HAL
open science

Graphe de connectivité cérébrale et longue dépendance

Florent Chatelain, Sophie Achard, Cedric Gouy-Pailler, Olivier J.J. Michel, Pierre-Olivier Amblard

To cite this version:

Florent Chatelain, Sophie Achard, Cedric Gouy-Pailler, Olivier J.J. Michel, Pierre-Olivier Amblard. Graphe de connectivité cérébrale et longue dépendance. GRETSI 2011 - XXIIIème Colloque francophone de traitement du signal et des images, Sep 2011, Bordeaux, France. hal-00639898

HAL Id: hal-00639898

https://hal.science/hal-00639898v1

Submitted on 10 Nov 2011

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.

Graphe de connectivité cérébrale et longue dépendance

Florent CHATELAIN[1], Sophie ACHARD[1], Cédric GOUY-PAILLER[2], Olivier MICHEL[1], Pierre Olivier AMBLARD[1,3]
[1]GIPSA-lab, CNRS, Université de Grenoble,
38402 Saint Martin d'Hères Cedex, FRANCE
prenom.nom@gipsa-lab.grenoble-inp.fr
[2]CEA, LIST,
Laboratoire d'Outlis pour l'Analyse de Données,
91191 Gif-sur-Yvette CEDEX, FRANCE
cedric.gouy-pailler@cea.fr
[3] Dept. Math&Stat, The University of Melbourne,
Parkville, VIC3010, Australia, bamblard@unimelb.edu.au

Résumé – Nous proposons une approche d'estimation au sens du maximum de vraisemblance de la matrice de corrélation partielle pour un modèle graphique gaussien décomposable, calculée à partir de séries temporelles présentant des caractéristiques de longue mémoire. Les estimateurs sont construits sur les séries de coefficients d'ondelettes des signaux à une échelle donnée, permettant de s'affranchir des biais d'estimation des corrélations introduits par la présence de longue mémoire. L'utilisation des graphes décomposables (markovien) permet de développer une forme explicite de l'estimateur de max. de vraisemblance de la matrice de corrélations partielles. La cohérence et la précision des résultats obtenus sur des données (surrogate data) de simulations réalistes de données IRMf illustrent le potentiel de cette approche; nous montrons que la décomposabilité 'forcée' des modèles graphiques utilisés n'affecte pas la qualité des estimations obtenues, même pour des ensembles de données de petite taille : les coefficients correspondant à des liens artificiellement introduits pour garantir la décomposabilité du graphe ne prennent que des valeurs non significatives. L'intérêt des modèles décomposables réside dans la possibilité de décomposer le problème d'estimation clique par clique, intégrant ainsi explicitement les contraintes de dépendance conditionnelle.

Abstract – In this communication, a decomposable graph based partial correlation matrix estimator is proposed for the analysis of time series exhibiting long memory properties. The estimator is derived from the set of wavelet coefficients computed at a given scale; the obtained coefficients are asymptotically uncorrelated, thus avoiding to introduce bias in the estimation. Decomposability of the graph leads to derive an explicit formulation of the maximum likelihood estimation of the partial correlation matrix; however, some spurious links must be introduced in order to insure graph decomposability. Surrogate data, that have similar properties as fMRI records, are used to illustrate the pertinence and accuracy of the proposed approach. The simulations show that even for small size samples, the proposed methods outperform the classical approaches. Furthermore, the added links do not affect the quality of the estimation. An appealing feature of decomposabilty resides in the possibility to account for conditional dependences through a clique based approach.

1 Introduction, motivation

Depuis la Grèce antique, l'étude du cerveau humain en fonctionnement a soulevé des débats enflammés. Si le XXème siècle aura notamment été le théâtre d'affrontements virulents entre partisans des théories *localisationnistes* (à chaque zone cérébrale est associée une fonction spécifique) et *holistiques* (le cerveau fonctionne comme un tout et aucun déficit spécifique ne peut être associé à une lésion localisée), ce siècle aura surtout vu l'avènement des dogmes portés par la 'neuroscience'. Cette approche tente la synthèse des hypothèses les plus plausibles des deux théories précédentes. Elle propose une modélisation globale du cerveau, appréhendé comme un système complexe fonctionnant de manière hiérarchique et s'appuyant sur des unités élémentaires présentant à la fois certaines formes de spécificité et de plasticité. L'avènement de l'ère neuroscientifique n'aurait pas été possible sans les progrès réalisés dans le domaine de l'imagerie cérébrale. Notamment, l'Imagerie fonctionnelle par Résonance Magnétique (IRMf) permet aujourd'hui d'observer le fonctionnement global du cerveau pendant plusieurs minutes avec des résolutions temporelle et spatiale de l'ordre de la seconde et du 1mm^3 (voxel) respectivement. Une modélisation et une caractérisation du cerveau par un graphe fonctionnel devient alors possible [**?**].

L'IRMf permet l'acquisition de séries temporelles associées à chaque voxel, formant une image de l'"information" traitée par ce voxel au cours du temps. Les signaux enregistrés montrent des comportements complexes, très variables et il est difficile de caractériser les relations de dépendance entre voxels à l'aide des techniques classiques d'analyse statistique. En particulier [**?**, **?**, **?**] la dynamique des signaux révèle la présence d'effet de longue mémoire associé à des propriétés d'invariance d'échelle ; de plus, le réseau fonctionnel cérébral présente une organisation de type 'small-world', conservée à travers plusieurs échelles d'analyse.

L'analyse en ondelettes de signaux de moyennes spatiales sur des zones anatomiques du cerveau [**?**] permet de prendre en compte les effets de longue mémoire. L'analyse de dépendance

[1]S. Achard est partiellement financée par l'ANR 2010 JCJC 0302 01. PO Amblard est financé par une bourse Marie Curie de la communauté européenne. Ce travail est partiellement financé par le PPF ISSO de l'université de Nice Sophia-Antiolis (F. Chatelain et O. Michel)

entre deux zones anatomiques peut alors être définie à l'aide des fonctions de corrélations entre les coefficients d'ondelettes -à une même échelle- associés à chaque zone. Dans ce contexte, les séries temporelles (de coefficients d'ondelettes) forment des séries aléatoires gaussiennes. Nous proposons dans cet article une méthode d'estimation du réseau fonctionnel cérébral pour les données IRMf, combinant les avantages de l'analyse en ondelettes et de la modélisation par graphe gaussien [**?**, **?**].

La notion de dépendance conditionnelle et sa relation avec la corrélation et la corrélation partielle sont réintroduits dans la section 2. La section 3 présente l'estimateur du maximum de vraisemblance (EMV) pour les modèles graphiques gaussiens. Ces deux approches sont comparées sur des données de synthèse présentant des propriétés de longue mémoire et des relations de dépendance décrites par un modèle graphique extrait de données IRMf réelles. Dans la suite de l'article, toutes les données considérées sont extraites d'une analyse en ondelettes à une échelle donnée, et seront notées $(\mathbf{x}^1, \ldots, \mathbf{x}^n)$. La série $\mathbf{x}^t$ sera supposée i.i.d. Cette hypothèse forte et approximative se fonde sur les propriétés de décorrélation de la décomposition en ondelette lorsque l'ondelette est bien choisie [**?**].

2 Dépendance conditionnelle instantanée

Soit x_i^t, $t \in \{1, \ldots, n\}$ le signal associé au voxel d'indice i, $i \in V = \{1, \ldots, p\}$. L'existence d'un lien actif entre deux voxels est dans cette étude, subordonnée à l'existence d'une relation de dépendance entre les signaux associés à ces voxels. En pratique, ces relations de dépendance sont souvent réduites à la notion de dépendance linéaire [**?**] : l'analyse consiste à estimer la matrice de covariance du processus p-dimensionnel $\mathbf{x}^t = [x_1^t, \ldots, x_p^t]^T$. De fortes valeurs de corrélations entre les signaux des voxels sont supposées signer l'existence d'un échange d'information entre ces derniers ; cependant, il est bien connu que deux séries peuvent être corrélées, sans qu'il n'existe de lien entre les voxels. L'importance de la prise en compte de l'ensemble des autres voxels pour déterminer l'existence de liens a été discutée dans l'article fondateur [**?**] et étendue au cas non gaussien dans [**?**].

On se limite dans cet article à ne considérer que des modèles de dépendance linéaire instantanée, et des processus gaussiens. La matrice de covariance fournit donc une information exhaustive. On introduit l'espace de Hilbert engendré par l'ensemble des observations, à l'exception de x_i^t et x_j^t : $\mathcal{H}_t^{V\setminus\{i,j\}} = \operatorname{span}\{x_k^t, k \in V \setminus \{i,j\}\}$. L'estimateur optimal de x_i^t au sens des moindres carrés, sur $\mathcal{H}_t^{V\setminus\{i,j\}}$ est donc $\hat{x}_i^t = \mathcal{P}\left(x_i^t | \mathcal{H}_t^{V\setminus\{i,j\}}\right)$, où $\mathcal{P}(x|\mathcal{H})$ est la projection orthogonale de x sur $\mathcal{H}$. Le processus d'innovation associé est $\varepsilon_i^t = \hat{x}_i^t - x_i^t$. La corrélation partielle est définie sur ces processus d'innovations par

$$\begin{cases} k_{i,j} = \operatorname{cov}[\varepsilon_i^t, \varepsilon_j^t] \\ p_{i,j} = -\frac{k_{i,j}}{\sqrt{k_{i,i}k_{j,j}}} \qquad \forall\, 1 \leq i,j \leq p \end{cases} \tag{1}$$

L'équation (1) est importante [**?**] : elle permet d'exprimer les coefficients de la matrice de corrélation partielle P à partir de l'inverse de la matrice de covariance Σ : $k_{i,j} = [\Sigma^{-1}]_{i,j}$ si $i \neq j$. Deux stratégies d'estimation de P se présentent : la première consiste à inverser une estimée de Σ (estimateur empirique classique, noté $\widehat{P}_{\text{emp}}$), la seconde qui fait l'objet de cet article, conduit à estimer P directement à partir d'une représentation globale du modèle de dépendance par un graphe décomposable.

3 EMV pour les modèles graphiques décomposables

Les modèles graphiques permettent de caractériser les relations de dépendance conditionnelle entre processus aléatoires. On note $G = (V, E)$ un graphe non dirigé, où $V = \{1, ..., p\}$ décrit l'ensemble des sommets et E l'ensemble des liens $(i,j) \in V \times V$. Un modèle multivarié est qualifié de markovien sur le graphe G si pour tout lien (i,j) de E, les variables d'indice i et j sont indépendantes conditionnellement à toutes les autres [**?**].

Soit $\mathbf{x}$ un vecteur gaussien centré p-dimensionnel, de covariance Σ, et dont les composantes sont indexées par V. La matrice de précision de $\mathbf{x}$ est définie par $K = \Sigma^{-1}$. Une propriété cruciale des modèles graphiques gaussien est la suivante : imposer une propriété de Markov sur G est équivalent à imposer une contrainte de parcimonie sur $K = (k_{i,j})_{1 \leq i,j \leq p}$: $k_{i,j} = 0 \;\forall\, (i,j) \notin E$. L'évaluation de l'EMV de K peut être conduite par des méthodes numériques itératives [**?**, p. 134] ; plus récemment, une pénalisation en norme ℓ_1 a été proposée pour accentuer la parcimonie des solutions [**?**, **?**, **?**]. Cependant, lorsque la structure de dépendance est connue et correspond à un graphe décomposable (G), il est possible d'obtenir une forme analytique de l'EMV.

Un graphe décomposable est un graphe qui admet une relation d'ordre sur ses cliques $(C_1, \ldots, C_k)$: pour $i = 2, \ldots, n$, il existe $l < i$ tel que

$$S_i = C_i \bigcap \left[\bigcup_{j=1}^{i-1} C_j\right] \subseteq C_l. \tag{2}$$

Les S_i sont qualifiés de séparateurs. Sous ces hypothèses, la vraisemblance du modèle graphique gaussien associé à G peut se factoriser :

$$p(\mathbf{x}; K) = \frac{\prod_{i=1}^{k} p(\mathbf{x}_{C_i}; \overline{K}_{C_i})}{\prod_{i=2}^{k} p(\mathbf{x}_{S_i}; \overline{K}_{S_i})}, \tag{3}$$

où $\mathbf{x}_A$ représente le sous-vecteur de $\mathbf{x}$ qui contient tous les éléments indexés par $A \subset V$. $\overline{K}_A$ est l'inverse de la sous matrice de covariance Σ_A qui contient tous les termes de Σ indexés par $A \times A$. Une condition nécessaire et suffisante de décomposabilité est que le graphe ne contienne aucun cycle de longueur strictement supérieur à trois (un graphe décomposable est aussi appelé graphe triangulé).

Cette condition de décomposabilité est compatible avec les notions de hiérarchie, de modularité et de 'quasi-décomposabilité' introduite dans [**?**] pour l'étude des systèmes complexes.

En conséquence, l'expression factorisée de la vraisemblance du modèle graphique gaussien permet ainsi d'exprimer analytiquement l'EMV de K :

$$\widehat{K}_{\text{EMV}} = \sum_{i=1}^{k} \left[\Gamma_{C_i}^{-1}\right]^0 - \sum_{i=2}^{k} \left[\Gamma_{S_i}^{-1}\right]^0, \tag{4}$$

où $[M_A]^0$, pour $A \in V$, est la complétée de M_A sur V construite en complétant les entrées de M_A par des zéros, Γ_A^{-1} désignant l'inverse de la sous-matrice Γ_A de la matrice de covariance empirique. L'EMV existe presque sûrement si et seulement si le nombre d'échantillons n est au moins égal au cardinal de la clique de taille maximale dans G. Dans [**?**], les auteurs ont étendu ce résultat et proposé un estimateur non biaisé de variance minimale (MVUE) défini par

$$\begin{aligned}\widehat{K}_{\text{MVUE}} &= \textstyle\sum_{i=1}^{k} \frac{n-|C_i|-1}{n} \left[\Gamma_{C_i}^{-1}\right]^0 \\ &\quad - \frac{n-|S_i|-1}{n} \textstyle\sum_{i=2}^{k} \left[\Gamma_{S_i}^{-1}\right]^0 .\end{aligned}$$

où $|C|$ est le cardinal de l'ensemble C.

Nous nous concentrons dans cet article sur le problème d'estimation de la matrice de corrélation partielle P. Cette matrice présente l'avantage d'être invariante par rapport aux changements d'échelles que l'on peut observer dans les expériences d'IRMf. Par ailleurs, la principe d'invariance garantit que l'EMV $\widehat{P}_{\text{EMV}}$ de P peut être obtenu en substituant $\widehat{K}_{\text{EMV}}$ à K dans l'équation (1). L'estimateur noté $\widehat{P}_{\text{PMVUE}}$ obtenu en remplaçant K par $\widehat{K}_{\text{MVUE}}$ dans (1) est lui aussi introduit, bien que ce dernier puisse être biaisé.

4 Résultats, Simulations

Synthèse des données. Un ensemble de données expérimentales réelles est utilisé afin de simuler une matrice de précision parcimonieuse, aussi réaliste que possible. Les voxels ont été regroupés en $p = 90$ composantes (90 noeuds) en respectant des critères spatiaux anatomiques. La matrice de corrélation considérée est la matrice de corrélation des coefficients d'ondelette correspondant à la bande 0,02-0,04 Hz. Elle est seuillée de sorte à garantir la parcimonie de la matrice de précision (env. 15% des entrées de K sont non nulles).

Les données de synthèses sont construites de sorte à posséder cette même matrice de corrélation ('surrogate data'). Les exposants de Hurst, fixés à 0,7 dans la simulation proposée, et les densités spectrales de puissances sont eux aussi contrôlés. Les simulations ont été développées en langage R[1] et à l'aide de l'algorithme décrit dans [**?**].

Il est important de noter que la présence d'exposants de Hurst supérieurs à 0,5 a été observée sur des données réelles, ce qui conduit à des estimations fortement biaisées des corrélations, dominées essentiellement par les basses fréquences [**?**]. Cela justifie de décomposer les signaux à différentes échelles successives par une transformation en ondelettes, et de ne considérer l'estimation de K qu'à une échelle spécifique. La simulation des graphes sur d'autres bandes spectrales peut être conduite de manière analogue. Une analyse conjointe des graphes de connectivité obtenus à chacune des échelles d'analyse n'est pas envisagée dans cet article.

Triangulation. La matrice de précision synthétisée n'est en général pas associée à un graphe décomposable. Il est naturel de considérer alors le graphe décomposable le plus proche. Ce problème connu est NP-difficile. Nous utilisons donc une méthode heuristique sous optimale de triangulation [**?**], conduisant à ajouter des liens de sorte qu'aucun cycle induit de longueur supérieure ou égale à 4 ne puisse exister (condition de décomposabilité de G). Nous avons eu recours pour ces étapes au paquetage Java "JPSGCS" [**?**].

Dans notre cas, la taille de la clique maximale du graphe décomposable étant 28, l'estimateur de maximum de vraisemblance dérivé du graphe décomposable est défini avec une probabilité un pour toutes tailles d'échantillon plus grande ou égale à 28. Notons que l'estimateur $\widehat{P}_{\text{emp}}$ n'est défini que pour des tailles d'échantillons $n \leq p$, i.e. lorsque la matrice de covariance empirique Γ est définie positive.

Résultats. Un patron de parcimonie pour la matrice de corrélation partielle P, obtenu est représenté sur la figure 1. Il est bien sûr attendu que les corrélations partielles estimées sur des liens qui ont été rajoutés pour satisfaire les conditions de décomposabilité, ne prennent que des valeurs peu significatives. Les performances des estimateurs de corrélations partielles proposés ont été estimés via des simulations de Monte Carlo, à l'aide de l'erreur quadratique moyenne globale définie par $\text{MSE}(\widehat{P}) = E\left[||\widehat{P} - P||^2\right]$ où $||A||$ est la norme matricielle de Frobenius.

Il est à noter que les erreurs quadratiques moyennes des estimateurs $\widehat{P}_{\text{MLE}}$ et $\widehat{P}_{\text{PMVUE}}$ sont égales à 0 pour toutes les valeurs de la matrice $(i,j) \notin E$. Comme nous n'imposons aucune hypothèse a priori sur la structure de covariance pour l'estimateur $\widehat{P}_{\text{emp}}$, nous ne comparons dans la suite seulement les valeurs non-nulles de la matrice de précision initiale K.

Fig. 2 (a) représente le logarithme décimal de l'erreur quadratique moyenne pour les valeurs non-nulles de la matrice de précision initiale K en fonction du logarithme décimal de la taille d'échantillon. Pour des tailles d'échantillons suffisamment grandes, tous les estimateurs ont des performances équivalentes. Cependant pour de petites tailles d'échantillons, l'EMV semble être l'estimateur le plus performant.

Pour finir, fig. 2 (b) représente les erreurs quadratiques moyennes calculées seulement sur les arêtes ajoutées lors de la triangulation. Les estimateurs ont un comportement similaires. De plus, l'estimation trouvée par les estimateurs basés sur le maximum de vraisemblance sur les arêtes rajoutées est bien proche de zéro.

[1] partiellement à l'aide de la boite à outils *brainwaver*

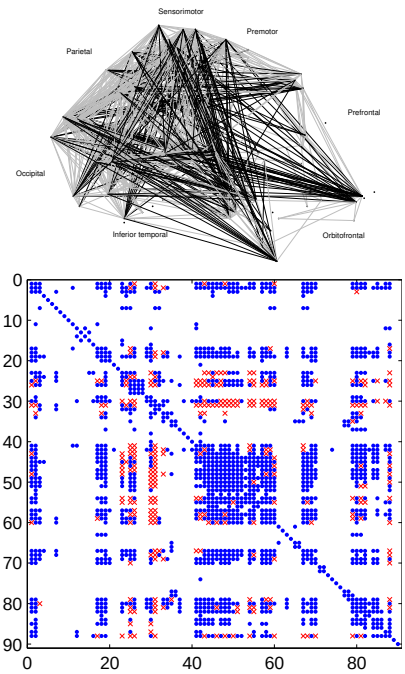

FIG. 1: Haut : représentation anatomique de 700 connexions retenues ; les connexions à courte distance, < 7.5cm, (resp. longues) sont indiquées en gris (resp. noir). Bas : Indicatrice des entrées non nulles dans la matrice de précision. En bleu : connexions physiques (avant triangulataion) ; en rouge : connexions ajoutées par l'algorithme de triangulation permettant de satisfaire la propriété de décomposabilité.

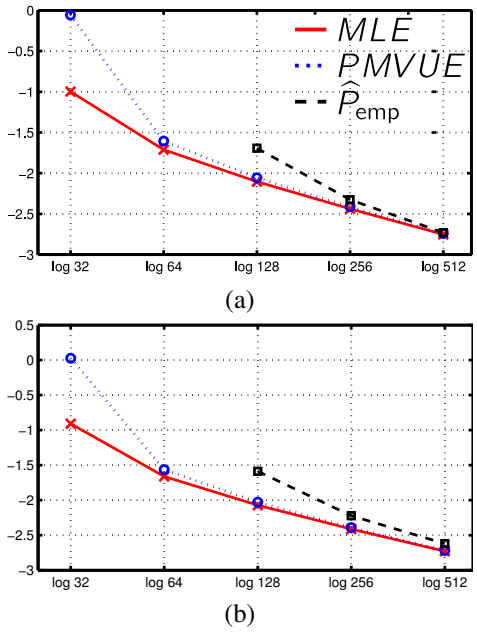

FIG. 2: Erreurs d'estimation quadratique normalisés (en bleu), en fonction de la taille d'échantillon (rouge : EMV, bleu : PMVUE, noir : estimateur $\widehat{P}_{\text{emp}}$). (a) Erreur quadratique moyenne sur la totalité du graphe décomposable. (b) Erreur quadratique moyenne sur les arêtes rajoutées lors de la triangulation.

5 Conclusions

Dans ce papier, nous avons montré l'intérêt des méthodes d'estimation basées sur des modéles graphiques dans le cadre d'étude de graphes de connectivité en IRMf. Le caractére longue mémoire des séries temporelles observées en IRMf est controlé grâce à une décomposition en ondelettes. Nous proposons dans la suite de poursuivre ces travaux dans le cadre de sélection de modèles graphiques par des méthodes lasso ou bayésiennes.

Références

[1] E. Bullmore and O. Sporns. Complex brain networks : graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3) :186–198, Mar 2009.

[2] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear. Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage*, 23(Sup 1) :234–249, 2004.

[3] V. Maxim, L. Şendur, M. J. Fadili, J. Suckling, R. Gould, R. Howard, and E. T. Bullmore. Fractional Gaussian noise, functional MRI and Alzheimer's disease. *NeuroImage*, 25 :141–158, 2005.

[4] S. Achard, D. S. Bassett, A. Meyer-Lindenberg, and E. Bullmore. Fractal connectivity of long-memory networks. *Physical Review E*, 77 :036104, 2008.

[5] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.*, 26(1) :63–72, Jan. 2006.

[6] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

[7] G. Marrelec and H. Benali. Asymptotic bayesian structure learning using graph supports for gaussian graphical models. *J. of Multivariate Analysis*, 97 :1451–1466, 2006.

[8] P. Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inform. Theory*, 38 :910–17, 1992.

[9] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37 :424–438, 1969.

[10] P.-O. Amblard and O. J. J. Michel. On directed information theory and Granger causality graphs. *J. Comput. Neurosci.*, pages 1–10, 2010.

[11] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley, 1990.

[12] N. Meinshausen and P. Buhlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3) :1436–1462, 2006.

[13] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3) :432–441, June 2008.

[14] N. Meinshausen and P. Buhlmann. Stability selection. *J. R. Stat. Soc. B.*, 72(4) :417–473, 2010.

[15] H. Simon. The architecture of complexity. *Proc. Am. Phil. Soc.*, 106 :467–482, 1962.

[16] A. Wiesel, Y.C. Eldar, and A.O. Hero. Covariance estimation in decomposable gaussian graphical models. *IEEE Trans. Signal Process.*, 58(3) :1482–1492, March 2010.

[17] M. J. Chambers. The simulation of random vector time series with given spectrum. *Math. Comput. Model.*, 22(2) :1–6, 1995.

[18] E. H. Bachoore and H. L. Bodlaender. New upper bound heuristics for treewidth. Technical Report UU-CS-2004-036, Department of Information and Computing Sciences, Utrecht University, 2004.

[19] Alun Thomas. JPSGCS : java programs for statistical genetics and computational statistics. http ://balance.med.utah.edu/wiki/index.php/JPSGCS.