

Estudo do artigo:

'A group analysis using the Multiregression Dynamic Models for fMRI networked time series'

Costa, L., Smith, J. Q. e Nichols, T.

André F. B. Menezes | Maio, 2020



Contents lists available at [ScienceDirect](#)

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi



A group analysis using the Multiregression Dynamic Models for fMRI networked time series

Lilia Costa^{a,*}, James Q. Smith^b, Thomas Nichols^c

^a Universidade Federal da Bahia, Brazil

^b The University of Warwick, UK

^c Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, UK

ARTICLE INFO

Article history:

Received 26 April 2016

Received in revised form 7 March 2018

Accepted 17 March 2018

Available online 6 April 2018

Keywords:

Multiregression Dynamic Model

Bayesian network

Group analysis

Cluster analysis

Functional Magnetic Resonance Imaging (fMRI)

ABSTRACT

Connectivity studies of the brain are usually based on functional Magnetic Resonance Imaging (fMRI) experiments involving many subjects. These studies need to take into account not only the interaction between areas of a single brain but also the differences amongst those subjects. In this paper we develop a methodology called the *group-structure* (GS) approach that models possible heterogeneity between subjects and searches for distinct homogeneous sub-groups according to some measure that reflects the connectivity maps. We suggest a GS method that uses a novel distance based on a model selection measure, the Bayes factor. We then develop a new class of Multiregression Dynamic Models to estimate individual networks whilst acknowledging a GS type dependence structure across subjects. We compare the efficacy of this methodology to three other methods, virtual-typical-subject (VTS), individual-structure (IS) and common-structure (CS), used to infer a group network using both synthetic and real fMRI data. We find that the GS approach provides results that are both more consistent with the data and more flexible in their interpretative power than its competitors. In addition, we present two methods, the Individual Estimation of Multiple Networks (IEMN) and the Marginal Estimation of Multiple Networks (MEMN), generated from the GS approach and used to estimate all types of networks informed by an experiment –individual, homogeneous subgroups and group networks. These methods are then compared both from a theoretical perspective and in practice using real fMRI data.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Organização

- Contextualização
- Abordagens para análise de grupo
- *Multiregression dynamic model* e análise de grupo
- Estudo de simulação
- Aplicação
- Considerações finais

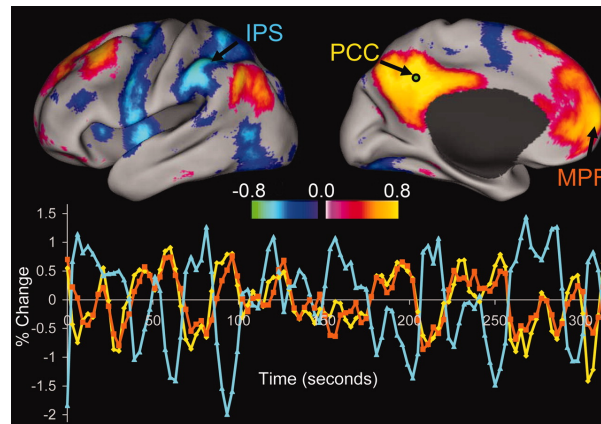
Contextualização

Resting-State

- **O que é?**
 - Área da neurociência que estuda o cérebro quando indivíduo está em estado de repouso.
- **Qual a importância?**
 - Compreender a conectividade do cérebro. Existem dois tipos:
 - *functional connectivity*: associação estatística entre as conexões.
 - *effective connectivity*: relação causal entre as conexões.

Resting-State

- Como fazer isso?
 - *function Magnetic Resonance Imaging* (fMRI).
 - Para cada voxel do cérebro obtém-se uma imagem que reflete o nível de oxigenação do sangue.
 - Esta relacionado à ativação dos neurônios cerebrais.

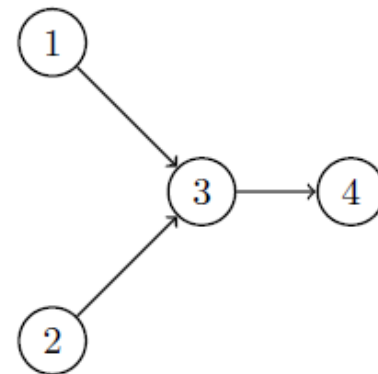
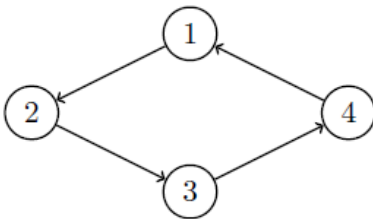


Contextualização

- **Objetivo do artigo:** Estudar a *effective connectivity* utilizando modelos multivariados dinâmicos e considerando a presença de diferentes indivíduos.
- **Notação:** $Y_{it}(r)$ denota o nível de oxigénio para a região r do indivíduo i no tempo t .

Directed acyclic graphs (DAG)

- A conectividade do cérebro pode ser estudada utilizando diferentes classes de modelos de grafos.
- Nós (*nodes*): representam cada voxel (parte) do cérebro.
- Arestas (*edges*): representam a conectividade entre os pares de voxels.
 - *Undirected*: reflete alguma dependência. **Sem direção!**
 - *Directed*: reflete uma relação de causa-efeito.
- DAG: classe de modelos de grafos acíclicos, i.e., nenhum caminho começa e termina no mesmo nó.



Abordagens para análise de grupo

Abordagens para análise de grupo

- *virtual-typical-subject* (VTS): ignora a variabilidade do indivíduo.
 - Trabalha com o valor médio por região do cérebro e tempo.
 - Concatena as observações supondo que as informações são de um único indivíduo.
- *common-structure* (CS): considera a mesma estrutura de rede (conexões), porém os parâmetros diferem entre indivíduos.
- *individual-structure* (IS): estima uma rede (conexão) para cada indivíduo.
 - A rede para o grupo de indivíduos é formada pelas conexões estimadas para o maior número de indivíduos.
- *group-structure* (GS): estuda a homogeneidade do grupo via análise de agrupamento, considerando alguma medida de similaridade entre os indivíduos.
 - Análises subsequentes devem considerar a heterogeneidade, se houver.

Multiregression dynamic model e análise de grupo

0 modelo

- **Para um único indivíduo.** Seja $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(n))$ as medidas de oxigenação em n regiões do cérebro no tempo t .
- $\text{Pa}(r)$ conjunto de variáveis pais de $Y_t(r)$.
- Equações de observação:

$$Y_t(r) \sim N(\mathbf{F}_t(r)^\top \boldsymbol{\theta}_t(r), V(r)), \quad r = 1, \dots, n.$$

- Equações de sistema:

$$\boldsymbol{\theta}_t \sim N(\boldsymbol{\theta}_{t-1}, \mathbf{W}_t)$$

- Em que
 - $\mathbf{F}_t(r)$ é um vetor contendo os pais $\text{Pa}(r)$ da região r .
 - $\mathbf{W}_t = \text{blockdiag}\{\mathbf{W}_t(1), \dots, \mathbf{W}_t(n)\}$.

0 modelo

- Seja $\phi(r) = V(r)^{-1}$ a precisão observacional da r -ésima região.
- Informação inicial:

$$\begin{aligned}(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) &\sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}) \\(\phi(r) \mid D_{t-1}) &\sim \text{Gamma}\left(\frac{n_0(r)}{2}, \frac{d_0(r)}{2}\right)\end{aligned}$$

- $\mathbf{C}_{t-1} = \text{blockdiag}\{\mathbf{C}_{t-1}(1), \dots, \mathbf{C}_{t-1}(n)\}$
- $\mathbf{W}_t(r) = \frac{1 - \delta(r)}{\delta(r)} \mathbf{C}_{t-1}(r)$, em que $\delta(r) \in (0, 1]$ é o fator de desconto para $r = 1, \dots, n$.
- $\delta(r)$ pode ser obtido maximizando a verossimilhança marginal do modelo.

Seleção de modelo

- Como escolher o melhor conjunto $\text{Pa}\{Y_t(r)\}$ para cada $Y_t(r)$ respeitando a condição do DAG?
- Log-predictive likelihood (LPL) do modelo m :

$$\text{LPL}(m) = \sum_{r=1}^n \sum_{t=1}^T \log p(y_t(r) \mid D_{t-1}, \text{Pa}_m(r))$$

- $\text{LPL}(m)$ é a soma de n escores locais, um para cada nó r .
- Escore local para $Y_t(r)$ é determinado pela escolha do conjunto $\text{Pa}_m(r)$ especificada pelo modelo m .
- Seja $c(r, \text{Pa}_m(r))$ o escore local do modelo m , de tal forma que

$$\text{LPL}(m) = \sum_{r=1}^n c(r, \text{Pa}_m(r)).$$

Seleção de modelo

- Costa et al. (2015) propõe o método MDM-IP que consiste em:
 - Buscar n subconjuntos $Pa(1), \dots, Pa(n)$ que **maximizem** o $LPL(m)$ **sujeito** a restrição que todo o sistema deve ser acíclico e direcional (DAG).
 - Os autores utilizam programação linear inteira para encontrar melhor conjunto $Pa_m(r)$.
- Costa et al. (2017) propõe outro método que busca pelo melhor subconjunto sem a restrição do DAG.
- Neste artigo o método MDM-IP foi empregado, conforme a abordagem de grupo, para selecionar o "melhor" modelo.

VTs

- Assume mesma rede de conexões para todos os indivíduos.
- O escore local é dado por

$$c_a(r, \bar{M}(r)) = \sum_{t=1}^T \log p [\bar{y}_t(r) \mid \bar{\mathbf{x}}_t(r), \bar{M}(r)] ,$$

em que:

- $\bar{y}_t(r)$ é a média no tempo t e nó r .
- $\bar{\mathbf{x}}_t(r) = (\bar{y}_t(1), \dots, \bar{y}_t(r-1))$
- $\bar{M}(r)$ é o modelo definido pelo conjunto de pais do nó r
- $\bar{M} = (\bar{M}(1), \dots, \bar{M}(n))$ é a rede de conexões para o grupo de indivíduos.
- A força da conectividade para o grupo de indivíduos é baseada na distribuição posteriori suavizada de $\boldsymbol{\theta}_t$.

CS

- Assume a mesma estrutura de rede, porém os parâmetros diferem conforme indivíduo.
- Dado um modelo (grafo) $\bar{\mathbf{M}}(r)$, o escore local é dado por

$$c(r, \bar{\mathbf{M}}(r)) = \sum_{i=1}^S \sum_{t=1}^T \log p [y_{it}(r) \mid \mathbf{x}_{it}(r), \bar{\mathbf{M}}(r)] ,$$

em que:

- S é o número de indivíduos.
- $y_{it}(r)$ é o valor observado para a região r e indivíduo i no tempo t .
- $\mathbf{x}_{it}(r) = (y_{it}(1), \dots, y_{it}(r-1))$

IS

- Assume diferentes redes para cada indivíduo.
- O escore local para o i -ésimo indivíduo é dado por

$$c_i(r, \mathbf{M}_i(r)) = \sum_{t=1}^T \log p[y_{it}(r) \mid \mathbf{x}_{it}(r), \mathbf{M}_i(r)] ,$$

em que:

- $\mathbf{M}_i(r)$ é o modelo definido pelo conjunto de variáveis pais do nó r para o sujeito i .
- $\mathbf{M}_i = (\mathbf{M}_i(1), \dots, \mathbf{M}_i(n))$ é a rede para o individuo i .
- A estrutura de rede do grupo ($\bar{\mathbf{M}}$) é formada pelas conexões obtidas para o maior número de indivíduos.

GS ou *pairwise log Bayes factor separation*

- Indivíduos são agrupados conforme uma medida de dissimilaridade de seus grafos.
- Seja M_i a rede estimada para o i -ésimo indivíduo.
- Seja m_G a rede de grupo dos pares de indivíduos i e j .
- A medida de dissimilaridade, $d(i, j)$, é calculada para todos os pares de indivíduos i e j .

$$d(i, j) = c_{ij}(m_I) - c_{ij}(m_G)$$

em que $m_I = (M_i, M_j)$,

$$c_{ij}(m_I) = \sum_{r=1}^n [c_i(r, M_i(r)) + c_j(r, M_j(r))]$$
$$c_{ij}(m_G) = \sum_{r=1}^n [c_i(r, m_G(r)) + c_j(r, m_G(r))]$$

para $i < j$ com $i \in \{1, \dots, S-1\}$ e $j \in \{2, \dots, S\}$.

Considerações sobre método GS

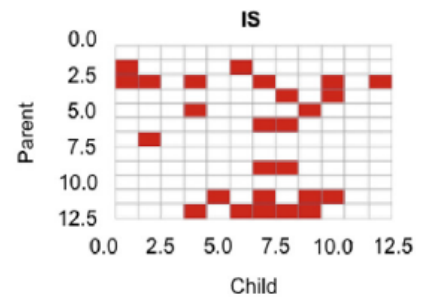
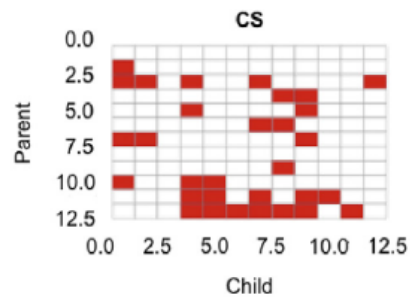
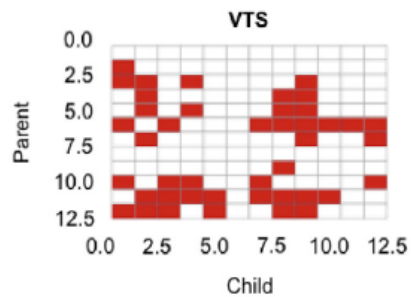
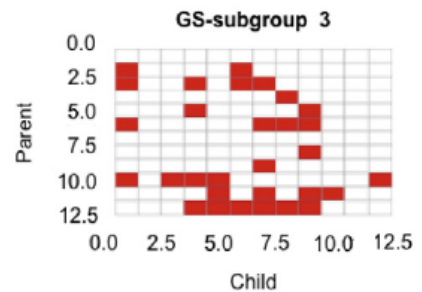
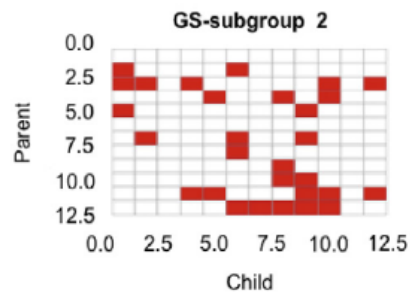
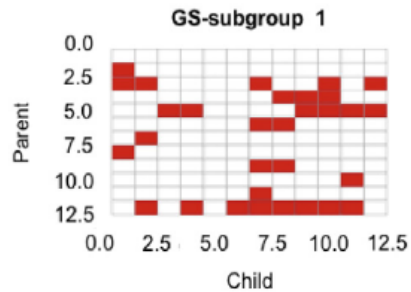
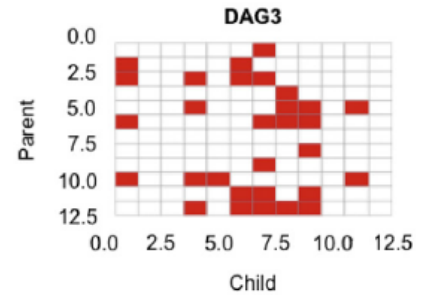
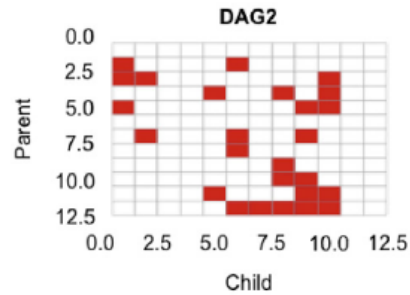
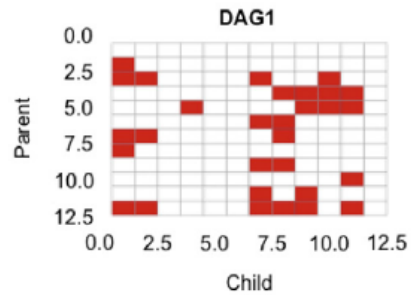
- Em relação as estruturas de rede dos modelos:
 - M_i é estimada maximizando os escores, i.e., conforme a abordagem IS.
 - m_G é estimada maximizando os escores dos indivíduos i e j , conforme a abordagem CS considerando $\bar{M}(r) = m_G(r)$.
- Se o algoritmo de busca (seleção de modelo) é o método MDM-IP os escores são exatamente a LPL, logo a medida $d(i, j)$ pode ser vista como sendo o fator de Bayes.
- Utilizando métodos de agrupamento e a medida de dissimilaridade definida, os indivíduos são agrupados conforme suas respectivas estruturas de redes (conexões).
- A abordagem CS é utilizada conforme os grupos estabelecidos para obter um escore local.
- Finalmente, a estrutura das conexões cerebrais são estimadas independentemente para cada subgrupo M_1, \dots, M_G .

Estudo de simulação

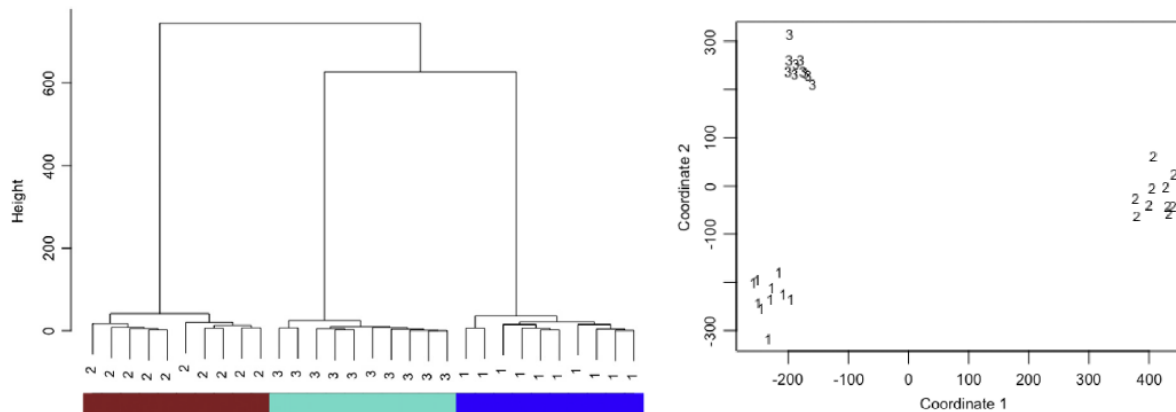
Estudo de simulação

- **Objetivo:** avaliar a eficácia dos métodos quando os indivíduos exibem diferentes estruturas de conexões (rede).
- **Logística:**
 - **3 DAGs** foram utilizados.
 - Sendo **10 indivíduos** para cada.
 - Considerando **12 nós**.
 - Resultando em **1158 observações** no tempo.
- **Abordagem GS:**
 - MDM-IP foi utilizado para encontrar "melhor" modelo.
 - Método de agrupamento hierarquico e escalonamento dimensional foram utilizados para avaliar a homogeneidade dos grupos.

Resultados



Resultados



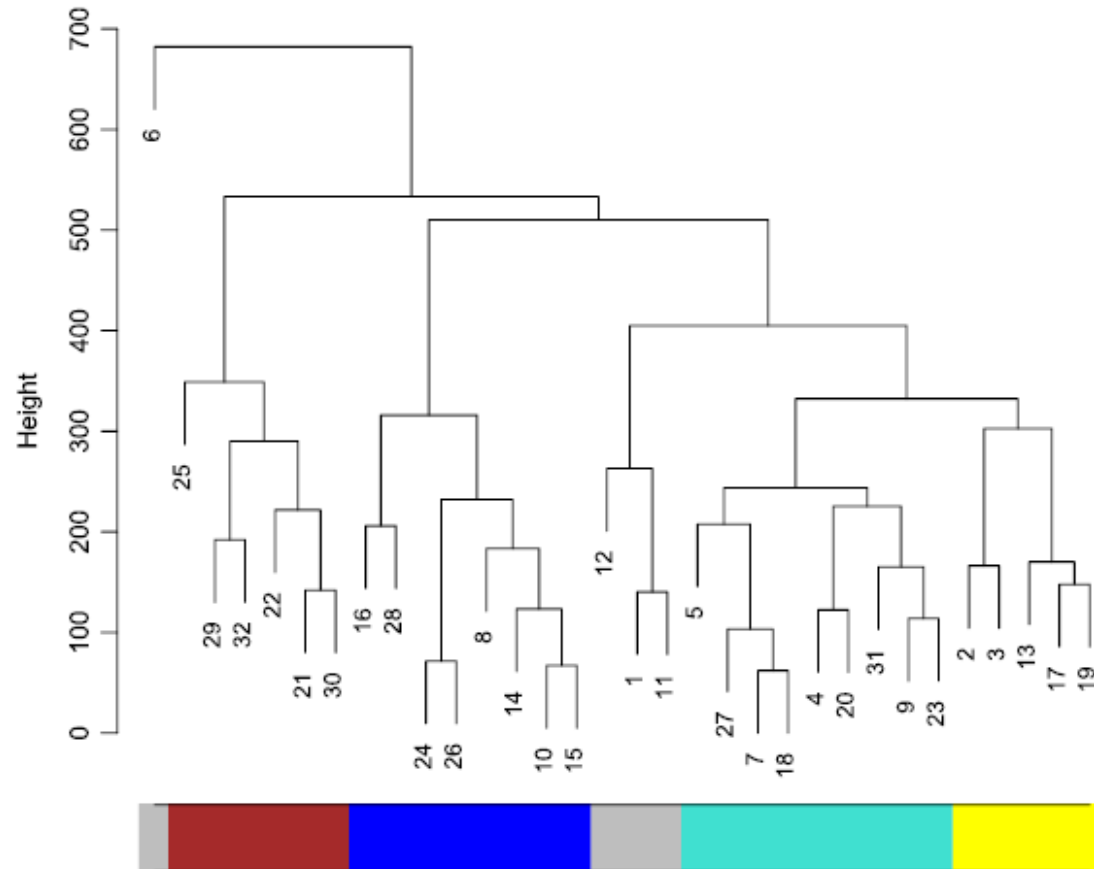
- Medidas de sensibilidade e especificidade também foram calculadas para cada método.
- Tais medidas são baseadas nas conexões reais e estimadas.

Aplicação

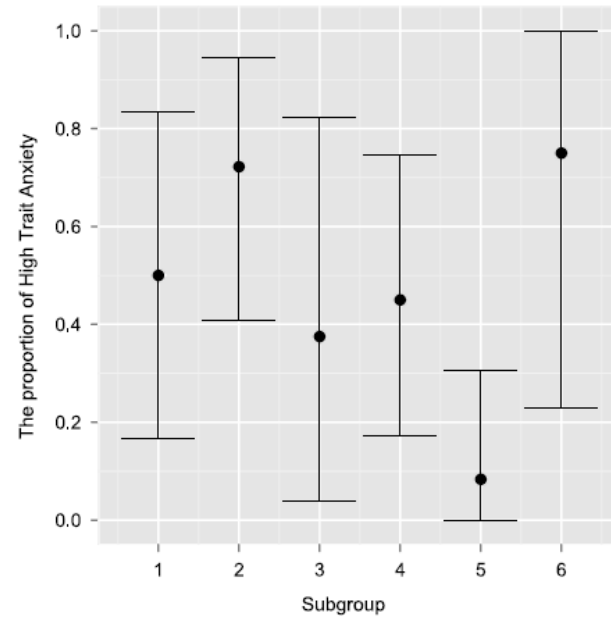
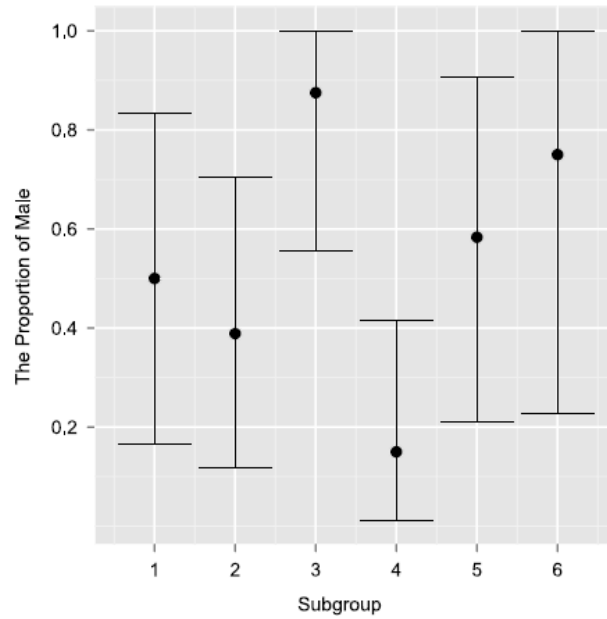
Os dados

- Exerimento *resting-state* fMRI conduzido em 32 indivíduos.
- O nível de oxigenação foi mensurado em 12 voxel (partes) do cérebro.
- O experimento demorou 15 min para cada pessoa.
- Também esta disponível informações sobre sexo e ansiedade para cada indivíduo.
- **Objetivo:** estimar as conectividades cerebrais de grupos de indivíduos.
- A estrutura de grafos foi estimada independentemente para cada grupo somando os escore dos indivíduos do mesmo subgrupo e utilizando o método MDM-IP.

Dendrograma



Características dos indivíduos



Conectividade dos subgrupos

- A seguinte medida foi utilizada para estimar as conectividades em cada subgrupo:

$$\bar{Z}_{ig} = \frac{\bar{\theta}_{ig}}{\sqrt{\frac{\bar{\sigma}_{ig}^2}{TS_g}}}$$

- em que:
 - $\bar{\theta}_{ig}$ e $\bar{\sigma}_{ig}^2$ são a média e variância da distribuição suavizada ao longo do tempo e para todos indivíduos do subgrupo g e conectividade i .
 - T é o tamanho amostral
 - S_g número de indivíduos no subgrupo g .

Conectividade dos subgrupos

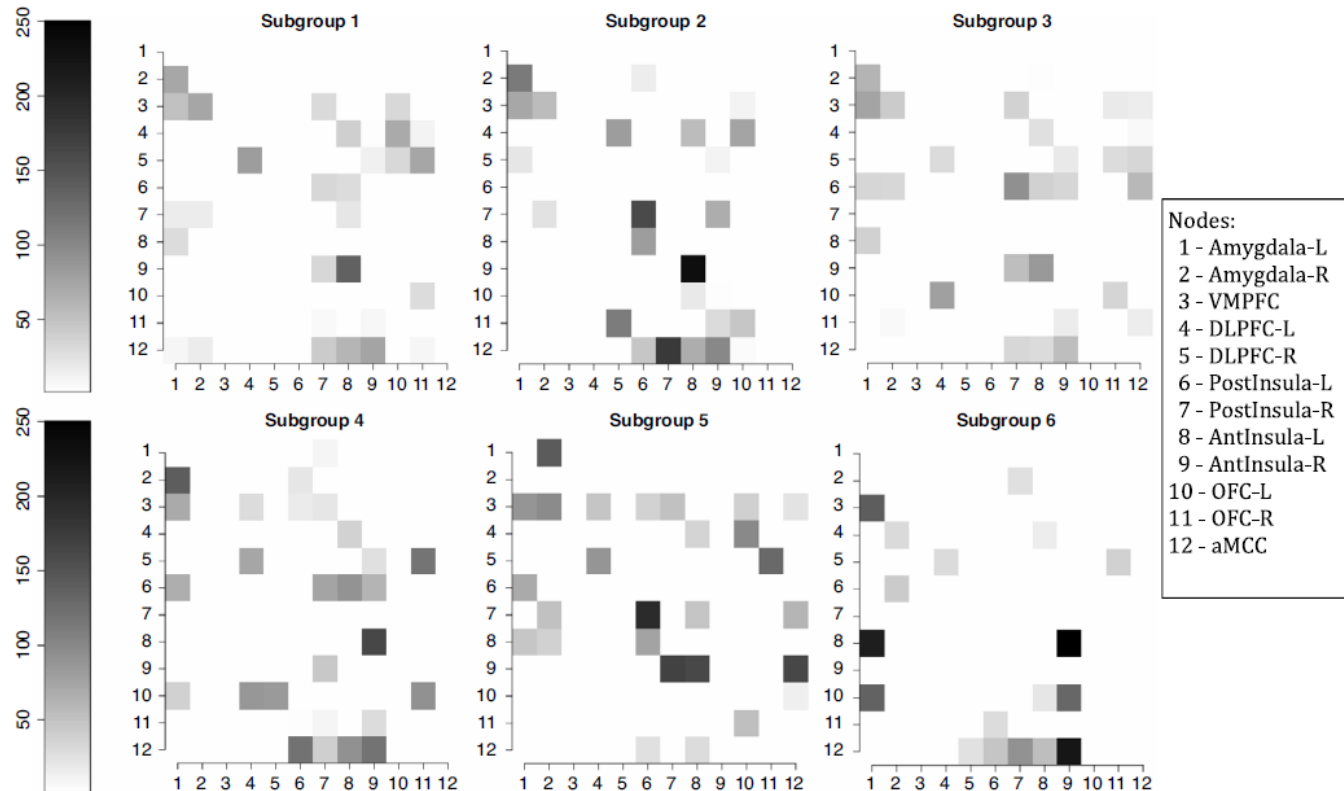


Fig. 5. The connectivity strength standardised mean (\bar{Z}_{ig}) for a particular edge $i \rightarrow j$, where i indexes rows and j columns, using the MDM-IPA per subgroup defined in Fig. 3, only for significant connectivities, i.e. $\bar{Z}_{ig} \geq 2$.

- Note que a existência e força da conectividade muda conforme o grupo.

Comparação de subgrupos

- A seguinte medida foi utilizada para estimar as diferenças de conectividades nos subgrupos g e l .

$$\bar{D}_{igl} = \frac{\bar{\theta}_{ig} - \bar{\theta}_{il}}{\sqrt{\frac{\bar{\sigma}_{ig}^2}{TS_g} + \frac{\bar{\sigma}_{il}^2}{TS_l}}}$$

- Duas comparações foram realizadas:
 - Subgrupo 3 e 4, em que proporção de homens é maior no primeiro.
 - Subgrupo 2 e 5, em que proporção indivíduos com alta ansiedade é maior no primeiro.

Diferenças entre subgrupos

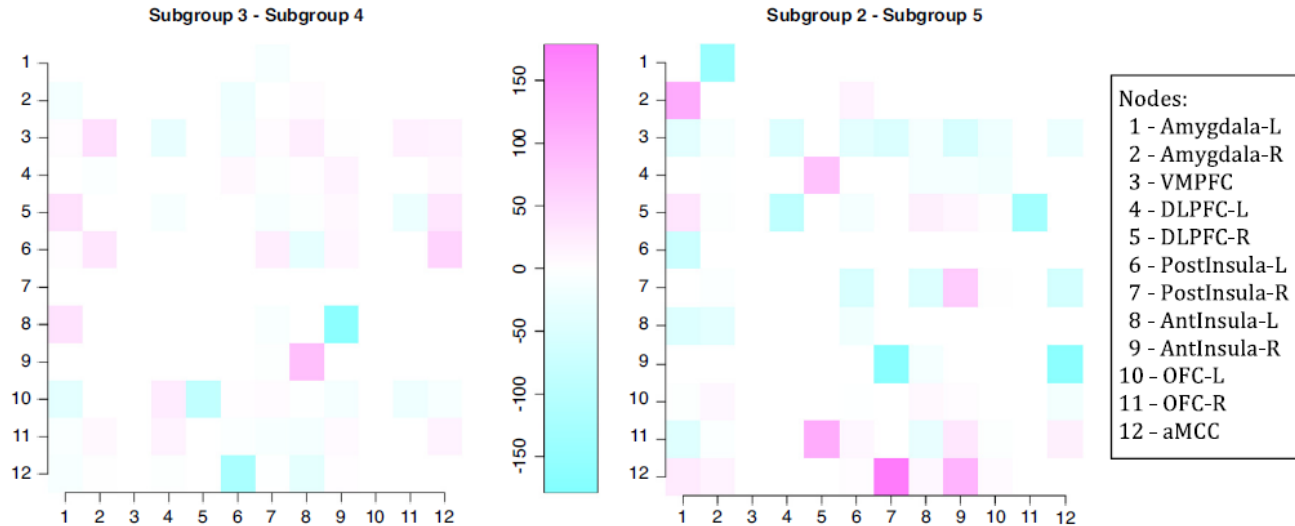


Fig. 6. The connectivity strength standardised difference ($\bar{D}_{t|g}$) for a particular edge $i \rightarrow j$, where i indexes rows and j columns, between subgroup 3 and subgroup 4 (*left*), and between subgroup 2 and subgroup 5 (*right*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Considerações Finais

- Este artigo foi parte da tese desenvolvida e defendida em 2014 por Lilia Costa:
 - **Studying Effective Brain Connectivity Using Multiregression Dynamic Models**
- Duas abordagens foram propostas para inferir de forma unificada a estrutura de grafos individual, subgrupos e geral. Especificamente:
 - *Individual Estimation of Multiple Networks* (IEMN).
 - *Marginal Estimation of Multiple Networks* (MEMN).

Referências

- COSTA, L. **Studying Effective Brain Connectivity Using Multiregression Dynamic Models**. Tese (Doutorado) — University of Warwick, 2014.
- COSTA, L. et al. **Searching multiregression dynamic models of resting-state fMRI networks using integer programming**. Bayesian Analysis, v. 10, n. 2, p. 441–478, 2015.
- COSTA, L.; NICHOLS, T.; SMITH, J. Q. **Studying the effective brain connectivity using multiregression dynamic models**. Brazilian Journal of Probability and Statistics, v. 31, n. 4, p. 765–800, 2017.
- COSTA, L.; SMITH, J. Q.; NICHOLS, T. **A group analysis using the multiregression dynamic models for fMRI networked time series**. Journal of Statistical Planning and Inference, v. 198, p. 43–61, 2019.

Obrigado!