

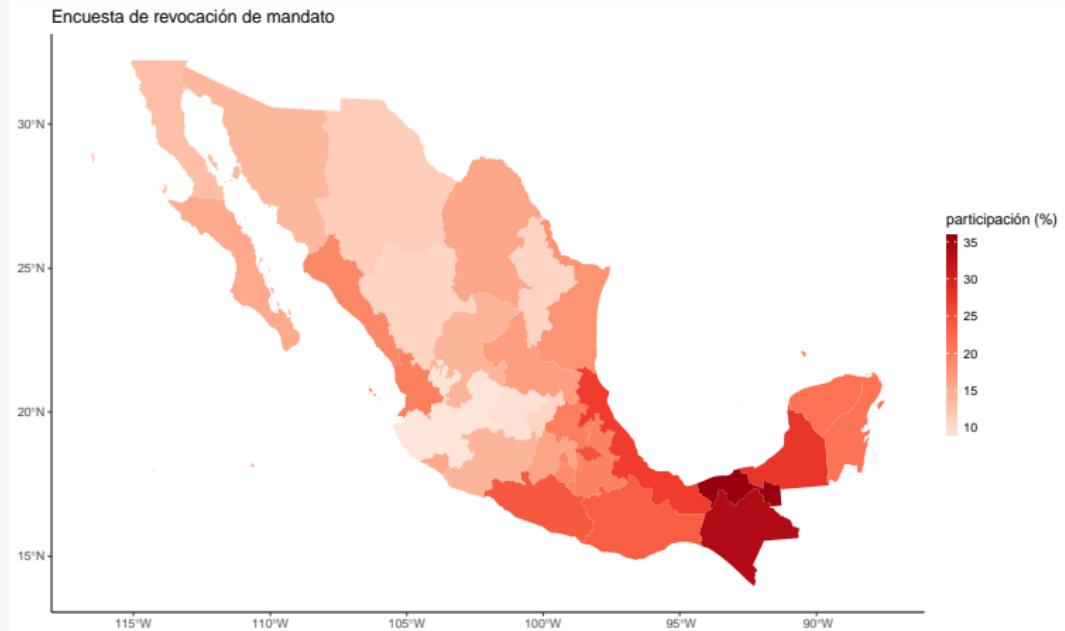
# **Modelos gráficos: Un vistazo a las estructuras de datos espacio-temporales**

**Rogelio Ramos Quiroga**

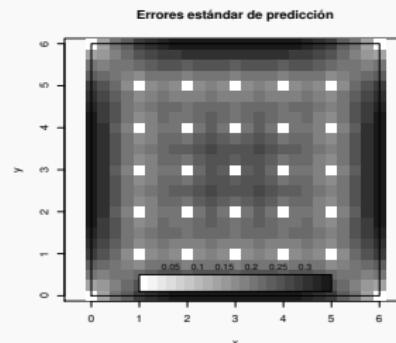
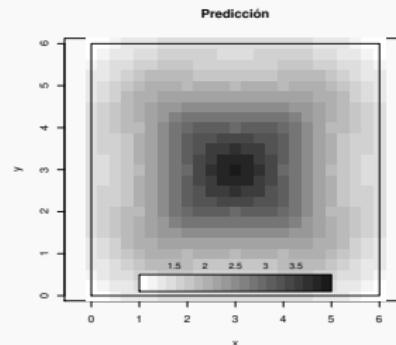
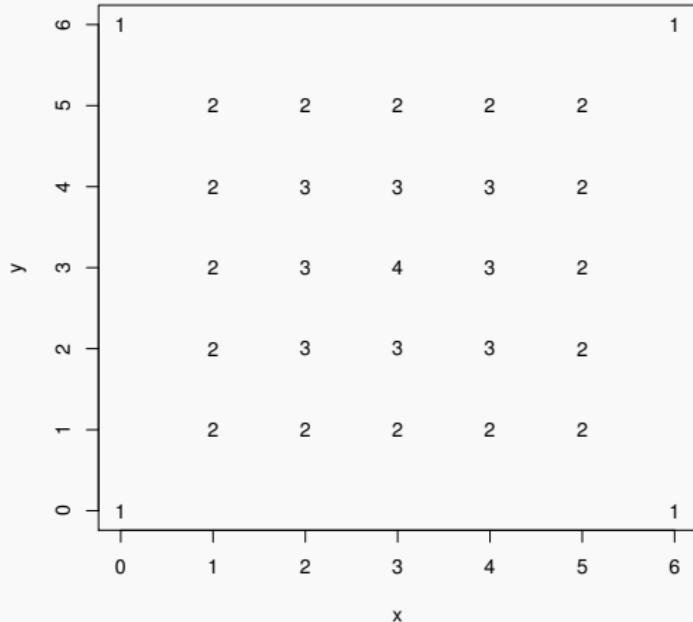
[raramosq@cimat.mx](mailto:raramosq@cimat.mx)

XX Escuela de Probabilidad y Estadística  
18-22 de Abril, 2022

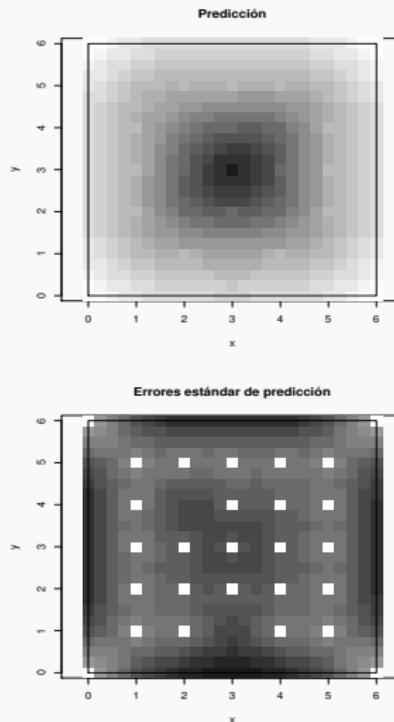
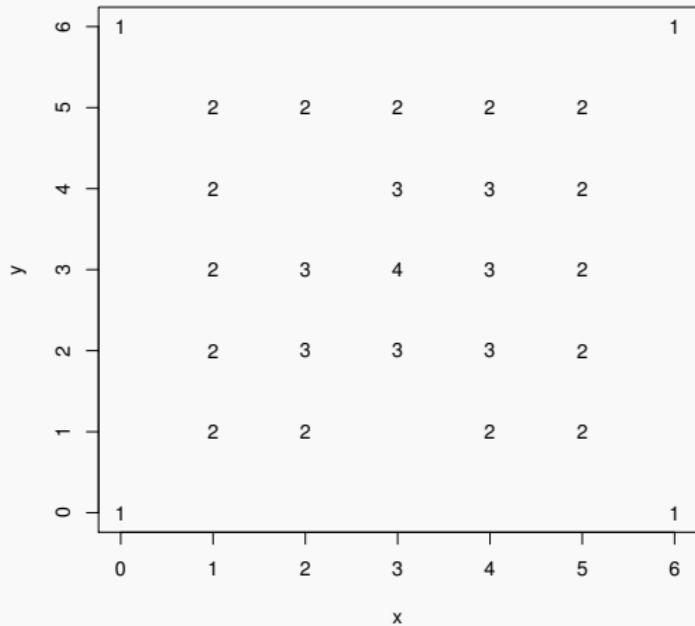
# Datos espaciales



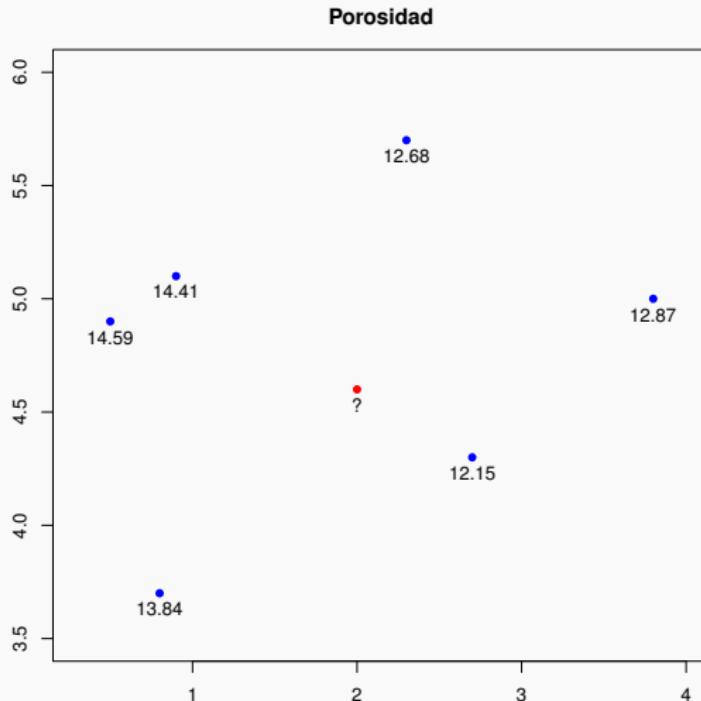
# Predicción espacial, versión simplificada 1



# Predicción espacial, versión simplificada 2



# Predicción espacial, versión no tan simplificada 1



## Predicción espacial, versión no tan simplificada 2

- Necesitamos estructura:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_6 \\ w \end{bmatrix} \equiv \begin{bmatrix} y \\ w \end{bmatrix} \sim N_7 \left( \begin{bmatrix} \mu_y \\ \mu_w \end{bmatrix}, \begin{bmatrix} V_{yy} & V_{yw} \\ V_{wy} & V_{ww} \end{bmatrix} \right)$$

- Entonces, una buena idea para predecir  $w$  en base a  $y$  es

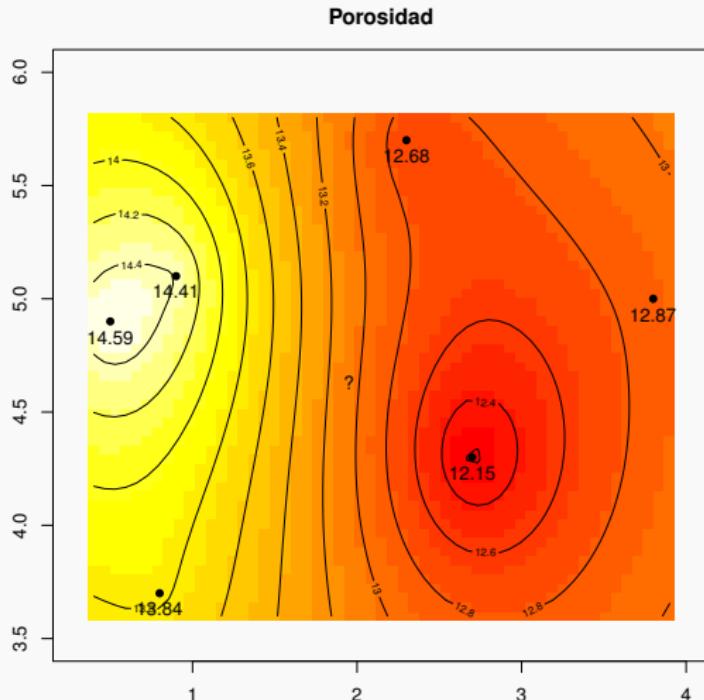
$$\hat{w} = E(w|y) = \mu_w + V_{wy} V_{yy}^{-1} (y - \mu_y)$$

el cual es un predictor con ciertas propiedades óptimas en el caso Normal y con buenas propiedades aún en el caso no-Normal<sup>1</sup>.

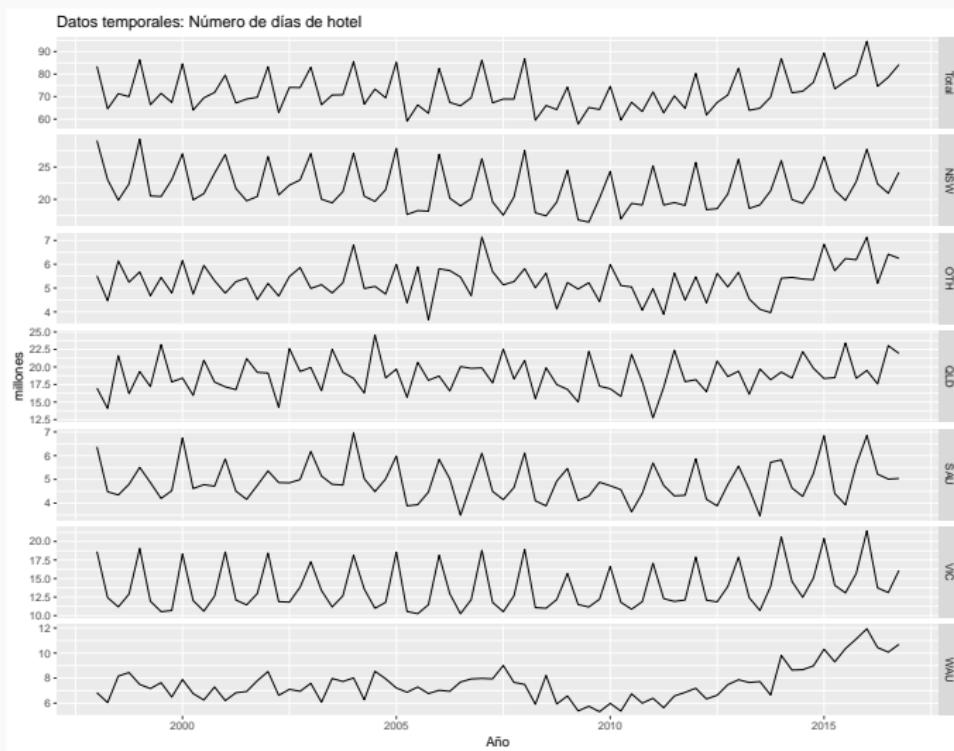
---

<sup>1</sup>Cressie, N. (1993) *Statistics for Spatial Data*. Wiley

# Predicción espacial, versión no tan simplificada 3



## Otra estructura de dependencia: Datos temporales<sup>2</sup>



<sup>2</sup>Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.

## Observaciones importantes

- Tanto en el caso espacial como en el temporal, los datos

$$y_1, y_2, \dots, y_n$$

**no** son i.i.d.

- Típicamente, sólo pueden considerarse como una sola observación

$$(y_1, y_2, \dots, y_n)$$

de una distribución  $n$ -variada (de ahí el interés agudo en suponer y verificar estacionariedad).

- Las estructuras de dependencia juegan un rol clave en la inferencia estadística.

## Datos espacio-temporales: Tasa de desempleo

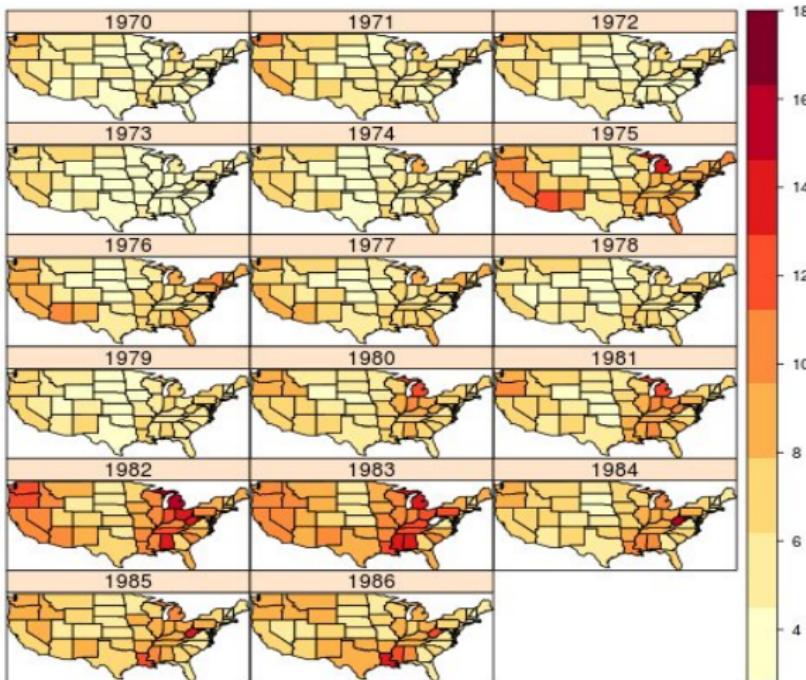


Figure 9: Unemployment rate per state, over the years 1970-1986.

# Datos espacio-temporales: Viento<sup>3</sup>

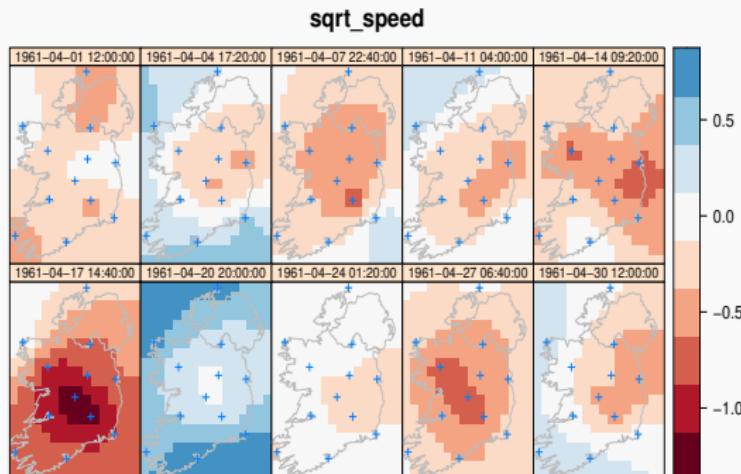


Figure 4: Space-time interpolations of wind (square root transformed, detrended) over Ireland using a separable product covariance model, for 10 time points regularly distributed over the month for which daily data was considered (April, 1961).

<sup>3</sup>Pebesma, E. (2012) spacetime: Spatio-Temporal Data in R. *Journal of Statistical Software*, 51, 7.

## Estructura de covarianza, datos espacio-temporales

- Proceso aleatorio  $Y(s, t)$  con función de covarianza separable<sup>4</sup>

$$\text{Cov}(Y(s, t), Y(s', t')) = C^S(s, s')C^T(t, t')$$

- Datos:  $Y = (Y_{ij})$  matriz  $m \times n$ , con  $Y_{ij} = Y(s_i, t_j)$
- Vectorización de matriz de datos

$$\text{Vec}(Y) =$$

$$(Y(s_1, t_1), \dots, Y(s_m, t_1), Y(s_1, t_2), \dots, Y(s_m, t_2), \dots, Y(s_1, t_n), \dots, Y(s_m, t_n))^T$$

- Covarianza: Producto de Kronecker de matrices de covarianza espacial y temporal

$$\text{Cov}(\text{Vec}(Y)) = C^T \otimes C^S$$

---

<sup>4</sup>No todo es separable: Cressie, N. & Huang, H. (1999) Classes of nonseparable, spatio-temporal stationary covariance functions. *JASA*.

## Una generalización: Datos espacio-temporales multivariados<sup>6</sup>

- Caso  $Y(s, t) \in \mathcal{R}^p$ . En cada punto espacial se recolecta información sobre  $p$  variables.
- Datos:  $Y = (Y_{ijk})$  arreglo  $m \times n \times p$  (Estructuras tensoriales<sup>5</sup>).
- Covarianza: Producto de Kronecker de matrices de covarianza espacial, temporal, multivariada

$$\text{Cov}(\text{Vec}(Y)) = C^T \otimes C^S \otimes C^M$$

---

<sup>5</sup>De la Peña Hnidey, Víctor Manuel (2022) Descomposición SVD para tensores (t-SVDM) y sus aplicaciones. Poster en XX Escuela de Probabilidad y Estadística. CIMAT.

<sup>6</sup>Bahadori, M.T., Yu, Q. & Liu, Y. (2014) Fast multivariate spatio-temporal analysis via low rank tensor learning. NIPS

## Modelos gráficos: Relaciones de dependencia

- Los modelos gráficos proveen la estructura para codificar las dependencias condicionales en un grupo de variables  $X_1, X_2, \dots, X_d$
- Han sido aplicados en muchos contextos, ciencias sociales, redes de regulación biológica, conectividad cerebral, etc.
- El concepto de dependencia condicional es mas adecuado para representar “ligas directas” entre variables, que otros conceptos de dependencia<sup>7</sup>.

---

<sup>7</sup>Giraud, C. (2015) *Introduction to High-Dimensional Statistics*. Chapman & Hall.

# Grafos

- Un grafo  $G = \{V, A\} \equiv \{\text{vértices, aristas}\}$
- No hay arista entre  $X$  y  $Y \iff X \perp Y \mid \text{resto}$
- $X \perp Y \mid Z \iff f(x, y|z) = f(x|z)f(y|z)$
- Una forma de codificar relaciones de dependencia condicional es mediante grafos no dirigidos

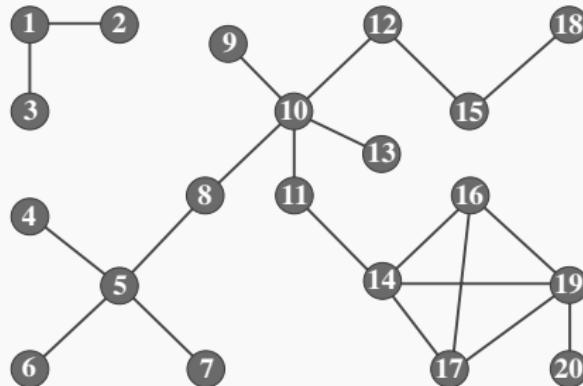


Figure 7.3 Non-directed graphical models:  $X_a \perp\!\!\!\perp \{X_b : b \not\sim a\} \mid \{X_c : c \sim a\}$ .

## Modelos gráficos no dirigidos

- Un resultado famoso (Teorema de Hammersley-Clifford) asegura que la distribución conjunta asociada a una gráfica no dirigida,  $g$ , es factorizable en “cliques”

$$f(x_1, x_2, \dots, x_d) = \prod_{c \in \text{cliques}(g)} f_c(x_c)$$

- En otras palabras, para conocer el comportamiento global, es suficiente conocer el comportamiento local de una gráfica mínima. Pero este es, en general, un problema difícil, el cual, sin embargo, es soluble en el caso Gaussiano.

## Caso Gaussiano

- La densidad Normal  $d$ -variada es

$$f(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left\{ -(x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

- Particionando  $x, \mu$  y  $\Sigma$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

- Puede verse que las condicionales son normales y que, por ejemplo:

$$E(X_1 | X_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2)$$

$$Var(X_1 | X_2) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

# Distribuciones condicionales bajo modelos Gaussianos

Veamos más de cerca la varianza condicional:

$$\text{Var}(X_1|X_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

La matriz de precisión,  $\Omega$ , es

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \Sigma^{-1} \equiv \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

Puede verse que (usando fórmulas para la inversa de una matriz particionada)

$$\Omega_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$$

## Distribuciones condicionales bajo modelos Gaussianos

Esto es,

$$\Omega_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}, \quad \text{o, equiv.} \quad \Omega_{11}^{-1} = \text{Var}(X_1|X_2)$$

Ahora, considere  $X_1$  de dimensión  $d_1 = 2$ . Vemos que

$$\text{Var}\left(\begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix} | X_2\right) = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}^{-1} = \frac{1}{\Delta} \begin{bmatrix} \omega_{22} & -\omega_{12} \\ -\omega_{21} & \omega_{11} \end{bmatrix}$$

El resultado importante es que

$$X_{11} \perp X_{12} | X_2, \quad \text{si y solo si} \quad \omega_{12} = 0$$

equivalentemente, **El nodo  $i$  es independiente del nodo  $j$ , condicionalmente al resto de los nodos, si y solo si  $\omega_{ij} = 0$**

# Estimación de modelos gráficos Gaussianos

Suponemos que se dispone de datos

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}_{n \times d}$$

bajo normalidad,  $x_i \sim N(\mu, \Omega^{-1})$  e independencia, la logverosimilitud perfil para  $\Omega$  es

$$\log L = \log |\Omega| - \text{tr}(S\Omega)$$

donde  $S$  es la varianza muestral

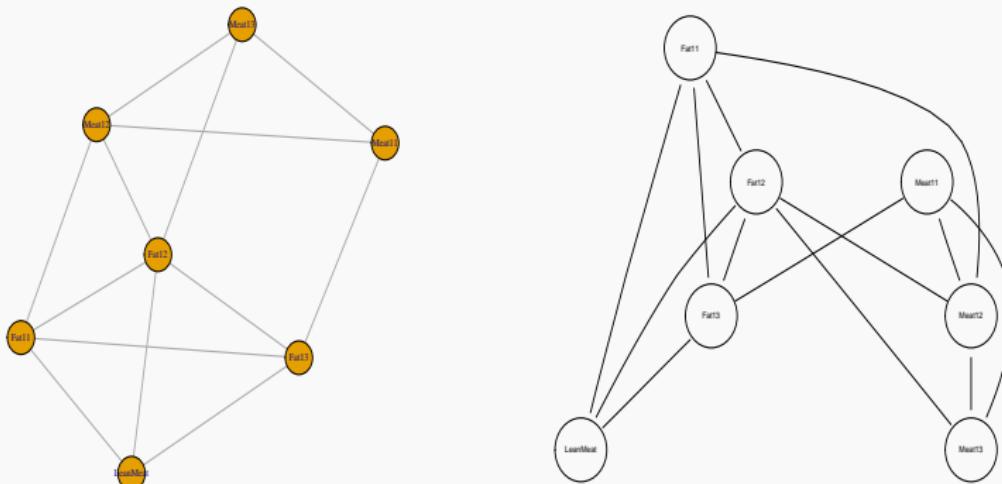
$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$

En este caso, se obtiene el máximo en forma analítica

$$\hat{\Omega} = S^{-1}$$

## Estimación de modelos gráficos Gaussianos

Usando algún umbral de corte para los elementos de  $\hat{\Omega}$ , o pruebas de hipótesis para conservar aristas significativas (ajustando por multiplicidad de pruebas<sup>8</sup>), podemos desplegar la estimación:



<sup>8</sup>Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society Series B*, Vol.85, 289–300.

## Estimación rala de modelos gráficos Gaussianos

Regularización es otra forma de inducir “sparseness” en un modelo gráfico, un ejemplo es el Lasso Gráfico:

$$\min_{\Omega > 0} \{-\log|\Omega| + \text{tr}(S\Omega) + \lambda\|\Omega\|_1\}, \quad \lambda \geq 0$$

- **Banerjee, O., El Ghaoui, L. & d'Aspremont, A.** (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data, *Journal of Machine Learning Research*
- **Friedman, J., Hastie, T. & Tibshirani, R.** (2008). Sparse inverse covariance estimation with the graphical Lasso, *Biostatistics*
- **Mazumder, R. & Hastie, T.** (2012). The Graphical Lasso: New insights and alternatives, *Electronic Journal of Statistics*

## Estimación rala de modelos gráficos Gaussianos

- Comentamos que la esperanza condicional es

$$E(X_1|X_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$$

- Considerando el caso  $d_1 = 1$  puede verse (s.p.g. medias cero) que las variables satisfacen un modelo de regresión

$$X_1 = \sum_{j \neq 1} \beta_j X_j + \epsilon$$

donde los coeficientes de regresión son funciones de las entradas de la matriz de precisión.

- Meinshausen y Bühlmann<sup>9</sup> usan este modelo básico para estimar en forma rala, usando regresión Lasso, los elementos de  $\Omega$ .

---

<sup>9</sup> Meinshausen, N. & Bühlmann, P. (2006) High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 1436–1462.

# Datos financieros<sup>10</sup>

Figure 9: Average monthly return of firms used in the empirical analysis of Section 5 over 3-year rolling windows spanning 1990 – 2012. In each window, 25 largest firms (in terms of market capitalization) from three sectors - Banks (BA), primary broker-dealers (PB), and insurance firms (INS), are included.

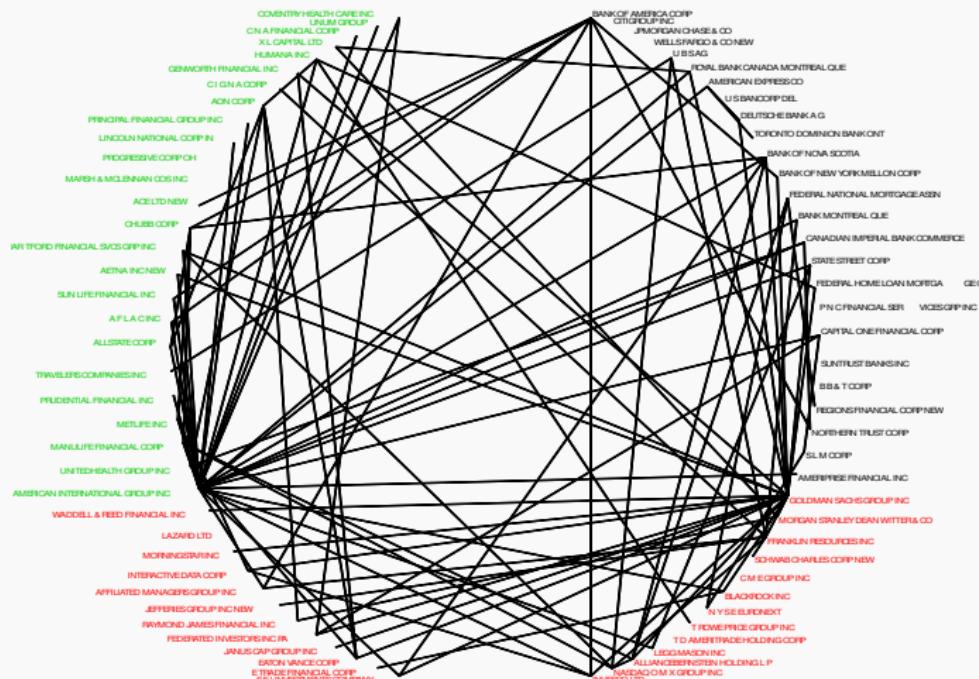



---

<sup>10</sup>Basu, S., Das, S., Michailidis, G. & Purnanandam, A. (2019) A system-wide approach to measure connectivity in the financial sector.  
<http://dx.doi.org/10.2139/ssrn.2816137>

# Interconectividad de instituciones financieras

200608-200907



DLVAR

## Interconectividad en regiones cerebrales<sup>12</sup>

- Datos públicos: “Autism Brain Imaging Data Exchange”
- Datos de fMRI en estado de reposo de 63 individuos
- 37 de ellos del grupo con ASD (Trastorno del espectro autista)
- 26 controles
- Cada sujeto tiene registros de resonancia magnética de 111 regiones de interés y 116 puntos en el tiempo:

$$X_i, \text{ matriz } 116 \times 111, \quad i = 1, \dots, 63$$

- Interés: Comparar la “conectividad” <sup>11</sup> del grupo ASD versus el grupo control

---

<sup>11</sup>Varoquaux, G., Gramfort, A., Poline, J.B. y Thirion, B. (2010). Brain covariance selection: better individual functional connectivity models using population prior, arXiv:1008.5071v4

<sup>12</sup>Urquiza Robles, Natalia Monserrath (2017) Comparación de modelos gráficos gaussianos con aplicación en datos de fMRI. Tesis de Maestría en Prob. y Est., CIMAT.

# Modelos gráficos individuales

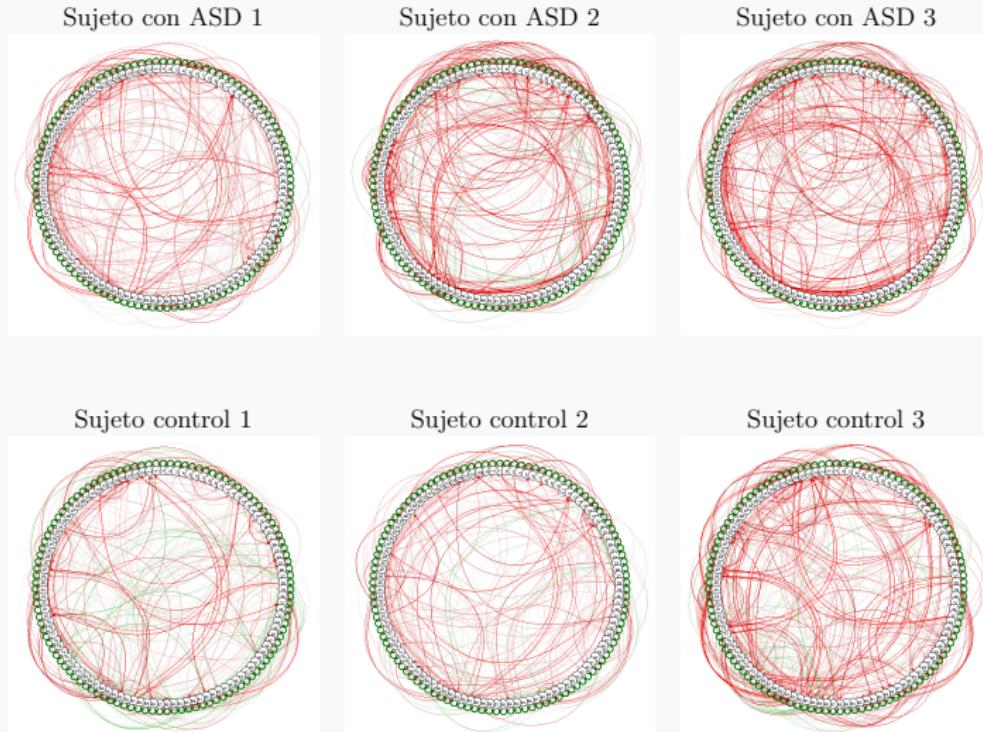


Figura 2. Gráfica estimada de 6 distintos sujetos.

## Hipótesis de interés: $H_0 : \Omega_A = \Omega_C$

- Las hipótesis de igualdad de conexión que fueron rechazadas son consistentes con teorías que se han desarrollados para personas con ASD
- Las aristas que resultaron tener más probabilidad de estar presentes en personas con ASD son aquellas que unen nodos cercanos, mientras que las aristas con menor probabilidad de existir en personas de este grupo unen nodos lejanos.
- Los modelos gráficos Gaussianos, además de aplicaciones como la presentada, se han usado en Estadística Espacial (Campos Aleatorios Markovianos), Ciencias Sociales y Economía.

# GRACIAS!!