



HAL
open science

Analysis of Brain Functional Connectivity by Frequent Pattern Mining in Graphs. Application to the Characterization of Murine Models

Aurelie Leborgne, Florence Le Ber, Laetitia Degiorgis, Laura Harsan, Stella Marc-Zwecker, Vincent Noblet

► To cite this version:

Aurelie Leborgne, Florence Le Ber, Laetitia Degiorgis, Laura Harsan, Stella Marc-Zwecker, et al.. Analysis of Brain Functional Connectivity by Frequent Pattern Mining in Graphs. Application to the Characterization of Murine Models. ISBI-Virtual, 2021, En ligne, France. hal-03238122

HAL Id: hal-03238122

<https://hal.science/hal-03238122>

Submitted on 21 Oct 2021

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.

ANALYSIS OF BRAIN FUNCTIONAL CONNECTIVITY BY FREQUENT PATTERN MINING IN GRAPHS. APPLICATION TO THE CHARACTERIZATION OF MURINE MODELS

*Aurélie Leborgne, Florence Le Ber, Laetitia Degiorgis,
Laura Harsan, Stella Marc-Zwecker, Vincent Noblet*

ICube, Université de Strasbourg, CNRS, ENGEES
{aurelie.leborgne,florence.le-ber,laura.harsan,stellamarc,zwecker,vincent.noblet}@unistra.fr,
laetitia.degiorgis@icube.unistra.fr
<http://icube.unistra.fr>

ABSTRACT

Functional Magnetic Resonance Imaging (fMRI) is an imaging technique that allows to explore brain function *in vivo*. Many methods dedicated to analyzing these data are based on graph modeling, each node corresponding to a brain region and the edges representing their functional link. The objective of this work is to investigate the interest of methods for extracting frequent pattern in graphs to compare these data between two populations. Results are presented in the context of the characterization of a mouse model of Alzheimer’s disease in comparison with a group of control mice.

Index Terms— Graph, frequent pattern mining, fMRI, brain function

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is widely used in the field of neuroscience to investigate brain function. Resting fMRI (rsfMRI) is more particularly used to highlight and characterize cerebral networks, *i.e.*, sets of regions, which activate synchronously over time. A large number of methods for analyzing these data rely on graph modelling, each node corresponding to a brain region and edges representing the correlation between the two regions. In order to compare this data across groups of subjects, two strategies are generally used, either by performing univariate or multivariate statistical tests on the correlation coefficients between the different regions, either by characterizing the global or local topology of these graphs, via parameters such as degree, hubness, clustering coefficient, average path length, that are extracted for each individual and then compared at group level using statistical tests [1].

The aim of this work is to show the potential of graph frequent pattern mining methods for comparing rsfMRI sequences between two populations. It focuses on the characterization of a murine model of Alzheimer’s disease in comparison with a group of control mice.

Searching for structures in graphs is a core problem with many applications, such as the characterization of chemical structures or the study of social networks. Different approaches have been proposed to extract frequent subgraphs, either from a single large graph, or from a set of graphs. One of the first generic algorithms, AGM [2] is based on an extension of Apriori frequent pattern mining algorithm [3]. It allows to find frequent subgraphs in a graph database. On the same basis, algorithms FSG [4] and gSpan [5] were then designed. We use the latter in this work. More recently, [6] have proposed an approach for frequent pattern mining in a single large graph.

In the following, we first describe the data and the preprocessing pipeline (Sect. 2). Then, we present the proposed approach for extracting and comparing common subgraphs between two groups (Sect. 3). Results are detailed and discussed in Sect. 4, while Section 5 gives some conclusions and perspectives of this work.

2. DATA DESCRIPTION AND PREPROCESSING

The study focuses on a mouse model of Alzheimer’s disease (Thy-Tau22) [7]. A group of $n_{\text{Alz}} = 16$ Alzheimer’s mice is compared to a group of $n_{\text{Ctrl}} = 13$ control mice. rsfMRI were acquired on a 7T MR scanner (BioSpec 70/30, Bruker, Germany) using a GE-EPI sequence with the following parameters: 500 volumes, TE: 15ms, TR: 2000ms, 27 axial slices, slice thickness: 0.4mm, in-plane spatial resolution: 0.14x0.22mm. Data were preprocessed with SPM8¹ and SPMmouse² libraries according to the following pipeline: (i) motion correction, (ii) affine registration on the Allen Brain atlas³, (iii) Gaussian spatial filtering (FWHM: 2mm), (iv) temporal bandpass filtering (0.01-0.1Hz) and (v) regressing out the signal of the cerebrospinal fluid (see [7] for details).

Twenty-two regions known to be involved in Alzheimer’s

¹<http://www.fil.ion.ucl.ac.uk/spm/>

²www.spmmouse.org

³<http://mouse.brain-map.org/static/atlas>

disease were considered. The partial correlation coefficients between the average time course of each pair of regions were calculated to build a correlation matrix for each mouse (see Fig. 1.a). These matrices were then binarized by retaining only the correlations greater than a given threshold θ (Fig. 1.b). To reduce the influence of choosing one arbitrary threshold, a range of threshold values was considered.

Each binarized matrix is then modeled by a graph $\mathcal{G}_x^i = (\mathcal{S}_x^i, \mathcal{A}_x^i)$, representative of the i^{th} mouse ($i \in \{1, \dots, n_x\}$) of group x ($x \in \{\text{Ctrl}, \text{Alz}\}$) (see Fig. 1.c). The set of vertices $\mathcal{S}_x^i \subseteq \mathcal{S}$ represents the brain areas and the set of edges \mathcal{A}_x^i represent the links between these areas. Note that each vertex of \mathcal{S} occurs at most only once in each graph \mathcal{G}_x^i . By this way, each mouse population is represented by a set of graphs, denoted D_{Alz} and D_{Ctrl} for Alzheimer and control group, respectively.

3. EXTRACTION AND COMPARISON OF FREQUENT PATTERNS

The proposed approach is composed of two steps: (i) the extraction of frequent patterns in all subjects regardless to their belonging group and then (ii) the statistical comparison of the frequency of these patterns within each group. The aim is to highlight brain networks that are altered (*i.e.*, less frequent or even absent) in one of the two groups.

3.1. Subgraph pattern extraction

Let D be a set of graphs, and \mathcal{M} a subgraph that is present in one or more graphs of D . The set of graphs of D , that contain the subgraph \mathcal{M} is referred as the support of \mathcal{M} and is denoted by $\text{support}_{\mathcal{M}}$. The frequency $f_{\mathcal{M}}$ of \mathcal{M} in D is defined as the ratio of \mathcal{M} support cardinality and the total number of graphs, *i.e.*, $f_{\mathcal{M}} = \frac{|\text{support}_{\mathcal{M}}|}{|D|}$. A subgraph \mathcal{M} is classified as a frequent pattern if its frequency $f_{\mathcal{M}}$ is greater than a given threshold σ [4].

To perform this task, we used an implementation⁴ of gSpan algorithm [5]. It is based on an in-depth search for a hierarchical space to eliminate candidate subgraphs. Technically, each graph is associated to a code defined as a sequence, so that isomorphic graphs (*i.e.*, graphs for which there is a bijection between the vertices preserving the edges) share the same code. By this way, the problem of frequent patterns mining in a set of graphs boils down to a mining problem of a set of sequences, which can be solved far more efficiently.

Practically, we applied gSpan algorithm to the set $D_{\text{Ctrl}} \cup D_{\text{Alz}}$ with a given frequency threshold σ , in order to extract subgraphs that are common to at least $(n_{\text{Ctrl}} + n_{\text{Alz}}) * \sigma$ subjects, regardless to the group they belong.

3.2. Subgraph frequency analysis

The second step of the proposed framework consists in analyzing the extracted subgraphs while now considering the belonging group of each subject. The goal is to identify the subgraphs that may be statistically less frequent in one of the two groups. To this end, the frequency of each pattern \mathcal{M} in group x is computed as follows:

$$f_{\mathcal{M}}^x = |\text{support}_{\mathcal{M}} \cap D_x| / |D_x|$$

Then, a bilateral z-test for comparing proportions between two independent samples is carried out:

$$Z = \frac{f_{\mathcal{M}}^{\text{Alz}} - f_{\mathcal{M}}^{\text{Ctrl}}}{\sqrt{\frac{f_{\mathcal{M}}^{\text{Alz}}(1-f_{\mathcal{M}}^{\text{Alz}})}{n_{\text{Alz}}} + \frac{f_{\mathcal{M}}^{\text{Ctrl}}(1-f_{\mathcal{M}}^{\text{Ctrl}})}{n_{\text{Ctrl}}}}} \sim \mathcal{N}(0, 1)$$

Thus, a pattern is labeled as statistically different at the significance level $\alpha = 0,05$ if $|Z| > 1.96$. In order to associate to each detected graph a degree of confidence, the p-value of the statistical test (*i.e.*, probability under hypothesis \mathcal{H}_0 that a pattern is detected by chance) is also calculated.

4. EXPERIMENTS

4.1. Results

To reduce the influence of the choice of the threshold θ on the correlations, experiments were conducted for correlation thresholds varying from 0.3 to 0.6 with 0.1 increment. The threshold for the support has been set to $\sigma = 0.25$. The patterns detected and labeled as different between the two populations are summarized in Tab. 1. At the threshold $\theta = 0.4$, two subgraphs are found to be statistically different between the two group: g_1 (with frequencies 0.375 for Alz group vs. 0.769 for Ctrl group) and g_2 (with frequencies 0.625 vs. 1). At the threshold $\theta = 0.3$, only the graph g_3 is found to be statistically different. Interestingly, g_3 includes g_2 , but in contrast to g_2 which was more frequent in the Ctrl group (13/13) as compared to the Alz group (10/16), g_3 appears to be quasi-absent in the Ctrl group (2/13) while the frequency remains almost the same in the Alz group (9/16).

To allow a more detailed interpretation of the results and to understand the seemingly contradictory behavior between g_2 and g_3 , it is necessary to analyze the frequencies of the edges constituting the graphs as well as the associated p-values (see the right side of Tab. 1). If we examine the edges of graph g_1 , we can notice that none of them have a statistically different frequency between the two populations. This is of great interest since the commonly used approach that relies on univariate statistical test conducted on each edge independently would fail to detect this network. In fact, this network appears to be remarkably stable in the Ctrl group while exhibiting much more variability in the Alzheimer group: most of the mice of Alz group have one or two altered connections

⁴<https://github.com/betterenvi/gSpan>

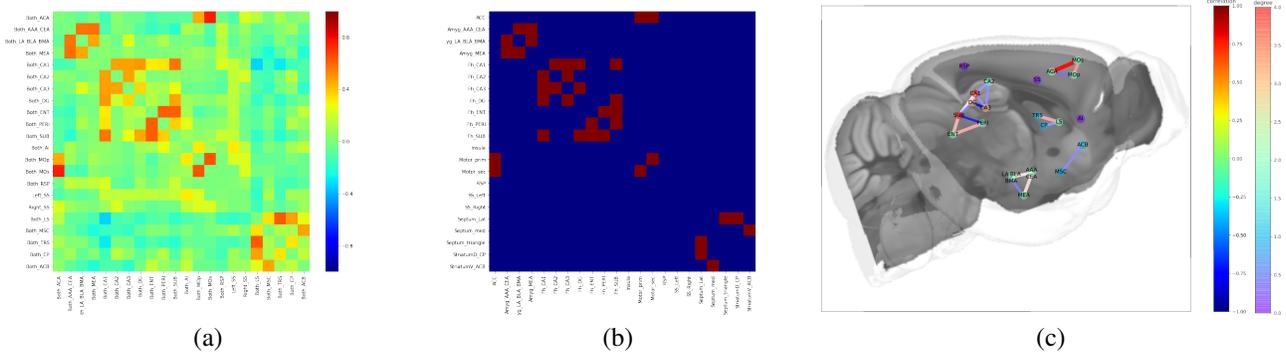


Fig. 1. Representations of the functional relationships between brain areas for a mouse: (a) correlation matrix ($c \in [-1; 1]$); (b) binarized correlation matrix (thresholding : $c \geq \theta = 0,35$); (c) corresponding graph representation.

but not always the same. For graph g_2 , only a single edge has a significantly lower frequency in the Alz group, thus explaining the lower frequency of the whole graph. For graph g_3 that includes graph g_2 , the edge Both_ACA - Both_MOp is no longer statistically different (decreasing the threshold on the correlation makes the frequency of this edge increase in Alz group), while the edge Both_MOp - Right_SS (which is not included in g_2) is, this time, less frequent in Ctrl mice. This emphasizes the main added value of the proposed framework as compared to the standard approach that focuses only on edges analysis: the proposed approach does not only highlight deficient connection, but rather altered networks, with sometimes only one or even none statistically different edge between the groups. Reporting alteration at both network and connection levels is really of great interest for neuroscientists to help them in the interpretation of the observed functional connectivity disorders.

4.2. Relevance of identified networks

Despite the fact that the analysis was conducted at the early beginning of the disease (prodromal stage) in the Thy-Tau22 mice - meaning that only subtle changes were expected - results show a significant decrease in the frequency of appearance of the network connecting the primary motor (MOp), secondary motor (MOs) and anterior cingulate cortex (ACA) (graph g_2). An alteration of the functional connectivity of the ACA, that plays an important role in particular in the control of motor decisions, has already been shown in patients with mild cognitive impairment (MCI) [8]. A second network (graph g_1), that appears significantly less frequently in Alz mice than Ctrl mice, is of particular interest since it gather key regions involved in memory, namely the hippocampus (CA1, CA2, CA3, DG), the entorhinal (ENT) and perirhinal (PERI) cortex, and the subiculum (SUB). This is consistent with microstructural alterations that were detected in the main connection pathway of these regions for the same mice, as well as a significantly higher concentration of pathologi-

cal proteins in the hippocampus [7]. These alterations could therefore reflect an overall dysfunction of the memory network, that may be responsible of potential memory disturbances at more advanced stages of the pathology. Contrary to the two previous networks, the third network (graph g_3) involving the somatomotor network is more frequent in Alz group. This suggests a hypersynchrony of this network which may be linked to a potential compensation mechanism in Alz mice, consistent with previous analyzes on the same animals revealing a functional hyperconnectivity of the somatosensory cortex (SS) [7]. A study conducted on MCI patients using magnetoencephalography also showed an increase in the amplitude of response of the SS in these patients [9].

5. CONCLUSION AND FUTURE WORKS

In this article, we have shown the interest of graph frequent pattern mining methods for the comparison of resting fMRI sequences between two populations. Unlike conventionally approaches which are based on univariate statistical tests conducted independently on each edge, the proposed method does not only detect deficient links between regions in one group as compared to the other, but also larger networks of altered regions. The method was even able to highlight a key network involved in Alzheimer's pathology in which no edge was statistically altered. The proposed approach has thus demonstrated a greater sensitivity to detect functional anomalies between populations. In addition, it offers a more global view of altered networks, making result analysis and interpretation easier for biologists and neuroscientists.

One of the limits of the current approach is the sensitivity to the choice of the threshold on the correlations. One perspective would be to explore all possible binarizations and then associate to each extracted subgraph a weight related to the size of the interval of correlation where it is statistically significant. Another limit is related to the fact the same threshold has been considered for all individuals. An

Graph	Param	graph Alz freq.	graph Ctrl freq.	graph pval	Edges	Edge Alz freq.	Edge Ctrl freq.	Edge pval
g_1	$\sigma = 0.25$ $\theta = 0.4$	6/16	10/13	0.034	Both_PERI - Both_ENT	16 / 16	13/13	1
					Both_CA2 - Both_CA1	15 / 16	13/13	0.359
					Both_CA2 - Both_CA3	13 / 16	13/13	0.099
					Both_DG - Both.SUB	15 / 16	13/13	0.359
					Both_DG - Both_CA1	14 / 16	12/13	0.672
					Both_DG - Both_CA3	10 / 16	10/13	0.404
					Both_CA3 - Both_CA1	14 / 16	12/13	0.672
					Both_ENT - Both.SUB	16 / 16	13/13	1
					Both_CA1 - Both.SUB	15 / 16	13/13	0.359
g_2	$\sigma = 0.25$ $\theta = 0.4$	10/16	13/13	0.013	Both_ACA - Both_MOp	10 / 16	13/13	0.013
					Both_ACA - Both_MOs	16 / 16	13/13	1
					Both_MOp - Both_MOs	16 / 16	13/13	1
g_3	$\sigma = 0.25$ $\theta = 0.3$	9/16	2/13	0.024	Both_ACA - Both_MOp	14 / 16	13/13	0.186
					Both_MOp - Both_MOs	16 / 16	13/13	1
					Both_MOp - Left_SS	12 / 16	7/13	0.233
					Both_MOp - Right_SS	15 / 16	4/13	0.000
					Both_MOs - Both_ACA	16 / 16	13/13	1

Table 1. Subgraphs detected as different between the two populations

alternative would be to consider a proportional thresholding, *i.e.*, a subject specific threshold that ensure that all individuals will have graphs with the same density of edges. Finally, another perspective would be to extend the approach to weighted graphs [10].

6. ACKNOWLEDGMENTS

The authors thank the Neuroscience & Cognition Center of Lille (Dr. Luc Buée and David Blum) and the Laboratory of Cognitive and Adaptive Neurosciences (LNCA) of Strasbourg (Dr. Anne-Laure Boutiller) for the mouse model. We also thank Jules Massart for his internship work. This project was supported by the Idex of Strasbourg University and ICube laboratory. The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

- [1] Fabrizio De Vico Fallani, Jonas Richiardi, Mario Chavez, and Sophie Achard, “Graph analysis of functional brain networks: Practical issues in translational neuroscience,” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 369, 06 2014.
- [2] Akihiro Inokuchi, Takshi Washio, and Hiroshi Motoda, “An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data,” in *PKDD 2000: Principles of Data Mining and Knowledge Discovery*, 2000, pp. 13–23.
- [3] Rakesh Agrawal and Ramakrishnan Srikant, “Fast algorithms for mining association rules,” in *Proc. of the 20th VLDB Conference*, 1994, pp. 487–499.
- [4] M. Kuramochi and G. Karypis, “Frequent subgraph discovery,” in *Proceedings 2001 IEEE International Conference on Data Mining*, 2001, pp. 313–320.
- [5] Xifeng Yan and Jiawei Han, “gSpan: graph-based substructure pattern mining,” in *2002 IEEE International Conference on Data Mining, 2002. Proceedings*, 2002, pp. 721–724.
- [6] Mohammed Elseidy, Ehab Abdelhamid, Spiros Skiadopoulos, and Panos Kalnis, “GRAMI: frequent subgraph and pattern mining in a single large graph,” *Proc. VLDB Endow.*, vol. 7, no. 7, pp. 517–528, Mar. 2014.
- [7] L. Degiorgis, M. Karatas, M. Sourty, E. Faivre, J. Lamy, V. Noblet, T. Bienert, M. Reiser, D. von Elverfeldt, L. Buée, D. Blum, AL. Boutillier, JP. Armspach, F. Blanc, and LA. Harsan, “Brain network remodelling reflects tau-related pathology prior to memory deficits in Thy-Tau22 mice,” *Brain*, 2020.
- [8] N. Cera, R. Esposito, F. Cieri, and A. Tartaro, “Altered cingulate cortex functional connectivity in normal aging and mild cognitive impairment,” *Front Neurosci*, vol. 13, 2019.
- [9] JM. Stephen, R. Montañó, CH. Donahue, JC. Adair, J. Knoefel, C. Qualls, Doug Ranken, and Cheryl J. Aine, “Somatosensory responses in normal aging, mild cognitive impairment, and alzheimer’s disease,” *J Neural Transm*, vol. 117, pp. 217–25, 2010.
- [10] Ngoc-Thao Le, Bay Vo, Lam B.Q. Nguyen, Hamido Fujita, and Bac Le, “Mining weighted subgraphs in a single large graph,” *Information Sciences*, vol. 514, pp. 149–165, Apr. 2020.

8. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. All experimental protocols were approved by the Regional Committee of Ethic in Animals Experiment of Strasbourg (CREMEAS, APAFIS n°2016033011298450).