

Motivación

Los datos tensoriales constituidos por mediciones obtenidas en múltiples direcciones han surgido en muchas áreas científicas, como la genómica y neurología. Un objetivo común es investigar la estructura de dependencia condicional de las variables de cada dirección, teniendo en cuenta la estructura multidimensional de los datos. Los enfoques multivariados tradicionales no son adecuados para este tipo de datos altamente estructurados; debido a la ineficiencia, pérdida de información y falta de interpretabilidad. Por tanto existe la necesidad de establecer procedimientos que aprovechen la estructura inherente de los datos.

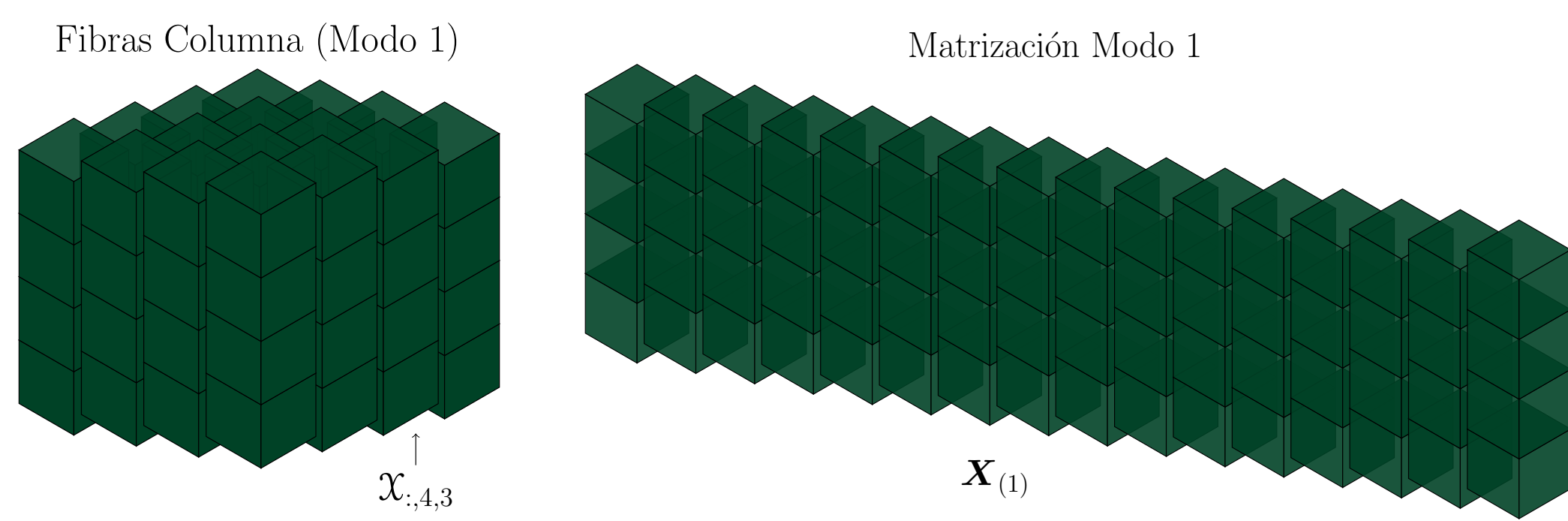


Figura 1: Matrización modo 1 de un tensor de orden tres.

Distribución Normal Tensorial

Definición 1. Un tensor aleatorio $\mathcal{X} \in \mathbb{R}^{q_1 \times \dots \times q_d}$ de orden d tiene una distribución normal tensorial, con tensor de medias $\mathcal{M} \in \mathbb{R}^{q_1 \times \dots \times q_d}$ y matrices de covarianzas $\Sigma_{q_k \times q_k} \succ 0, k = 1, 2, \dots, d$, denotada como $\mathcal{X} \sim TN_{q_1, \dots, q_d}(\mathcal{M}, \Sigma_1, \dots, \Sigma_d)$, si su función de densidad de probabilidad es

$$f(\mathcal{X}|\mathcal{M}, \Sigma_1, \dots, \Sigma_d) = \frac{\exp\left(-\frac{1}{2} \left\| \left[\mathcal{X} - \mathcal{M}; \Sigma^{-1/2} \right] \right\|^2\right)}{(2\pi)^{q/2} \prod_{k=1}^d |\Sigma_k|^{q/(2q_k)}} \quad (1)$$

donde $\|\cdot\|$ es la norma del tensor, $[\cdot; \cdot]$ el operador Tucker, $\Sigma^{-1/2} = \{\Sigma_1^{-1/2}, \dots, \Sigma_d^{-1/2}\}$ y $q = q_1 \dots q_d$.

El siguiente teorema resume algunas de las propiedades más interesantes de la distribución normal tensorial.

Teorema 1. $\mathcal{X} \sim TN_{q_1, \dots, q_d}(\mathcal{M}, \Sigma_1, \dots, \Sigma_d)$

1. si y solo si $\mathbf{X}_{(n)} \sim MN_{q_n, q/q_n}(\mathbf{M}_{(n)}, \Sigma_n, \Sigma_{-n})$, para $n = 1, 2, \dots, d$,
2. si y solo si $\text{vec}(\mathcal{X}) \sim N_{q_1 \dots q_d}(\text{vec}(\mathcal{M}), \Sigma_d \otimes \dots \otimes \Sigma_1)$,
3. $\mathbf{X}_{(n)}^{i_1} \perp \mathbf{X}_{(n)}^{i_2} | \mathbf{X}_{(n)}^{-\{i_1, i_2\}} \iff (\Sigma_n^{-1})_{i_1 i_2} = 0$, para $n = 1, 2, \dots, d$.
4. si $\mathbf{B}_k \in \mathbb{R}^{p_k \times q_k}$ para $k = 1, 2, \dots, d$ y $\mathcal{C} \in \mathbb{R}^{p_1 \times \dots \times p_d}$, entonces $\mathcal{Y} = \mathcal{C} + [\mathcal{X}; \mathbf{B}] \sim TN_{p_1, \dots, p_d}(\mathcal{C} + [\mathcal{M}; \mathbf{B}], \mathbf{B}_1 \Sigma_1 \mathbf{B}_1', \dots, \mathbf{B}_d \Sigma_d \mathbf{B}_d')$.

donde $\Sigma_{-n} = \bigotimes_{k \neq n} \Sigma_k$ y $\mathbf{X}_{(n)}^i$ denota la fila i de la matrización modo n del tensor \mathcal{X} .

DAG Tensorial

Definición 2. (Ni et al., 2017) La distribución conjunta de un tensor aleatorio \mathcal{X} se factoriza de acuerdo con un DAG Gaussiano tensor-variable (tDAG) si y solo si $\mathbf{y} = \text{vec}(\mathcal{X})$ cumple la ecuación

$$(\mathbf{A}_d \otimes \dots \otimes \mathbf{A}_1) \mathbf{y} = \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{D}_d \otimes \dots \otimes \mathbf{D}_1), \quad (2)$$

con la función de densidad de probabilidad (1), donde $\Omega_k = \mathbf{A}_k' \mathbf{D}_k^{-1} \mathbf{A}_k$ es la descomposición LDL de la k -ésima matriz de precisión.

- Problema de identificabilidad $\mathbf{D}_d \otimes \dots \otimes \mathbf{D}_1 = (c_1 \dots c_{d-1} \mathbf{D}_d) \otimes (c_{d-1}^{-1} \mathbf{D}_{d-1}) \otimes \dots \otimes (c_1^{-1} \mathbf{D}_1)$ puede resolverse fijando $\lambda_{k, q_k} = 1$, para $k = 1, 2, \dots, d-1$.
- La factorización (2) implica que

$$\mathbf{X}_{(k)}^i \perp \mathbf{X}_{(k)}^{\text{nd}(i) \setminus \text{pa}_k(i)} | \mathbf{X}_{(k)}^{\text{pa}_k(i)}$$

donde $\text{pa}_k(i)$ y $\text{nd}_k(i)$ son los padres y no descendientes del nodo i con respecto al grafo \mathcal{G}_{Ω_k} de la dirección $k = 1, 2, \dots, d$.

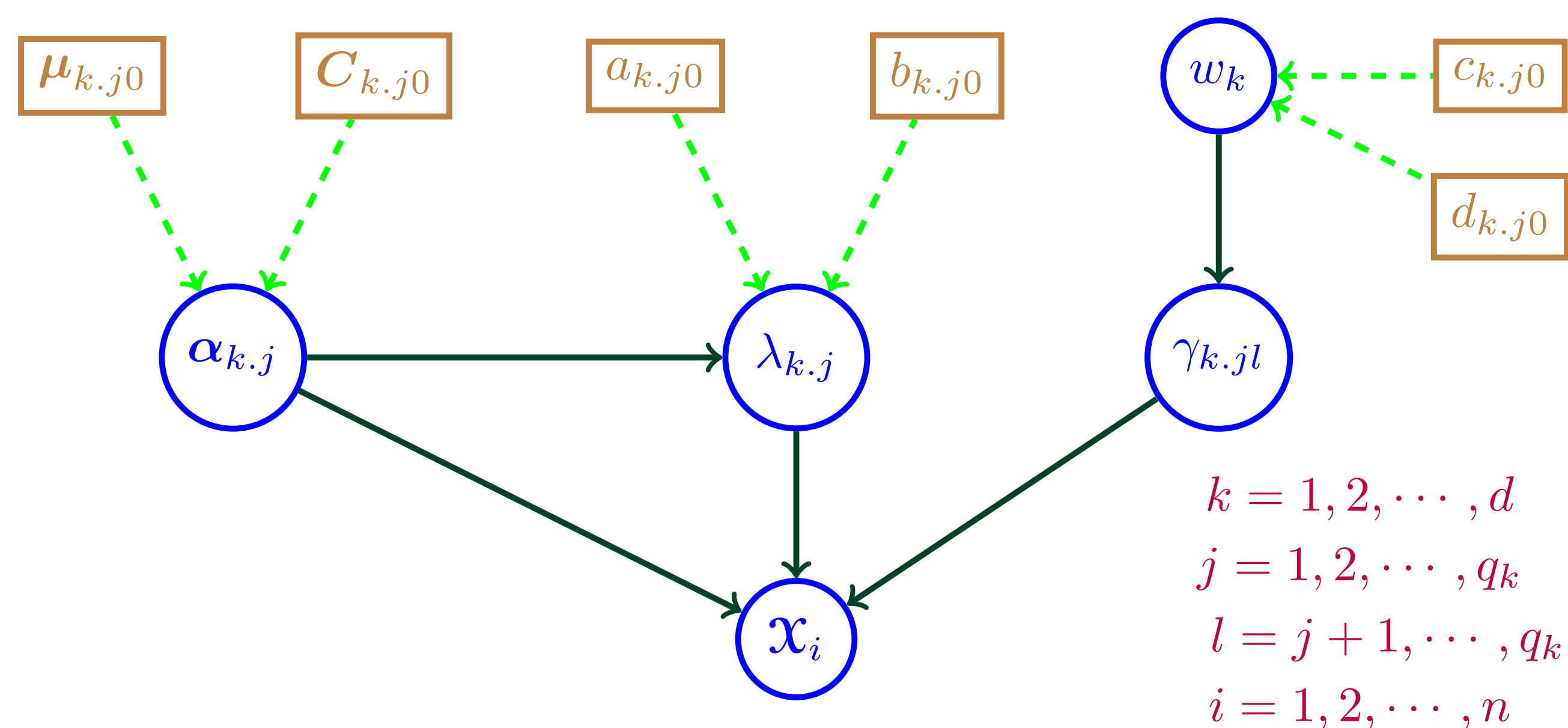


Figura 2: Modelo jerárquico para tDAG.

- Distribuciones a priori y condicionales completas para $\alpha_{k,j}$, $\lambda_{k,j}$, $\gamma_{k,jl}$ y w_k son distribuciones normales, gamas inversas, Bernoullis y betas, respectivamente.
- El número de parámetros a estimar en (2) es $\sum_{k=1}^d q_k^2 + 1$ para $d \geq 2$.
- Para un grafo no dirigido, siempre que se propone una nueva arista $\gamma_{k,jl}$ se debe verificar si el nuevo grafo es cordal y permutar los nodos del grafo de acuerdo con un orden perfecto.

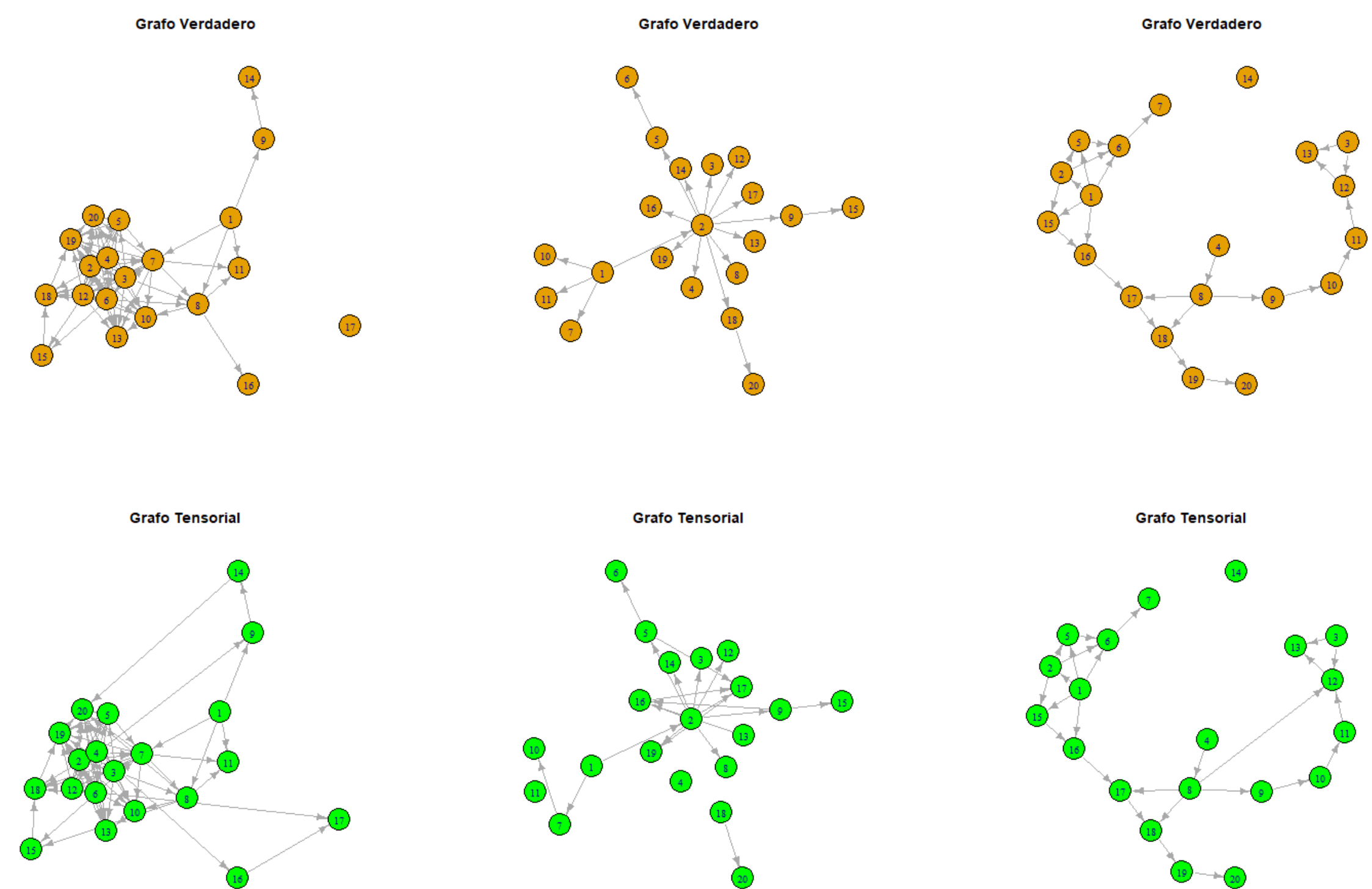


Figura 3: Estimación de grafo tensorial con tDAG.

Lasso Tensorial

Si $\mathcal{X}_1, \dots, \mathcal{X}_n \sim TN_{q_1, \dots, q_d}(\mathbf{0}, \Sigma_1^*, \dots, \Sigma_d^*)$, entonces el negativo de la log verosimilitud penalizada (hasta una constante) se puede escribir como

$$\ell_{\lambda}(\Omega) = \frac{1}{q_k} \text{tr}(\mathbf{S}_k \Omega_k) - \sum_{k=1}^d \frac{1}{q_k} \log |\Omega_k| + \sum_{k=1}^d P_{\lambda_k}(\Omega_k), \quad (3)$$

donde $\mathbf{S}_k = \frac{q_k}{nq} \sum_{i=1}^n \mathbf{Y}_{k,i} \mathbf{Y}_{k,i}'$ y $\mathbf{Y}_{k,i} = [\mathcal{X}_i; \Omega_k^{1/2}]_{(k)}$ con $\Omega_k = \mathbf{I}_{q_k}$, $P_{\lambda_k}(\cdot)$ es una función de penalización indexada por el parámetro de penalización λ_k .

- La función (3) no es conjuntamente convexa en $(\Omega_1, \dots, \Omega_d)$, pero es biconvexa.
- Si se usa penalización lasso $P_{\lambda_k}(\Omega_k) = \lambda_k \|\Omega_k\|_1$, se puede resolver usando el algoritmo lasso gráfico (Friedman et al., 2008).

Se puede demostrar que un algoritmo iterativo con base en lasso gráfico para optimizar (3) tiene las siguientes propiedades

- Estima la matriz de precisión verdadera hasta una constante.
- Genera un estimador consistente después de una única iteración, sin importa el estado inicial.
- Puede ser consistente incluso si solamente se tiene una observación, lo cual es frecuente en la practica.
- Al estimar la k -ésima matriz de precisión toma ventaja de la información de la j -ésima dirección ($j \neq k$) del dato tensorial.

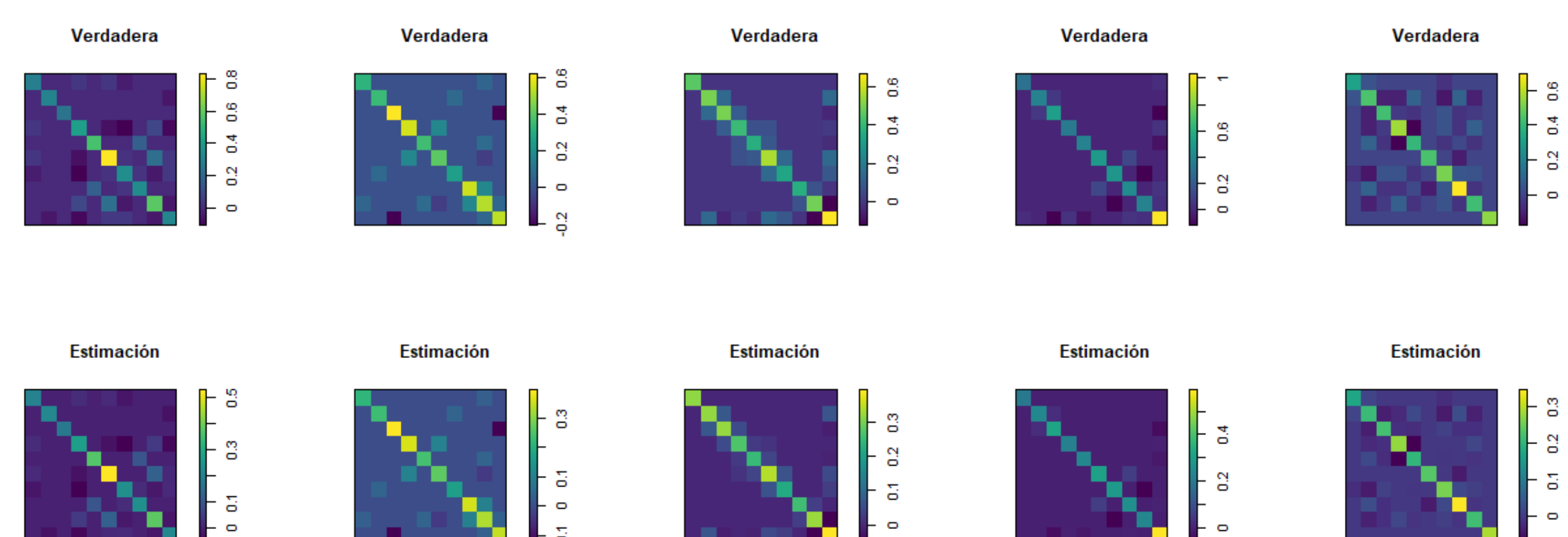


Figura 4: Estimación de matrices de precisión con TLasso.

Conclusiones

- Se han establecido dos procedimientos de estimación de modelos gráficos que aprovechan la estructura inherente de datos multidimensionales, donde uno de ellos permite combinar grafos dirigidos y no dirigidos.
- Se ha conseguido un algoritmo eficiente para generar observaciones normales de dimensiones altas, con matrices de covarianzas con estructura de producto de Kronecker.

Referencias

- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.
- Ni, Y., Stingo, F., and Baladandayuthapani, V. (2017). Sparse multi-dimensional graphical model: A unified Bayesian framework. *Journal of the American Statistical Association*, 112(518):779–793.