\gamma)$  est choisi parmi la famille de lois normales de la forme :

$$q(\gamma) = \prod_{t=1}^T \prod_{s=1}^S \prod_{k=1}^K \mathcal{N}(\gamma_{sk}^{(t)}; \hat{\gamma}_{sk}^{(t)}, \hat{\sigma}_{sk}^{2(t)}),$$

afin de dériver des expressions analytiques dans l'étape E. Comme la dernière composante de chaque vecteur  $\gamma_s^{(t)}$  doit rester égale à zéro pour préserver les contraintes de bijectivité de la transformée  $f(\cdot)$ , les termes  $\hat{\gamma}_{sK}^{(t)}$  et  $\hat{\sigma}_{sK}^{2(t)}$  sont fixés à zéro, de manière à obtenir une masse de Dirac en zéro. Tous les autres termes de moyenne et de variance  $(\hat{\gamma}_{sk}^{(t)}, \hat{\sigma}_{sk}^{2(t)}), \forall s, k \neq K, t$  sont des paramètres à estimer.

## 2.6. Algorithme VEM pour l'inférence du modèle dRSM

Dans cette section, les paramètres variationnels dans  $\xi$  sont supposés connus. Grâce aux approximations proposées précédemment, il est maintenant possible de proposer un algorithme VEM (algorithme 1) permettant de maximiser itérativement la borne  $\mathcal{L}(q, \theta, \xi)$  par rapport à  $q(Z, \gamma, \nu)$  (étape E) et par rapport aux paramètres du modèle  $\theta$  (étape M). Les algorithmes VEM standards considèrent généralement un seul ensemble de variables latentes. Dans notre cas, trois ensembles différents coexistent et par conséquent l'étape E nécessite elle-même des étapes successives d'optimisation. A chaque étape, toutes les distributions sont fixées, exceptée une, qui est mise à jour. Cette séquence est répétée pour chacune des distributions dans  $q(Z, \gamma, \nu)$ .

Nous donnons ci-dessous les formules de mise à jour de ces étapes. Les détails des calculs sont donnés dans l'annexe A.1.

**Proposition 1.** *L'étape de mise à jour pour la distribution  $q(Z_i^{(t)})$  est donnée par :*

$$q(Z_i^{(t)}) \sim \mathcal{M}(Z_i^{(t)}; 1, \tau_i^{(t)}) \quad \forall i, t,$$

où :

$$\begin{aligned} \tau_{ik}^{(t)} \propto \exp \left( \sum_{l=1}^K \sum_{c=0}^C \sum_{j \neq i}^N \delta(X_{i,j}^{(t)} = c) \tau_{jl}^{(t)} \left[ \log(\Pi_{kl}^c) + \log(\Pi_{lk}^c) \right] \right. \\ \left. + \sum_{s=1}^S y_{is} \left( \hat{\gamma}_{sk}^{(t)} - \left( \xi_s^{-1(t)} \sum_{l=1}^K \exp(\hat{\gamma}_{sl}^{(t)} + \frac{\hat{\sigma}_{sl}^{2(t)}}{2}) - 1 + \log(\xi_s^{(t)}) \right) \right) \right). \end{aligned}$$

Notons que  $\tau_{ik}^{(t)}$  est une approximation de la probabilité a posteriori que le nœud  $i$  appartienne à la classe  $k$  à l'instant  $t$ .

**Proposition 2.** *L'étape de mise à jour pour la distribution  $q(\nu)$  est donnée par :*

$$q(\nu) \propto p(\nu^{(1)} | \mu_0, V_0) \left[ \prod_{t=2}^T p(\nu^{(t)} | \nu^{(t-1)}, A, \Phi) \right] \left[ \prod_{t=1}^T \mathcal{N}(x^{(t)}; B, \nu^{(t)}, \frac{\Sigma}{S}) \right],$$

$$\text{où } x^{(t)} = \sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)} / S.$$

Nous rappelons qu'à cette étape, les termes  $\hat{\gamma}_s^{(t)}$  sont fixés. Cette formulation de  $q(\nu)$  est donc remarquable puisqu'il s'agit de celle associée à la loi a posteriori d'un modèle à espace d'état où  $\nu$  est l'ensemble des variables latentes et  $x = (x^{(t)})_t$  l'ensemble des variables observées. En effet, chaque  $x^{(t)}$  suit une loi normale  $\mathcal{N}(B\nu^{(t)}, \Sigma/S)$  et il est possible de faire correspondre  $q(\nu)$  à un système linéaire de la forme :

$$\begin{cases} \nu^{(t)} = A\nu^{(t-1)} + \omega \\ \nu^{(1)} = \mu_0 + u, \end{cases}$$

avec

$$\begin{cases} \omega \sim \mathcal{N}(0, \Phi) \\ u \sim \mathcal{N}(0, V_0). \end{cases}$$

Par conséquent, l'approximation variationnelle de  $q(\nu)$  conduit à un nouveau modèle à espace d'état dont les paramètres  $\theta' = (\mu_0, A, B, \Phi, V_0, \Sigma/S)$  peuvent être estimés grâce au filtre de Kalman standard et au lisseur de Rauch-Tung-Striebel (RTS) [Minka \(1998\)](#); [Rauch et al. \(1965\)](#). Le paquet MARSS pour le logiciel R permet notamment de faire l'estimation d'un tel modèle à espace d'état. Notons que le vecteur  $x$  rassemble l'ensemble des variables observées  $x^{(t)}$ , et que  $(\hat{\nu}^{(t)}, \hat{V}^{(t)})$  sont les estimations de l'espérance et la matrice de covariance de la variable latente  $\nu^{(t)}$  sachant  $x$ .

**Proposition 3.** *Après l'étape E de l'algorithme VEM, la borne  $\tilde{\mathcal{L}}(q, \theta, \xi)$  prend la forme suivante :*

$$\begin{aligned} \tilde{\mathcal{L}}(q, \theta, \xi) &= \sum_{t=1}^T \sum_{k,l}^K \sum_{c=0}^C \sum_{i \neq j}^N \delta(X_{i,j}^{(t)} = c) \tau_{ik}^{(t)} \tau_{jl}^{(t)} \log(\Pi_{kl}^c) \\ &+ \sum_{t=1}^T \sum_{s=1}^S \left( (r_s^{(t)})^\top \hat{\gamma}_s^{(t)} - N_s \xi_s^{-1(t)} \sum_{l=1}^K \exp(\hat{\gamma}_{sl}^{(t)} + \frac{\hat{\sigma}_{sl}^{2(t)}}{2}) + N_s - N_s \log(\xi_s^{(t)}) \right) \\ &+ \sum_{t=1}^T \sum_{s=1}^S \left( \log \mathcal{N}(\hat{\gamma}_{s \setminus K}^{(t)}; B \hat{\nu}_s^{(t)}, \Sigma) - \frac{1}{2} \text{tr}(\Sigma^{-1} B^\top \hat{V}^{(t)} B) - \frac{1}{2} \text{tr}(\Sigma^{-1} \hat{\sigma}_s^{2(t)}) \right) \\ &- \sum_{t=1}^T \sum_{s=1}^S \sum_{k=1}^{K-1} -\log \left( (2\pi)^{\frac{1}{2}} \hat{\sigma}_{sk}^{(t)} \right) + \frac{T(K-1)S}{2} \\ &- \sum_{t=1}^T \left( \log \mathcal{N}(x^{(t)}; B \hat{\nu}^{(t)}, \frac{\Sigma}{S}) + \frac{1}{2} \text{tr}(\Sigma^{-1} S B^\top \hat{V}^{(t)} B) \right) \\ &- \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^K \tau_{ik}^{(t)} \log(\tau_{ik}^{(t)}) \\ &+ \log p(x|\theta'), \end{aligned}$$

où  $r_s^{(t)} = \sum_{i=1}^N \tau_i^{(t)} y_{is}$ ,  $N_s$  est le nombre de nœuds dans le sous graphe  $s$  et  $\log p(x|\theta')$  est la log-vraisemblance du système dynamique linéaire qui est associé à la distribution variationnelle  $q(\nu)$ . Le détail du calcul de la borne  $\tilde{\mathcal{L}}(q, \theta, \xi)$  est donné en annexe A.2.

La maximisation de cette borne permet d'obtenir les formules de mise à jour pour la matrice  $\Pi$  comme suit :

$$\hat{\Pi}_{kl}^c = \frac{\sum_{t=1}^T \sum_{i \neq j}^N \delta(X_{i,j}^{(t)} = c) \tau_{ik}^{(t)} \tau_{jl}^{(t)}}{\sum_{t=1}^T \sum_{c=0}^C \sum_{i \neq j}^N \delta(X_{i,j}^{(t)} = c) \tau_{ik}^{(t)} \tau_{jl}^{(t)}}, \forall k, l, c.$$

Les paramètres  $\hat{\gamma}_{sik}^{(t)}$  et  $\hat{\sigma}_{sl}^{2(t)}$  doivent quant à eux être obtenus par une maximisation numérique de la borne. Cela peut notamment être fait grâce à un algorithme de type quasi-Newton.

## 2.7. Optimisation de $\xi$

Jusqu'à présent, nous avons vu qu'un algorithme de type VEM pouvait être mis en œuvre à partir d'approximations dépendant de paramètres variationnels  $\xi_s^{(t)}$ . Cependant, le problème de l'estimation de ces paramètres à partir des données n'a pas encore été abordé. Nous nous appuyons

sur les travaux de [Svensén and Bishop \(2004\)](#) sur les mélanges hiérarchiques Bayésiens d'experts. Ainsi, la borne inférieure  $\mathcal{L}(q, \theta, \xi)$  est optimisée par rapport à chaque terme variationnel  $\xi_s^{(t)}$  de manière à améliorer la qualité des approximations variationnelles. Cette optimisation conduit à de nouveaux estimateurs  $\hat{\xi}_s^{(t)}$  des  $\xi_s^{(t)}$  :

$$\hat{\xi}_s^{(t)} = \sum_{l=1}^K \exp(\hat{\gamma}_{sl}^{(t)} + \hat{\sigma}_{sl}^{2(t)}), \forall s, t.$$

Cette procédure donne lieu à un schéma d'optimisation en trois étapes. Pour des paramètres  $\xi = (\xi_s^{(t)})_{st}$  fixés, l'algorithme VEM décrit précédemment est utilisé pour maximiser la borne inférieure par rapport à  $q(Z, \gamma, \nu)$  et  $\theta$ . Ces termes sont ensuite fixés et une nouvelle estimation des  $\xi_s^{(t)}$  est proposée. Les trois étapes sont répétées jusqu'à convergence de la borne inférieure.

### 3. Simulation numérique

Ce paragraphe vise, d'une part, à mettre en avant les caractéristiques du modèle dRSM présenté dans le paragraphe 2 et, d'autre part, à démontrer la validité de l'algorithme d'inférence que nous avons proposé. Pour ce faire, nous considérons dans cette partie des données simulées.

### 4. Simulation numérique

Ce paragraphe vise, d'une part, à mettre en avant les caractéristiques du modèle dRSM présenté dans le paragraphe 2 et, d'autre part, à démontrer la validité de l'algorithme d'inférence que nous avons proposé. Pour ce faire, nous considérons dans cette partie des données simulées.

#### 4.1. Configuration expérimentale

Afin de valider notre approche et de faciliter la reproductibilité des expériences, nous utilisons dans ce paragraphe des données artificielles générées selon une configuration expérimentale commune. Les données sont simulées selon le modèle dRSM avec  $T = 10$  pas de temps,  $N = 300$  nœuds distribués dans  $K = 3$  ou 4 groupes latents et selon  $S = 2$  sous-graphes. En fonction des expérimentations, les réseaux sont générés soit avec des arêtes binaires (*i.e.*  $C=1$ ), soit avec des arêtes catégorielles à 2 types (*i.e.*  $C=2$ ).

Pour simplifier le processus de génération des données, nous supposons que les matrices  $A, B$  et  $V_0$ , servant à la génération de  $\gamma$ , sont fixées à la matrice identité  $I_{K-1}$ . De même,  $\Sigma = 0.1 \times K \times I_{K-1}$  et  $\Phi = 0.01 \times I_{K-1}$ . Enfin, le paramètre  $\Pi$ , qui définit les probabilités de connexion entre les groupes pour les différents types d'arêtes, est défini de sorte que la probabilité  $1 - \Pi_{ll}^0$  d'avoir une arête entre deux nœuds d'un même groupe soit plus grande que la probabilité  $1 - \Pi_{kl, k \neq l}^0$  d'avoir une arête entre des nœuds de groupes différents. La matrice  $\Pi^0$  est fixée de sorte que  $\text{diag}(\Pi^0) = (0.1, 0.4, 0.6)$  et  $\Pi_{kl}^0 = 0.99, \forall k \neq l$  dans les expériences. Notons qu'un tel choix de paramètres conduit à des réseaux structurés en communautés. En cas de connexion entre deux nœuds, les différents types d'arêtes sont équiprobables, *i.e.*  $\Pi_{kl}^{c \neq 0} = (1 - \Pi_{kl}^0)/C$ .

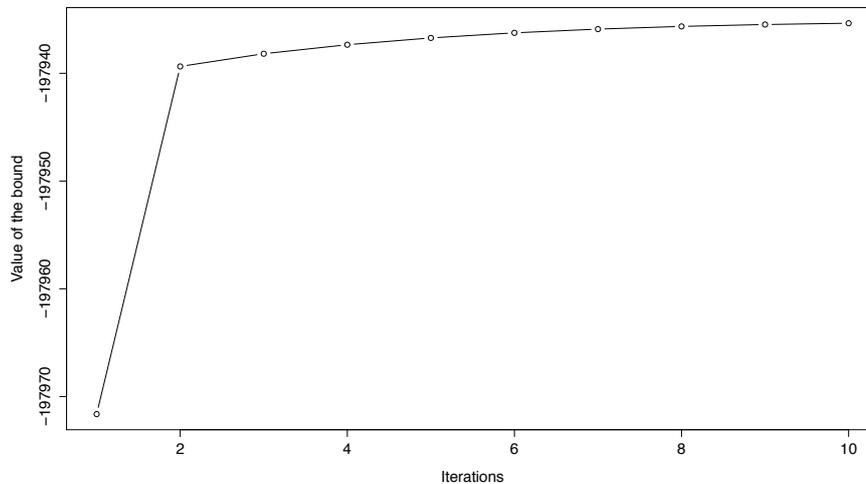


FIGURE 3. Evolution de la borne  $\mathcal{L}$  au cours des itérations de l'algorithme VEM sur l'exemple introductif.

#### 4.2. Exemple introductif

Nous considérons tout d'abord un exemple introductif sous la forme d'un unique réseau généré selon la configuration expérimentale détaillée ci-dessus. Le réseau considéré ici a des arêtes catégorielles à 2 types (*i.e.*  $C=2$ ). L'algorithme VEM que nous avons proposé dans les paragraphes précédents a été appliqué à ce réseau pour  $K = 3$  groupes.

La figure 3 permet tout d'abord d'observer l'évolution de la borne  $\mathcal{L}$  en fonction des itérations de l'algorithme VEM. Nous pouvons remarquer que l'algorithme VEM que nous avons proposé permet bien de maximiser à chaque étape la borne  $\mathcal{L}$ .

Il est ensuite possible de s'intéresser aux estimations des paramètres du modèle fournies par l'algorithme VEM. La table 2 permet de comparer les valeurs réelles de la matrice  $\Pi_{kl}^c$  pour  $c = (0, 1, 2)$  (partie de gauche) avec les valeurs estimées (partie de droite) par notre algorithme d'inférence. Les estimations apparaissent très proches des valeurs réelles.

De même, la figure 4 permet de comparer visuellement les estimations des proportions  $\alpha$  des groupes au cours du temps avec les valeurs simulées, et cela pour les  $S = 2$  sous-graphes. Les estimations sont représentées par des lignes pointillées et les proportions réelles sont représentées par des lignes pleines. A nouveau, l'algorithme VEM semble avoir réussi à retrouver la dynamique des proportions des groupes puisque les estimations sont très proches des valeurs simulées.

#### 4.3. Etude de l'influence de la taille du réseau

Nous nous intéressons à présent à l'influence de la taille du réseau (*i.e.* le nombre  $N$  de nœuds) sur la performance en classification et sur le temps de calcul. Pour ce faire, nous avons simulé des réseaux selon la configuration expérimentale commune avec  $K = 4$  groupes et ce pour des tailles de réseaux variant entre 100 et 400 nœuds. Pour chaque taille de réseaux, l'algorithme VEM a été appliqué à 20 réseaux simulés et la performance de classification a été mesurée grâce au critère ARI Rand (1971) en comparant la partition trouvée avec la partition simulée. Le critère ARI est

TABLEAU 2. Valeurs réelles (à gauche) et estimées (à droite) pour la matrice  $\Pi_{kl}^c$  avec  $c \in (0, 1, 2)$  (de haut en bas).

groupe	1	2	3	groupe	1	2	3
1	0.100	0.990	0.990	1	0.109	0.990	0.989
2	0.990	0.400	0.990	2	0.990	0.609	0.989
3	0.99	0.990	0.600	3	0.989	0.989	0.407
Valeurs réelles pour $C = 0$				Valeurs estimées pour $C = 0$			
groupe	1	2	3	groupe	1	2	3
1	0.450	0.005	0.005	1	0.442	0.004	0.005
2	0.005	0.300	0.005	2	0.004	0.197	0.005
3	0.005	0.005	0.200	3	0.005	0.005	0.294
Valeurs réelles pour $C = 1$				Valeurs estimées pour $C = 1$			
groupe	1	2	3	groupe	1	2	3
1	0.450	0.005	0.005	1	0.448	0.004	0.004
2	0.005	0.300	0.005	2	0.004	0.193	0.004
3	0.005	0.005	0.200	3	0.004	0.005	0.297
Valeurs réelles pour $C = 2$				Valeurs estimées pour $C = 2$			

TABLEAU 3. Temps moyens d'exécution de l'algorithme VEM en fonction de la taille  $N$  des réseaux dans le cas binaire (gauche) et catégoriel (droite) pour  $T = 10$  pas de temps.

Taille du réseau (N)	Temps d'exécution ( $C = 1$ )	Temps d'exécution ( $C = 2$ )
100	0.14 min	0.22 min
150	0.29 min	0.32 min
200	0.43 min	0.46 min
250	0.55 min	0.58 min
300	0.78 min	0.75 min
350	0.96 min	0.94 min
400	1.26 min	1.21 min

certainement la mesure d'adéquation la plus populaire actuellement en classification. Le critère se base sur l'étude de la classification de toutes les paires de nœuds entre les deux partitions. Une valeur proche de 1 pour l'ARI signifie que les partitions sont similaires et signifie dans notre cas que l'algorithme VEM réussit à retrouver la partition simulée.

La figure 5 présente, sous forme de boxplots, les valeurs du critère ARI en fonction de la taille  $N$  des réseaux dans le cas binaire (gauche) et catégoriel (droite). Il apparaît que dans le cas binaire ( $C = 1$ ), les résultats de classification sont satisfaisants pour  $N = 100$  et très bons à partir de  $N = 150$  nœuds. Dans le cas catégoriel, la tâche semble significativement plus difficile et il faut attendre des tailles de réseaux de l'ordre de  $N = 250$  nœuds pour obtenir de bons résultats de classification.

Enfin, la table 3 donne les temps moyens d'exécution de l'algorithme VEM en fonction de la taille  $N$  du réseaux dans le cas binaire (gauche) et catégoriel (droite) pour  $T = 10$  pas de temps. Il apparaît ici que l'algorithme VEM que nous proposons est exécutable sur un ordinateur de bureau en un temps raisonnable. Nous remarquons en outre que le passage au cas catégoriel n'implique pas un surcoût significatif par rapport au cas binaire.

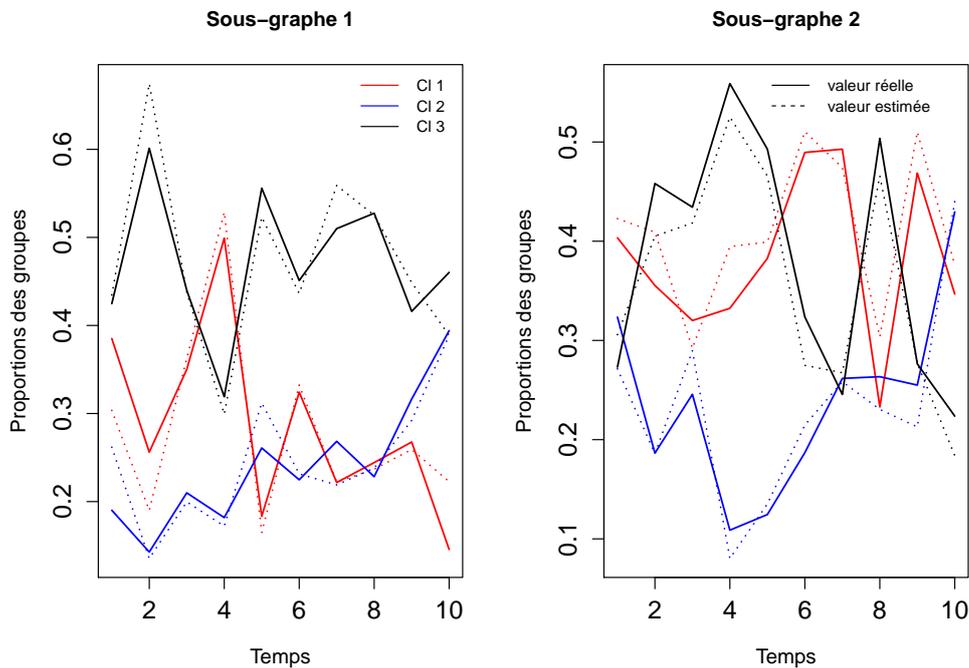


FIGURE 4. Evolution des valeurs réelles (lignes pleines) des proportions des groupes et leur estimation (lignes pointillées) au cours du temps.

## 5. Application aux réseaux du scandale Enron

Nous appliquons à présent la méthodologie décrite dans ce papier au jeu de données des emails Enron. L'entreprise Enron était spécialisée dans l'énergie (gaz, électricité, ...) et est devenue célèbre au début des années 2000 du fait d'un scandale financier lié à ses activités de courtage. Enron avait développé une activité de spéculation autour de l'électricité et des manipulations comptables visant à couvrir les pertes de l'entreprise ont mené à sa faillite en décembre 2001. A la suite de ce scandale financier, l'agence de régulation de l'énergie américaine a rendu public l'ensemble des emails de l'entreprise dans le cadre de ses investigations.

### 5.1. Données et protocole d'étude

Nous disposons ainsi de tous les échanges d'emails entre 148 personnes d'intérêt ayant travaillé pour l'entreprise à chaque temps  $t$ . Afin de faire apparaître les changements structurels intervenus au sein de l'entreprise, les données sont d'abord regroupées par mois, puis par période de temps, de manière à ce que deux personnes soient considérées comme connectées si elles ont échangé au moins un email pendant la période associée. Nous nous intéressons ici à cinq périodes de temps notées  $t_1, t_2, \dots, t_5$  (voir table 4), incluant les événements clés du scandale Enron. Nous avons choisi de découper les périodes de sorte que la densité du réseau soit suffisante à chaque période de temps. Les opérations de maquillage des pertes occasionnées par des opérations

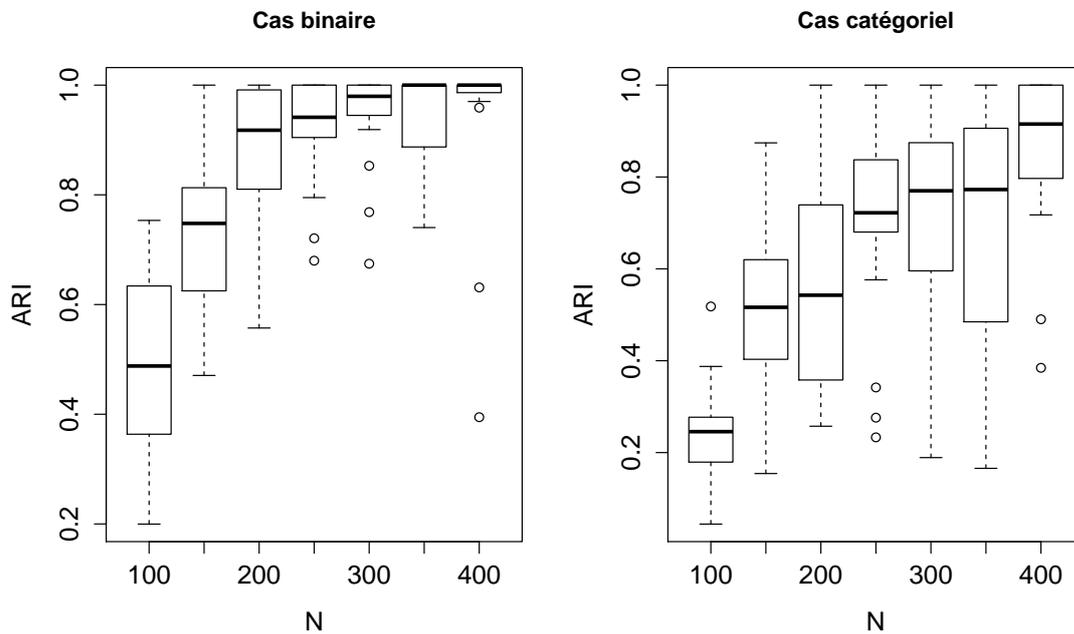


FIGURE 5. Valeurs du critère ARI en fonction de la taille  $N$  des réseaux dans le cas binaire (gauche) et catégoriel (droite).

TABLEAU 4. Périodes de temps considérées pour l'analyse des échanges d'emails au sein de l'entreprise Enron.

t	périodes
$t_1$	du 01/01/2000 au 01/12/2000
$t_2$	du 01/01/2001 au 01/03/2001
$t_3$	du 01/04/2001 au 01/06/2001
$t_4$	du 01/07/2001 au 01/09/2001
$t_5$	du 01/10/2001 au 01/03/2002

spéculatives furent révélées en octobre 2001 suite à l'ouverture d'une enquête par l'agence de régulation de l'énergie. L'entreprise fit finalement faillite en décembre 2001. Ces deux événements correspondent à la période  $t_5$ .

Nous disposons d'autre part d'une partition des employés en trois sous-graphes selon leur statut dans l'entreprise ( $s_1$  : cadre,  $s_2$  : employés,  $s_3$  : autres). Notons que le sous-graphe  $s_1$  comprend tous les cadres de l'entreprise, c'est à dire le directeur général, les présidents, vice-présidents, directeurs, managers, et directeurs managers. Les traders sont également associés à ce sous-graphe. Les individus du sous-graphe  $s_2$  sont tous employés par l'entreprise mais n'ont pas le statut de cadre. Enfin,  $s_3$  rassemble tous les autres individus en contact avec l'entreprise ayant également été touchés lors de la crise d'octobre 2001.

En résumé, le réseau considéré ici est dirigé, sans boucle et binaire :  $C = 1$  et  $X_{ij}^t = 1$  si  $i$  et  $j$  ont échangé au moins un email durant la période  $t$ , 0 sinon, avec  $t \in \{1, \dots, T\}$  et  $T = 5$ . Nous considérons en outre la partition connue du réseau en trois sous-graphes correspondant aux statuts des employés.

TABLEAU 5. Termes  $\Pi_{kl}^1$  de la matrice  $\Pi$  estimée à l'aide de l'algorithme variationnel EM pour  $K = 4$  groupes.

	groupe 1	groupe 2	groupe 3	groupe 4
groupe 1	0.478	0.037	0.005	0.023
groupe 2	0.020	0.181	0.006	0.012
groupe 3	0.001	0.002	0.001	0.003
groupe 4	0.012	0.012	0.024	0.119

## 5.2. Résultats

L'algorithme VEM décrit précédemment a été appliqué sur le jeu de données afin de rechercher  $K = 4$  groupes latents. Ce choix est motivé par des considérations empiriques et permet d'obtenir un modèle dRSM décrivant l'apparition et surtout la gestion de la crise suite à l'ouverture de l'enquête.

Parmi tous les résultats fournis par notre approche, intéressons-nous tout d'abord à la topologie des groupes obtenus. Le réseau étant binaire, le modèle dRSM utilisé correspond à un cas particulier du modèle décrit au paragraphe 2 où  $C = 1$ . Par construction la matrice  $\Pi$  vérifie  $\Pi_{kl}^0 + \Pi_{kl}^1 = 1, \forall(k, l)$  et par conséquent seules sont données les probabilités  $\Pi_{kl}^1$  d'apparition d'arêtes dans la table 5. Les groupes 1, 2, et 4 sont définis par des termes diagonaux significativement plus forts que les termes extra-diagonaux. Ces trois groupes correspondent donc à des communautés où la probabilité de connexion entre deux nœuds d'une même communauté est plus forte qu'entre des nœuds de communautés différentes. Ces groupes se distinguent principalement par le fait qu'ils ont des probabilités intra-groupe différentes de connexion. Ainsi, le groupe 1 est le groupe ayant la densité  $\Pi_{kk}^1$ , la plus forte, suivi du groupe 2 et du groupe 4. Finalement, notons que le groupe 3 est construit à partir de probabilités de connexion faibles. Il rassemble en fait tous les individus participant à des échanges d'emails peu structurés dans le réseau.

Le principal avantage du modèle dRSM est qu'il permet de caractériser l'évolution des sous-graphes en fonction de groupes latents estimés par l'approche d'inférence variationnelle. Toutes les proportions estimées sont données dans la figure 6 et nous nous concentrons ici sur l'interprétation de ces résultats. Comme indiqué précédemment le groupe 3 rassemble les nœuds peu structurés du réseau. Les personnes associées à ce groupe à un moment  $t$  échangent des emails avec d'autres personnes du réseau, sans profil type de connexion. Notons que l'augmentation de la proportion de ce groupe coïncide avec la diminution de la proportion du groupe 2 (densité intra-groupe moyenne), quelques soient les sous-graphes et temps  $t$ , et inversement. Ces deux proportions renseignent donc de manière inversée sur la structuration des échanges d'emails dans le réseau.

Nous observons également une chute importante de la proportion du groupe 3, dans tous les sous-graphes, entre  $t_4$  et  $t_5$  c'est à dire juste avant et après l'ouverture de l'enquête par l'organisme fédéral américain. Cette structuration du réseau est ici une réaction à la crise d'octobre 2001. Les personnes échangent des emails sur le sujet et contactent des personnes de manière préférentielle. Sur cette période, la proportion du groupe 4 (densité intra-groupe plus faible), comme celle du groupe 3, augmentent. Il est fondamental de noter que la structuration du réseau commence plus tôt (à  $t_3$ ) chez les cadres que chez les employés et les autres. Il y a bien une légère réaction à  $t_3$  pour ces deux derniers sous-graphes, mais elle disparaît à  $t_4$ . Ces observations laissent penser que les cadres ont eu connaissance des conditions d'arrivée de la crise avant les employés et les autres personnes considérées qui ont eu une légère réaction à  $t_3$ , ont été rassurées en  $t_4$ , et ont finalement

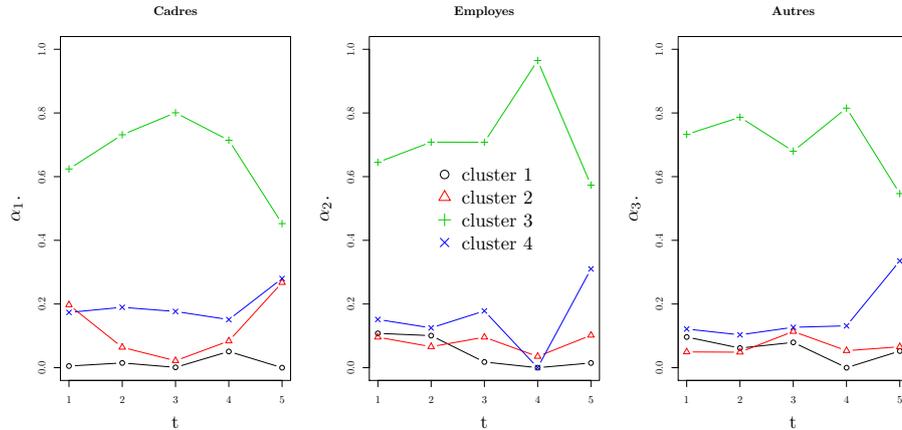


FIGURE 6. Proportions de chacun des  $K = 4$  groupes, à chaque temps  $t \in \{1, \dots, 5\}$ . Sous-graphe 1 (cadres), figure de gauche ; sous-graphe 2 (employés), figure du milieu ; sous-graphe 3 (autres), figure de droite.

réagi à  $t_5$ , plusieurs mois plus tard.

Intéressons-nous maintenant au groupe 1 (densité intra-groupe forte). Pour les sous-graphes 2 et 3, la proportion de ce groupe a une tendance générale à diminuer jusqu'à  $(t_4, t_5)$  où au contraire elle augmente. Cette remarque va également dans le sens d'une structuration du réseau liée à l'ouverture de l'enquête. Les cadres sont les seuls individus du réseau pour lesquels nous observons au contraire une diminution de la proportion du groupe 1 à ce moment là. En d'autres termes, le noyau dur des cadres, où l'échange d'emails se fait de manière très préférentielle, se désolidarise du reste réseau.

Finalement, la matrice  $\Phi$  estimée est de la forme  $\Phi = \sigma^2 I_{K-1}$  avec  $\sigma^2 = 7.11$ . Ces données sont donc marquées par une variance forte du processus temporel témoignant ainsi de changements profonds au cours du temps qu'un modèle dynamique tel que dRSM est capable de capturer.

## 6. Conclusion

Nous avons considéré dans ce travail le problème de l'analyse de réseaux dynamiques avec des arêtes catégorielles et pour lesquels une partition en sous-graphe est connue. Pour ce faire, nous avons proposé une extension au cadre dynamique du modèle RSM. Le nouveau modèle, appelé dRSM, est basé sur l'utilisation d'un modèle à espace d'état pour modéliser l'évolution au cours du temps des proportions des groupes latents. Un algorithme de type EM variationnel (VEM) a été proposé pour en effectuer l'inférence. Nous avons en particulier montré que les approximations variationnelles conduisent à un nouveau modèle à espace d'état à partir duquel les paramètres ainsi que les états cachés peuvent être estimés en utilisant le filtre de Kalman et le lisseur de Rauch-Tung-Striebel (RTS). Des expériences numériques ont permis de souligner les caractéristiques principales du modèle dRSM et ont démontré l'efficacité de l'algorithme VEM. La méthodologie a été finalement appliquée au jeu des données emails de l'entreprise Enron et a permis de mettre en évidence une réaction anticipée des cadres par rapport aux autres employés concernant le scandale à venir. Concernant les travaux futurs, nous souhaitons nous

intéresser au problème du choix du nombre  $K$  de groupes latents et cela pourrait être fait en utilisant une approche de choix de modèles. Cela permettrait également de sélectionner le modèle le plus adapté parmi les sous modèles de dRSM que l'on pourrait obtenir en ajoutant différentes contraintes.

## Références

- Ahmed, A. and Xing, E. P. (2007). On tight approximate inference of logistic-normal admixture model. *In Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1–8.
- Airoldi, E., Blei, D., Fienberg, S., and Xing, E. (2008). Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, 9 :1981–2014.
- Bickel, P. and Chen, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50) :21068–21073.
- Blei, D. and Lafferty, J. (2007a). A correlated topic model of science. *The Annals of Applied Statistics*, pages 17–35.
- Blei, D. and Lafferty, J. (2007b). A correlated topic model of science. *Annals of Applied Statistics*, 1(1) :17–35.
- Csisz, I., Tusnády, G., et al. (1984). Information geometry and alternating minimization procedures. *Statistics and decisions*.
- Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing*, 18(2) :173–183.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.
- Dubois, C., Butts, C., and Smyth, P. (2013). Stochastic blockmodelling of relational event dynamics. *In International Conference on Artificial Intelligence and Statistics*, volume 31 of the Journal of Machine Learning Research Proceedings, pages 238–246.
- Fienberg, S. and Wasserman, S. (1981). Categorical data analysis of single sociometric relations. *Sociological Methodology*, 12 :156–192.
- Girvan, M. and Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12) :7821.
- Handcock, M., Raftery, A., and Tantrum, J. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society : Series A (Statistics in Society)*, 170(2) :301–354.
- Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge, UK.
- Hathaway, R. J. (1986). Another interpretation of the EM algorithm for mixture distributions. *Statistics & Probability Letters*, 4(2) :53–56.
- Heaululani, C. and Ghahramani, Z. (2013). Dynamic probabilistic models for latent feature propagation in social networks. *In Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 275–283.
- Hofman, J. and Wiggins, C. (2008). Bayesian approach to network modularity. *Physical review letters*, 100(25) :258701.
- Jernite, Y., Latouche, P., Bouveyron, C., Rivera, P., Jegou, L., and Lamassé, S. (2014). The random subgraph model for the analysis of an ecclesiastical network in merovingian gaul. *Annals of Applied Statistics*, 8(1) :55–74.
- Jordan, M., Ghahramani, Z., Jaakkola, T., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2) :183–233.
- Krishnan, T. and McLachlan, G. (1997). *The EM algorithm and extensions*. John Wiley, New York.
- Lafferty, J. D. and Blei, D. M. (2006). Correlated topic models. In Weiss, Y., Schölkopf, B., and Platt, J., editors, *Advances in Neural Information Processing Systems 18*, pages 147–154. MIT Press.
- Latouche, P., Birmelé, E., and Ambroise, C. (2011). Overlapping stochastic block models with application to the french political blogosphere. *Annals of Applied Statistics*, 5(1) :309–336.
- Mariadassou, M., Robin, S., and Vacher, C. (2010). Uncovering latent structure in valued graphs : a variational approach. *Annals of Applied Statistics*, 4(2) :715–742.
- Matias, C. and Robin, S. (2014). Modeling heterogeneity in random graphs through latent space models : a selective review. *Esaim Proc. and Surveys*, 47 :55–74.
- Minka, T. (1998). From hidden markov models to linear dynamical systems. Technical report, MIT.
- Moreno, J. (1934). *Who shall survive ? : A new approach to the problem of human interrelations*. Nervous and Mental Disease Publishing Co.

- Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455) :1077–1087.
- Rand, W. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, pages 846–850.
- Rauch, H., Tung, F., and Striebel, T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIASS Journal*, 3(8) :1445–1450.
- Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, 7(2) :31–40.
- Svensén, M. and Bishop, C. (2004). Robust bayesian mixture modelling. *Neurocomputing*, 64 :235–252.
- Wang, Y. and Wong, G. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82 :8–19.
- White, H., Boorman, S., and Breiger, R. (1976). Social structure from multiple networks. i. blockmodels of roles and positions. *American Journal of Sociology*, pages 730–780.
- Xing, E., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2) :535–566.
- Xu, K. S. and Hero III, A. O. (2013). Dynamic stochastic blockmodels : Statistical models for time-evolving networks. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 201–210. Springer.
- Yang, T., Chi, Y., Zhu, S., Gong, Y., and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—a bayesian approach. *Machine learning*, 82(2) :157–189.

## Annexe A: Détails des calculs

### A.1. Détails des calculs de l'étape E du VEM

**Distribution  $q(Z)$  :** Afin de déterminer la distribution  $q(Z)$ , nous calculons l'espérance  $E_{\gamma, v, Z^i}$  par rapport aux lois variationnelles sur toutes les variables latentes privées du nœud  $i$ . Nous désignons ci après la probabilité que le nœud  $i$  appartienne au groupe  $k$  à un instant  $t$  par  $\tau_{ik}^{(t)}$ .

$$\begin{aligned}
\log q(Z_i) &= E_{\gamma, v, Z^i} [\log p(X|Z, \Pi) + \log h(Z, \gamma, \xi)] + \text{const} \\
&= \sum_{t=1}^T \sum_{k=1}^K Z_{ik}^{(t)} \left( \sum_{l=1}^K \sum_{c=0}^C \sum_{j \neq i}^N \delta(X_{i,j}^{(t)} = c) \tau_{jl}^{(t)} \left[ \log(\Pi_{kl}^c) + \log(\Pi_{lk}^c) \right] \right) \\
&\quad + \sum_{t=1}^T \sum_{k=1}^K \sum_{s=1}^S y_{is} E_{\gamma} \left[ Z_{ik}^{(t)} \log h(Z^{(t)}, \gamma^{(t)}, \xi^{(t)}) \right] + \text{const} \\
&= \sum_{t=1}^T \sum_{k=1}^K Z_{ik}^{(t)} \left( \sum_{l=1}^K \sum_{c=0}^C \sum_{j \neq i}^N \delta(X_{i,j}^{(t)} = c) \tau_{jl}^{(t)} \left[ \log(\Pi_{kl}^c) + \log(\Pi_{lk}^c) \right] \right) \\
&\quad + \sum_{t=1}^T \sum_{k=1}^K \sum_{s=1}^S y_{is} Z_{ik}^{(t)} E_{\gamma} \left[ \gamma_{sk}^{(t)} - \left( \xi_s^{-1(t)} \sum_{l=1}^K \exp(\gamma_{sl}^{(t)}) - 1 + \log(\xi_s^{(t)}) \right) \right] + \text{const} \\
&= \sum_{t=1}^T \sum_{k=1}^K Z_{ik}^{(t)} \left( \sum_{l=1}^K \sum_{c=0}^C \sum_{j \neq i}^N \delta(X_{i,j}^{(t)} = c) \tau_{jl}^{(t)} \left[ \log(\Pi_{kl}^c) + \log(\Pi_{lk}^c) \right] \right) \\
&\quad + \sum_{t=1}^T \sum_{k=1}^K \sum_{s=1}^S Z_{ik}^{(t)} y_{is} \left( \hat{\gamma}_{sk}^{(t)} - \left[ \xi_s^{-1(t)} \sum_{l=1}^K E(\exp(\hat{\gamma}_{sl}^{(t)}) - 1 + \log(\xi_s^{(t)})) \right] \right) + \text{const} \\
&= \sum_{t=1}^T \sum_{k=1}^K Z_{ik}^{(t)} \left( \sum_{l=1}^K \sum_{c=0}^C \sum_{j \neq i}^N \delta(X_{i,j}^{(t)} = c) \tau_{jl}^{(t)} \left[ \log(\Pi_{kl}^c) + \log(\Pi_{lk}^c) \right] \right) \\
&\quad + \sum_{s=1}^S y_{is} \left( \hat{\gamma}_{sk}^{(t)} - \left( \xi_s^{-1(t)} \sum_{l=1}^K \exp(\hat{\gamma}_{sl}^{(t)} + \frac{\hat{\sigma}_{sl}^{2(t)}}{2}) - 1 + \log(\xi_s^{(t)}) \right) \right) + \text{const}.
\end{aligned}$$

Les termes ne dépendant pas de  $Z_i$  ont tous été mis dans un terme générique const ci-dessus. En outre, concernant  $\gamma_{sk}^{(t)}$ , rappelons que  $\gamma_{sk}^{(t)} \sim \mathcal{N}(\hat{\gamma}_{sk}^{(t)}, \hat{\sigma}_{sk}^{2(t)})$  et, par conséquent, il est facile de montrer que :

$$\mathbb{E}[\exp(\gamma_{sk}^{(t)})] = \exp(\hat{\gamma}_{sk}^{(t)} + \frac{\hat{\sigma}_{sk}^{2(t)}}{2}).$$

On reconnaît alors la forme fonctionnelle d'une loi multinomiale :

$$q(Z_i^{(t)}) \sim \mathcal{M}(Z_i^{(t)}; 1, \tau_i^{(t)}), \forall i, t.$$

**Distribution  $q(v)$  :** On a :

$$\begin{aligned}
\log q(\mathbf{v}) &= E_{Z,\gamma} \left( \log p(\gamma_{\setminus K} | \mathbf{B}, \mathbf{v}, \Sigma) + \log p(\mathbf{v} | \mu_0, V_0, A, \Phi) \right) + \text{const} \\
&= \sum_t \sum_s \left( E_\gamma \left( \log \mathcal{N}(\gamma_{s \setminus K}^{(t)}; \mathbf{B}\mathbf{v}^{(t)}, \Sigma) \right) \right) + \log p(\mathbf{v}^{(1)} | \mu_0, V_0) \\
&\quad + \sum_{t=2}^T \log p(\mathbf{v}^{(t)} | \mathbf{v}^{(t-1)}, A, \Phi) + \text{const} \\
&= \sum_{t=1}^T \sum_{s=1}^S \left( E_\gamma \left( -\frac{1}{2} (\gamma_{s \setminus K}^{(t)})^\top \Sigma^{-1} (\gamma_{s \setminus K}^{(t)}) + (\gamma_{s \setminus K}^{(t)})^\top \Sigma^{-1} \mathbf{B}\mathbf{v}^{(t)} - \frac{1}{2} (\mathbf{v}^{(t)})^\top \mathbf{B}^\top \Sigma^{-1} \mathbf{B}\mathbf{v}^{(t)} \right) \right) \\
&\quad + \log p(\mathbf{v}^{(1)} | \mu_0, V_0) + \sum_{t=2}^T \log p(\mathbf{v}^{(t)} | \mathbf{v}^{(t-1)}, A, \Phi) + \text{const} \\
&= \sum_{t=1}^T \left( \sum_{s=1}^S \left( \hat{\gamma}_{s \setminus K}^{(t)} \Sigma^{-1} \mathbf{B}\mathbf{v}^{(t)} \right) - \frac{1}{2} (\mathbf{v}^{(t)})^\top \mathbf{B}^\top (\mathbf{S}\Sigma^{-1}) \mathbf{B}\mathbf{v}^{(t)} \right) \\
&\quad + \log p(\mathbf{v}^{(1)} | \mu_0, V_0) + \sum_{t=2}^T \log p(\mathbf{v}^{(t)} | \mathbf{v}^{(t-1)}, A, \Phi) + \text{const},
\end{aligned}$$

où toutes les variables qui ne dépendent pas de  $\mathbf{v}$  ont été mis dans le terme const. Nous reconnaissons la forme fonctionnelle de la distribution d'un modèle à espace d'état :

$$\begin{aligned}
\log q(\mathbf{v}) &= \sum_{t=1}^T \left( \log \mathcal{N} \left( \frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S}; \mathbf{B}\mathbf{v}^{(t)}, \frac{\Sigma}{S} \right) \right) \\
&\quad + \log p(\mathbf{v}^{(1)} | \mu_0, V_0) + \sum_{t=2}^T \log p(\mathbf{v}^{(t)} | \mathbf{v}^{(t-1)}, A, \Phi) + \text{const},
\end{aligned}$$

ou plus synthétiquement :

$$q(\mathbf{v}) \propto p(\mathbf{v}^{(1)} | \mu_0, V_0) \left[ \prod_{t=2}^T p(\mathbf{v}^{(t)} | \mathbf{v}^{(t-1)}, A, \Phi) \right] \left[ \prod_{t=1}^{(t)} \mathcal{N} \left( \frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S}; \mathbf{v}^{(t)}, \frac{\Sigma}{S}, \mathbf{B} \right) \right]. \quad (3)$$

Comme  $q(\mathbf{v})$  correspond à la loi a posteriori associée à un modèle à espace d'état, on a  $q(\mathbf{v}) = p(x, \mathbf{v} | \theta) / p(x | \theta)$  où  $p(x, \mathbf{v} | \theta)$  et  $p(x | \theta)$  désignent respectivement les vraisemblances jointes et simples du modèle. Le terme  $p(x, \mathbf{v} | \theta)$  correspond à (3) donc la normalisation de  $q(\mathbf{v})$  est exactement donnée par  $p(x | \theta)$ .

### A.2. Détails des calculs de la borne inférieure

Dans ce qui suit, nous notons  $x^{(t)} = \frac{\sum_{s=1}^S \gamma_{s \setminus K}}{S}$ . Le détail du calcul de la borne est comme suit :

$$\begin{aligned} \tilde{\mathcal{L}}(q, \theta, \xi) &= \sum_z \int_{\gamma} \int_{\mathbf{v}} q(Z, \gamma, \mathbf{v}) \log \frac{p(X|Z, \Pi) h(Z, \gamma, \xi) p(\gamma_K | B, \mathbf{v}, \Sigma) p(\mathbf{v} | \mu_0, A, \Phi, V_0)}{q(Z, \gamma, \mathbf{v})} d\mathbf{v} d\gamma \\ &= E_{Z, \gamma, \mathbf{v}} \left[ \log \frac{p(X|Z, \Pi) h(Z, \gamma, \xi) p(\gamma_K | B, \mathbf{v}, \Sigma) p(\mathbf{v} | \mu_0, A, \Phi, V_0)}{q(\gamma) q(\mathbf{v}) \prod_{i=1}^N q(Z_i)} \right] \\ &= E_Z (\log p(X|Z, \Pi)) + E_{Z, \gamma} (\log h(Z, \gamma, \xi)) + E_{\gamma, \mathbf{v}} (\log p(\gamma_K | B, \mathbf{v}, \Sigma)) \\ &\quad + E_{\mathbf{v}} (\log p(\mathbf{v} | \mu_0, A, \Phi, V_0)) - E_{\gamma} (\log q(\gamma)) - E_{\mathbf{v}} (\log q(\mathbf{v})) - E_Z (\log (\prod_{i=1}^N q(Z_i))). \end{aligned}$$

Comme,

$$E_{\mathbf{v}} (\log q(\mathbf{v})) = E_{\mathbf{v}} (\log p(\mathbf{v} | \mu_0, A, \Phi, V_0)) + E_{\mathbf{v}} (\log \mathcal{N}(\frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S} | B, \mathbf{v}^{(t)}, \frac{\Sigma}{S})),$$

alors,

$$E_{\mathbf{v}} (\log p(\mathbf{v} | \mu_0, A, \Phi, V_0)) - E_{\mathbf{v}} (\log q(\mathbf{v})) = -E_{\mathbf{v}} (\log \mathcal{N}(\frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S} | B, \mathbf{v}^{(t)}, \frac{\Sigma}{S})),$$

et nous pouvons réécrire  $\tilde{\mathcal{L}}(q, \theta, \xi)$  sous la forme :

$$\begin{aligned} \tilde{\mathcal{L}}(q, \theta, \xi) &= \sum_z \int_{\gamma} \int_{\mathbf{v}} q(Z, \gamma, \mathbf{v}) \log \frac{p(X|Z, \Pi) h(Z, \gamma, \xi) p(\gamma_K | B, \mathbf{v}, \Sigma) p(\mathbf{v} | \mu_0, A, \Phi, V_0)}{q(Z, \gamma, \mathbf{v})} \\ &= E_{Z, \gamma, \mathbf{v}} \left[ \log \frac{p(X|Z, \Pi) h(Z, \gamma, \xi) p(\gamma_K | B, \mathbf{v}, \Sigma) p(\mathbf{v} | \mu_0, A, \Phi, V_0)}{q(\gamma) q(\mathbf{v}) \prod_{i=1}^N q(Z_i)} \right] \\ &= E_Z (\log p(X|Z, \Pi)) + E_{Z, \gamma} (\log h(Z, \gamma, \xi)) + E_{\gamma, \mathbf{v}} (\log p(\gamma_K | B, \mathbf{v}, \Sigma)) \\ &\quad - E_{\gamma} (\log q(\gamma)) - E_{\mathbf{v}} (\log \mathcal{N}(\frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S} | B, \mathbf{v}^{(t)}, \frac{\Sigma}{S})) - E_Z (\log (\prod_{i=1}^N q(Z_i))) \\ &\quad + \log p(x | \theta'). \end{aligned}$$

Nous explicitons ci-dessous chacun des termes de la borne  $\tilde{\mathcal{L}}(q, \theta)$ . Le terme de log-vraisemblance  $\log p(x | \theta)$  du modèle SSM est quant à lui obtenu lors de l'inférence du modèle à espace d'état.

1.  $E_Z (\log p(X|Z, \Pi))$  :

$$\begin{aligned} E_Z (\log p(X|Z, \Pi)) &= \sum_{t=1}^T \sum_{k,l}^K \sum_{c=0}^C \sum_{i \neq j}^N E_z (\delta(X_{i,j}^{(t)} = c) Z_{ik}^{(t)} Z_{jl}^{(t)} \log(\Pi_{kl}^c)) \\ &= \sum_{t=1}^T \sum_{k,l}^K \sum_{c=0}^C \sum_{i \neq j}^N \delta(X_{i,j}^{(t)} = c) \tau_{ik}^{(t)} \tau_{jl}^{(t)} \log(\Pi_{kl}^c) \end{aligned}$$

2.  $E_{Z,\gamma}(\log h(Z, \gamma, \xi)) :$

$$\begin{aligned} E_{Z,\gamma}(\log h(Z, \gamma, \xi)) &= E_{Z,\gamma} \left[ \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N \sum_{s=1}^S y_{is} Z_{ik}^{(t)} \left( \gamma_{sk}^{(t)} - (\xi_s^{-1(t)} \sum_{l=1}^K \exp(\gamma_{sl}^{(t)})) \right. \right. \\ &\quad \left. \left. - 1 + \log(\xi_s^{(t)}) \right) \right] \\ &= \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N \sum_{s=1}^S y_{is} \left( \tau_{ik}^{(t)} \hat{\gamma}_{sk}^{(t)} - \tau_{ik}^{(t)} \xi_s^{-1(t)} \sum_{l=1}^K \exp(\hat{\gamma}_{sl}^{(t)} + \frac{\hat{\sigma}_{sl}^{2(t)}}{2}) \right. \\ &\quad \left. + \tau_{ik}^{(t)} - \tau_{ik}^{(t)} \log(\xi_s^{(t)}) \right) \\ &= \sum_{t=1}^T \sum_{s=1}^S \left( (r_s^{(t)})^T \hat{\gamma}_s^{(t)} - N_s \xi_s^{-1(t)} \sum_{l=1}^K \exp(\hat{\gamma}_{sl}^{(t)} + \frac{\hat{\sigma}_{sl}^{2(t)}}{2}) + N_s - N_s \log(\xi_s^{(t)}) \right) \end{aligned}$$

où  $r_s^{(t)}$  est désigne la quantité  $\sum_{i=1}^N \tau_i^{(t)} y_{is}$ .

3.  $E_{\gamma,v}(\log p(\gamma_K | B, v, \Sigma)) :$

$$\begin{aligned} E_{\gamma,v}(\log p(\gamma_K | B, v, \Sigma)) &= E_{\gamma,v} \left( \log \prod_{t=1}^T \prod_{s=1}^S \mathcal{N}(\gamma_{s \setminus K}^{(t)}; B v_s^{(t)}, \Sigma) \right) \\ &= \sum_{t=1}^T \sum_{s=1}^S \left( \log \mathcal{N}(\hat{\gamma}_{s \setminus K}^{(t)}; B \hat{v}_s^{(t)}, \Sigma) - \frac{1}{2} \text{tr}(\Sigma^{-1} B^T \hat{V}^{(t)} B) - \frac{1}{2} \text{tr}(\Sigma^{-1} \hat{\sigma}_s^{2(t)}) \right) \end{aligned}$$

où  $\hat{V}^{(t)}$  est la matrice de variance-covariance de la variable latente  $v^{(t)}$  sachant toutes les variables observées.

4.  $E_{\gamma}(\log q(\gamma)) :$

$$\begin{aligned} E_{\gamma}(\log q(\gamma)) &= E_{\gamma} \left( \log \prod_{t=1}^T \prod_{s=1}^S \prod_{k=1}^{K-1} \mathcal{N}(\gamma_{sk}^{(t)}; \hat{\gamma}_{sk}^{(t)}, \hat{\sigma}_{sk}^{2(t)}) \right) \\ &= \sum_{t=1}^T \sum_{s=1}^S \sum_{k=1}^{K-1} -\log \left( (2\pi)^{\frac{1}{2}} \hat{\sigma}_{sk}^{(t)} \right) - \frac{T(K-1)S}{2}. \end{aligned}$$

5.  $E_v(\log p(\frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S} | B, v^{(t)}, \frac{\Sigma}{S})) :$

$$E_v(\log p(\frac{\sum_{s=1}^S \hat{\gamma}_{s \setminus K}^{(t)}}{S} | B, v^{(t)}, \frac{\Sigma}{S})) = \sum_{t=1}^T \left( \log \mathcal{N}(x^{(t)}; B \hat{v}^{(t)}, \Sigma/S) - \frac{1}{2} \text{tr}(\Sigma^{-1} S B^T \hat{V}^{(t)} B) \right).$$

6.  $E_Z(\log(\prod_{i=1}^N q(Z_i))) :$

$$\begin{aligned} E_Z(\log(\prod_{i=1}^N q(Z_i))) &= \sum_{i=1}^N E_Z(\log q(Z_i)) \\ &= \sum_{i=1}^N E_Z(\sum_{t=1}^T \sum_{k=1}^K Z_{ik}^{(t)} \log(\tau_{ik}^{(t)})) \\ &= \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^K \tau_{ik}^{(t)} \log(\tau_{ik}^{(t)}). \end{aligned}$$