



HAL
open science

Génération de modèles graphiques

Kévin Polisano, Irène Gannaz, Sophie Achard

► **To cite this version:**

Kévin Polisano, Irène Gannaz, Sophie Achard. Génération de modèles graphiques. GRETSI 2022 - XXVIII ème Colloque Francophone de Traitement Du Signal et Des Images, Sep 2022, Nancy, France. . hal-04256146

HAL Id: hal-04256146

<https://hal.science/hal-04256146>

Submitted on 27 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONTEXTE

Génération des matrices de corrélations partielles

Difficultés :

- ▶ Obtenir des matrices définies positives
- ▶ Imposer des structures sous-jacentes de graphes
- ▶ Obtenir des valeurs comparables à celles observées en pratique

Méthode utilisée : Córdoba et al. [1].

MODÈLES GRAPHIQUES

Définition

- $\Sigma \in \mathbb{R}^{p \times p}$ matrice de corrélation
- $\Omega = \Sigma^{-1} = (\omega_{ij})$ matrice de corrélation partielle

Modèle graphique : graphe avec p nœuds et arêtes E , $E = \{(i, j), i \sim j\}$,
 $i \sim j$, si $\omega_{ij} \neq 0$.

Objectif : pour $E \subset \mathbb{R}^2$ donné, construire Ω^E telle que pour tout $(i, j) \notin E$, $\omega_{ij}^E = 0$.

Contrainte de positivité : décomposition de Cholesky.

- ▶ Idée : $\Omega = \mathbf{U}\mathbf{U}^\top$ avec $\mathbf{U} \in \mathcal{U}_1^p$, où \mathcal{U}_1^p est l'ensemble des matrices triangulaires supérieures de dimension $p \times p$, à diagonale positive et de lignes normalisées à 1.
- ▶ Modèle graphique : $\mathcal{U}_1^p(G) \subset \mathcal{U}_1^p$ le sous-ensemble où $u_{ij} = 0$ pour tout $(i, j) \notin E$.

On introduit

$$\mathcal{E}^p(G) = \{\mathbf{U}\mathbf{U}^\top \mid \mathbf{U} \in \mathcal{U}_1^p \text{ et } u_{ij} = 0 \text{ si } (i, j) \notin E\} \subset \mathcal{E}^p.$$

Contrainte du graphe sous-jacent

Supposons que le graphe G est **ordonné**, $1 \prec 2 \prec \dots \prec p$ avec $\{i \prec j\} \subseteq \text{pa}(i) = \{j : \omega_{ij} \neq 0\}$ les parents du nœud i . Soient $\{X_i, i = 1, \dots, p\} \sim \mathcal{N}(0, \Sigma)$, alors

$$X_i = \sum_{j \in \text{pa}(i)} \beta_{ij} X_j + \varepsilon_i = \sum_{j \succ i} \beta_{ij} X_j + \varepsilon_i,$$

où $j \succ i$ si $j \in \text{pa}(i)$. En ce cas, Pourhamadi [4] fait remarquer que $\beta_{ij} = -\frac{u_{ij}}{u_{ii}}$. Nous avons ainsi

$$\omega_{ij} = 0 \iff u_{ij} = 0.$$

Conclusion : $\mathcal{E}^p(G)$ donne bien des matrices de corrélation partielles associées au graphe G si le graphe est ordonné.

ALGORITHME DE SIMULATION

Méthode de Córdoba et al. [1] :

1. Triangulation du graphe $G \mapsto \tilde{G}$ rendu *cordal*.
2. Ordonnement de \tilde{G} .
3. Simulation de $\Omega^{\tilde{E}}$ via la décomposition de Cholesky et le changement de variable :

$$\Phi : \begin{matrix} \mathcal{U}_1^p(\tilde{G}) & \longrightarrow & \mathcal{E}^p(\tilde{G}) \\ \mathbf{U} & \longmapsto & \mathbf{U}\mathbf{U}^\top \end{matrix}$$

4. Construction de la matrice \mathbf{U} finale associée à E en fixant $u_{ij} = 0$ si $(i, j) \in \tilde{E} \setminus E$. Pour tout $i = 1, \dots, p$, orthogonalisation et renormalisation de l'ensemble $\{(u_{ij})_j, (i, j) \notin E, i \succ j\}$.

- ▶ Échantillonner uniformément $\Omega^{\tilde{E}}$ sur $\mathcal{E}^p(\tilde{G})$ en échantillonnant $\mathbf{U} \in \mathcal{U}_1^p(\tilde{G})$ à partir d'une densité proportionnelle au jacobien de Φ [3] :

$$\det(J\Phi(\mathbf{U})) = 2^p \prod_{i=1}^p u_{ii}^{\text{pa}(i)+1}.$$

- ▶ L'algorithme de Metropolis-Hastings permet, grâce à la factorisation ci-dessus, d'échantillonner de manière indépendante les lignes de \mathbf{U} à partir d'une densité $f(\mathbf{u}_i) \propto u_{ii}^i$.

RÉPARTITION DES CORRÉLATIONS

Données simulées

- ▶ 4 structures de graphe : Erdős-Rényi, Barabási-Albert, "Petit-monde" (*small-world*) et Blocs stochastiques à 2 communautés.
- ▶ Différents degré de sparsité : $d = |E|/p \in \{0.1, 0.5\}$
- ▶ 2 sources d'aléatoire :
 - génération de 50 graphes dans chaque classe;
 - échantillonnage de 10 matrices de corrélation $\Omega^{\tilde{E}}$ pour chaque graphe.

Données de neurosciences

- ▶ Données d'IRMf sur 4 rats morts [2].
- ▶ 51 séries temporelles, correspondant chacune à une région cérébrale \Rightarrow Corrélation empirique entre les coefficients d'ondelettes à l'échelle 4 (fréquence [0.06 ; 0.12] Hz).

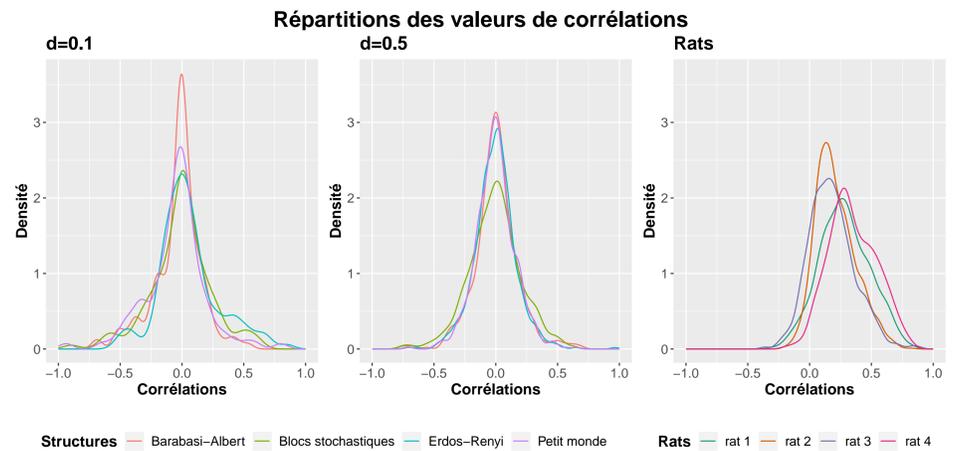


FIGURE 1 – Distributions des corrélations théoriques obtenues par simulations et empiriques sur données réelles. Les graphes ont $p = 51$ nœuds.)

- ▶ Plus il y a d'arêtes, plus les valeurs de corrélation sont faibles (évolution non régulière).
- ▶ Avec (p, d) égaux, les graphes de Erdős-Rényi et de blocs stochastiques ont les valeurs les plus élevées, et les graphes de Barabási-Albert plus faibles.
- ▶ Données simulées symétriques / Données empiriques fortement asymétriques.
- ▶ Modification nécessaire de l'algorithme pour reproduire des matrices réalistes.

ETUDE DE L'ÉTAPE 1

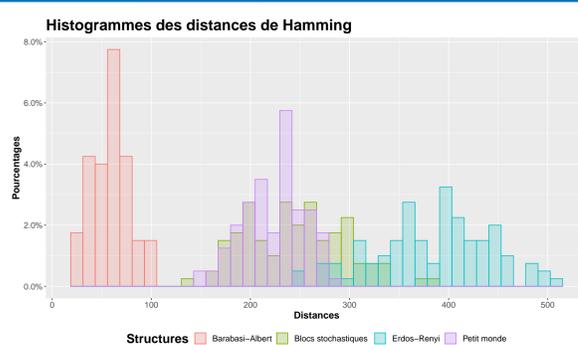


FIGURE 2 – Histogrammes des distances de Hamming entre les graphes G et les graphes \tilde{G} avec $(p, d) = (51, 0.1)$.

Biais possible lors de l'étape 1 de l'algorithme transformant G en \tilde{G} .

- ▶ Écarts entre les graphes G et \tilde{G} variables pour les différentes structures de graphes.
- ▶ Petit-monde et Blocs stochastiques ont des comportements similaires à l'étape 1 mais pas dans les distributions : les écarts observés sont *a priori* bien liés à la structure.

CONCLUSION

- ▶ Étude de l'algorithme proposé dans Córdoba et al. [1] et permettant la simulation de matrices de corrélations associées à des graphes.
- ▶ Mise en évidence de l'influence des structures de graphes sur les valeurs de corrélations générées, à nombres d'arêtes et de nœuds fixés.
- ▶ Limitations à des valeurs de corrélations générées faibles et symétriques, que l'on retrouverait également dans la méthode des oignons (n'autorisant pas de contrainte de graphe).
- ▶ Perspective d'étendre la contrainte de graphe sous-jacent à d'autres algorithmes de génération de matrices de corrélations.

REFERENCES

- [1] Córdoba, I., Varando, G., Bielza, C. et Larrañaga, P. (2020). On generating random Gaussian graphical models. *International Journal of Approximate Reasoning*, 125, 240–250.
- [2] Becq, G., J-PC, Habet, T., Collomb, N., Faucher, M., Delon-Martin, C., Coizet, V., Achard, S. et Barbier, E.L. (2020). Functional connectivity is preserved but reorganized across several anesthetic regimes. *NeuroImage*, 219, 116945.
- [3] Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10), 2177–2189.
- [4] Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data Unconstrained parameterisation. *Biometrika*, 86(3), 677–690.