Uncovering communities of pipelines in the task-fMRI analytical space

Elodie Germani, Elisa Fromont, Camille Maumet

Uncovering communities of pipelines in the task-fMRI analytical space

Elodie Germani
Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn ERL U 1228
F-35000 Rennes, France

Elisa Fromont*
Univ Rennes, IUF, Inria, CNRS, IRISA UMR 6074
F-35000 Rennes, France

Camille Maumet*
Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn ERL U 1228
F-35000 Rennes, France
camille.maumet@inria.fr

***Abstract*—Functional magnetic resonance imaging analytical workflows are highly flexible with no definite consensus on how to choose a pipeline. While methods have been developed to explore this analytical space, there is still a lack of understanding of the relationships between the different pipelines. We use community detection algorithms to explore the pipeline space and assess its stability across different contexts. We show that there are subsets of pipelines that give similar results, especially those sharing specific parameters (e.g. number of motion regressors, software packages, etc.), with relative stability across groups of participants. By visualizing the differences between these subsets, we describe the effect of pipeline parameters and derive general relationships in the analytical space.**

***Index Terms*—neuroimaging, pipeline, variability, communities, stability**

I. Introduction

Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that studies brain activity during the execution of specific tasks. This imaging modality is widely used to understand cognitive processes and neurological diseases [1], [2]. A typical fMRI data analysis can be split into three main steps: pre-processing, subject-level (first-level), and group-level (second-level) statistics. The sequence of steps performed during the analysis is called a pipeline. A large number of software packages and methods can be used at each step, which makes the choice of fMRI analysis pipeline a challenging process for researchers. In a meta-analysis [3], Carp examined the different pipelines used in more than 240 fMRI studies and found more than 200 unique combinations of analysis techniques. Recent studies [4], [5] also assessed the impact of these different pipelines in the results of an fMRI study. For instance, in [4], 70 research teams analyzed the same fMRI dataset using their favorite pipeline and found substantial differences in statistic maps and conclusions to binary research hypotheses (e.g. about the activation of the brain in a particular area during a particular task). This lack of robustness to different analytical conditions questions the validity of published results and calls for a better understanding of the analytical space. Methods have been developed for this purpose. For instance, Dafflon et al. [6] proposed a new method for identifying subsets of pipelines that are best to answer a problem for which ground truths are available, such as predicting the age of participants. A low-dimensional space was built to represent pipelines based on how they capture individual variability, using functional connectivity (FC) matrices. Such a (latent) space allows the measurement of distances between different analysis methods, which can help in understanding homogeneity (i.e. pipelines that give similar results) across pipelines but also heterogeneity (i.e. pipelines that have a different behavior). This allows to identify the extent to which different methods deviate from each other and to understand the variability within the pipeline space. It can also facilitate the selection of a set of pipelines that cover the diversity of the analysis space for further evaluation. Here, we focus on the use of community detection algorithms (i.e. clustering on graphs) to explore the pipeline space and assess its stability across different groups of subjects. We aim at identifying analysis pipelines that give similar results regardless of the context of the study (i.e. different contrasts or group of participants). If two pipelines are located in the same community across different contexts (i.e. similarity between results regardless of the context), we can consider that these are relatively stable and produce equivalent statistic maps.

We used raw data from the *Human Connectome Project* (HCP) [7] that we processed using several analysis pipelines which differed on a set of predefined parameters. For each pipeline, we carried out group analyses. Pipeline communities and their stability were assessed based on similarity graphs of statistic maps across pipelines for each group.

II. Material

This study was performed using data from the HCP. Written informed consent was obtained for all participants. The original study was approved by the Washington University Institutional Review Board and we agreed to the Data User Agreement. Unprocessed anatomical T1-weighted (T1w) and task-fMRI data [8]–[10] were used in this work. Among the 7 tasks performed, we selected data from the motor task, which was the simplest task performed in the study, and for which the protocol was very standard and robust. We used unprocessed data for the $N = 1080$ participants who completed this task.

III. Methods

In order to study the stability of the pipeline space across different groups of subjects, statistic maps were computed

Joint senior authorship.

for different groups of 50 randomly sampled participants (amongst 1080) with different pipelines. We computed graphs of similarity between the results of different pipelines for each group and used the *Louvain* community detection algorithm [11] to partition each graph. *Stability* was measured for each pair of pipelines as the number of groups (out of 1000) for which the two pipelines were located in the same community. All pipelines were implemented in Python3.8 using Nipype version 1.6.0 (RRID: SCR_002502) [12], FSL version 6.0.3 and SPM12 version r7771 on a NeuroDocker (RRID: SCR_017426) [13] environment. Graphs and communities were computed using NumPy (RRID:SCR_008633) and Networkx (RRID:SCR_016864). The code produced to run the experiments and to create the figures and tables of this paper is available in the Software Heritage public archive [14].

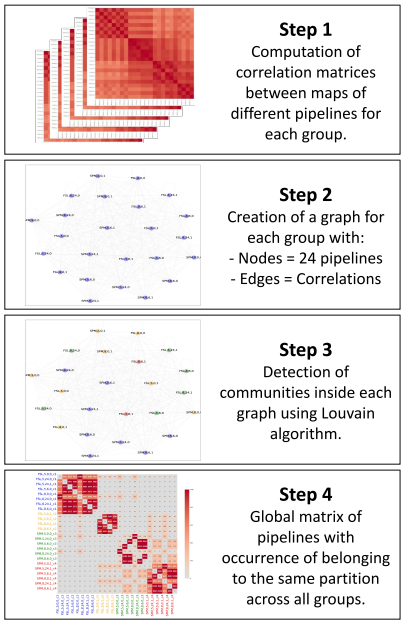

Fig. 1. Workflow of community detection in the pipeline space across different groups of participants and contrasts.

A. Analysis pipelines

24 different pipelines were applied to these data, with variations in terms of preprocessing and first-level analysis. Default values were used at each step except for 4 varying parameters:

- Software package: SPM (Statistical Parametric Mapping, RRID: SCR_007037) [15] or FSL (FMRIB Software Library, RRID: SCR_002823) [16].
- Smoothing kernel: Full-Width at Half-Maximum (FWHM) was equal to either 5mm or 8mm.
- Number of motion regressors included in the General Linear Model (GLM) for the first-level analysis: 0, 6 (3 rotations, 3 translations) or 24 (the 6 previous regressors + 6 derivatives and their 12 corresponding squares).
- Presence (1) or absence (0) of the derivatives (temporal+dispersion for SPM and temporal for FSL) of the Hemodynamic Response Function (HRF).

In the following, we refer to these parameters respectively as 'software', 'FWHM', 'motion regressors' and 'hrf derivatives'. For instance, pipelines built using the FSL software, smoothing with a kernel FWHM of 8mm, no motion regressors and no hrf derivatives will be denoted by 'fsl, 8, 0, 0'.

All pipelines were applied to each subject for the different motor contrasts studied (i.e. right-hand, right-foot, left-hand, left-foot and tongue). Within-group analyses were performed using SPM with default parameters. We used the same second-level analysis method for all pipelines in order to focus on differences in the first-level analyses. For second-level analysis, SPM and FSL provide equivalent statistical approaches thus, we arbitrarily chose to use SPM. 1,000 groups of 50 participants were randomly sampled among the 1,080 participants, leading to 1,000 statistic maps for each pipeline and contrast.

B. Graph computation and community detection

We computed the similarity for each pair of pipelines in terms of Pearson's correlation coefficient between statistic maps. This correlation matrix was used as an adjacency matrix to build an undirected weighted multi-graph for each group, with nodes representing the statistic maps of the different pipelines (V = 'fsl,0,0,0', 'fsl,0,0,1', etc.) and edges labeled by the correlation coefficient between each pipeline (E=('fsl,0,0,0','fsl,0,0,1'), etc.). After computation, each graph was partitioned using the Louvain algorithm [11] to detect the best partitions based on *modularity* optimization, which represents the density of links inside communities as compared to links between communities. The detected communities in each graph thus represented the pipelines that give similar results for the corresponding group.

To explore the stability of the pipeline space, we counted, for each pair of pipelines, the number of groups for which the two pipelines were located in the same community. A high value reflected a high similarity and stability of the two pipelines across groups. This matrix was used to build a second graph, global across groups, in which nodes represent the different pipelines and edges represent the stability measure mentioned above. Louvain community detection algorithm was again applied to this second graph to detect communities in which pipeline statistic maps were similar across different groups. These global graphs were computed for each contrast.

C. Communities-specific features

We studied the differences across communities by computing the mean statistic map of each community across groups for each contrast. We hence averaged the value of each voxel of the map across all groups and pipelines of the community to obtain a statistic map that was representative of the community for a given contrast. These were thresholded using a False Discovery Rate (FDR) of $p < 0.05$ and both maps (unthresholded and thresholded) were compared. We also computed the mean statistic maps across all groups for

each pipeline and compared them with the maps of other pipelines inside each community and across communities. For each pipeline map, we computed the number of activated voxels in the thresholded maps but also inside the Region of Interest (ROI) of the Primary Motor Cortex (M1), extracted from the probabilistic *Juelich Atlas*, available in *Nilearn* [17] (RRID: SCR_001362). The goal was to identify the specific patterns of each community, to understand why a pipeline was located inside a community, and to explore the potential impact on the results of the pipelines.

IV. Results

A. Pipeline community detection

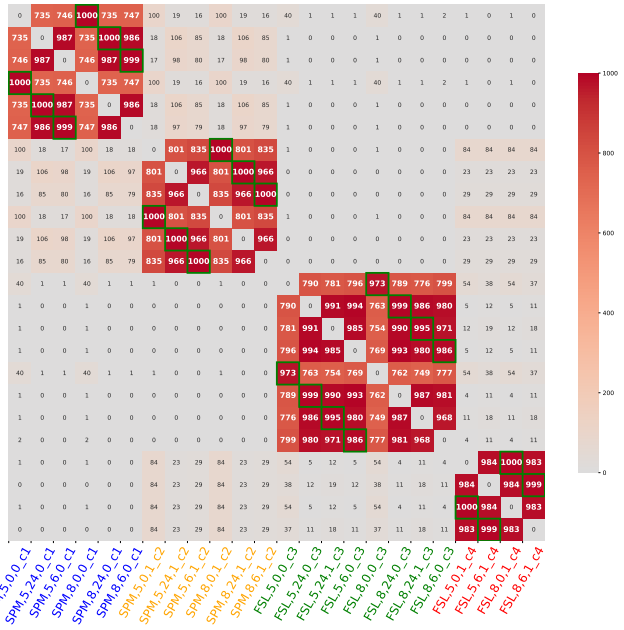

Fig. 2. Adjacency matrix representing the number of times each pair pipelines belong to the same community across different group-level statistic maps of the contrast right hand.

The adjacency matrix representing pairs of pipelines depending on the number of times they belonged to the same community across different group-level statistic maps of the contrast *right-hand* is shown in Fig. 2. The graph corresponding to this adjacency matrix was partitioned using the Louvain community algorithm and 4 communities were identified. They correspond to groups of pipelines that are frequently in the same community across groups (i.e. that give similar results for a high number of groups). The partitioning of this graph achieved a good modularity of 0.64 (modularity [11] is included between -0.5 and 1 and considered high above 0.3).

Inside each partition, pipelines shared specific parameters: for instance, the first community (in blue) contains all SPM pipelines without HRF derivatives.

Some pairs of pipelines are almost always located in the same community: for instance, every pair of pipelines that used the same software, the same number of motion regressors, and the same use of HRF derivatives but different smoothing kernel (5 and 8 mm FWHM) belong to the same community more than 950 times across $1,000$ groups (see cells highlighted in green in Fig.2). Other pairs of pipelines are also located in the same community for a high number of groups (more than 700), which indicates a high stability of pipeline relationships across groups for this contrast.

B. Visualization of communities specificities

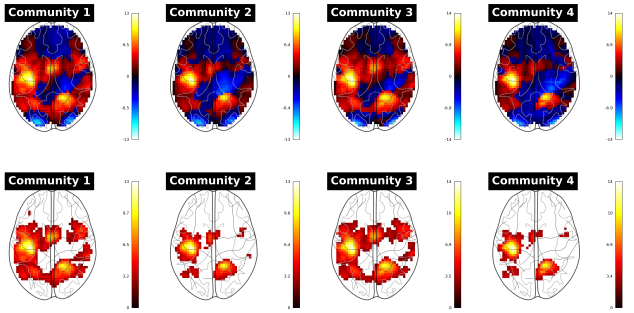

Fig. 3. Community statistic maps for the contrast right-hand. Unthresholded maps (upper) and thersholded maps (lower) with FDR-corrected $p < 0.05$.

Mean unthresholded (upper) and thresholded (lower) statistic maps of communities are displayed in Fig.3. The global activation patterns are similar across communities, with differences in the extent of the activation area between communities 1-3 and communities 2-4.

TABLE I
MEAN NUMBER OF ACTIVATED VOXELS IN THE THRESHOLDED STATISTIC MAPS (1ST ROW) AND INSIDE THE ROI OF THE PRIMARY MOTOR CORTEX (2ND ROW) FOR EACH COMMUNITY FOR THE CONTRAST RIGHT-HAND.

Community	**1**	**2**	**3**	**4**
Whole maps	2,585	808	3,009	978
ROI	354	239	422	298

This observation is confirmed by the number of activated in the thresholded maps of the pipelines inside each community (Tab.I). Statistic maps of communities 1 and 3 show a high number of activated voxels ($N = 2,585$ and $3,009$) compared to communities 2 and 4 ($N = 808$ and 978). The numbers of activated voxels inside the ROI of the Primary Motor Cortex for each community are consistent with this observation.

V. Discussion

In this work, we used a community detection method to explore the task-fMRI analytical space and understand relationships between pipelines. Pipelines were grouped in communities based on the similarity of their results and these communities were relatively stable across different groups of participants. We found that pipelines with similar values for specific parameters gave similar results for almost all groups, e.g. same software, motion regressors and HRF derivatives but different smoothing kernel FWHM. Some pairs of pipelines showed a slightly lower number of co-occurrence in the same community compared to others. This could be due to specific

properties of the data (high values for motion regressors in some groups, etc.) and their interaction with the pipeline parameters. The pipeline space we studied was restricted to changes in 4 parameters, leading to 24 different pipelines, but the fMRI analytical space is much larger and this method could also be applied to a higher number of pipelines, leading to a larger graph representing other pipelines parameters. Here, the parameters that seemed to separate pipelines across communities were the software package and the use of HRF derivatives to model signal. Communities containing pipelines with different use of HRF derivatives (e.g. communities 1 and 3 vs communities 2 and 4) were distinguishable by the extent of the activation area in the resultant statistic maps. We can suppose that the use of HRF derivatives (communities 2 and 4) in pipelines leads to a more restricted activation area in the resultant statistic maps. Statistic maps of communities with different software packages show very similar activation patterns but differs in terms of number of activated voxels. This can be explained by the differences in terms of signal demeaning before or after fitting the model [18] and pre-whitening methods [19]. FSL analyses tend to lead to higher statistical values and thus, more significantly activated voxels, which might explain the lower correlations between the maps of pipelines coming from different software packages. Using this method, we were able to identify subsets of pipelines that give similar results in terms of statistic maps across groups. Here, we only studied the contrast *right-hand* of the motor task but other contrasts might lead to different communities. Pipelines between communities might also lead to different results and have an impact on study results. For instance, here, the conclusion about activation or not of the Primary Motor Cortex during the task would not be impacted by the pipeline choice since we found activated voxels in the ROI for each pipeline and community, which might not be the case with other contrasts. By studying other contrasts, we could derive more general relationships between the different pipelines and explore in more depth the effects of pipelines' parameters.

VI. Conclusion

Our study presents a method to explore the analytical space and its stability across different contexts (here, groups of participants). In future work, this workflow may be used with other sets of pipelines and other paradigms to assess the generalizability of our results. These results could thus be used to tackle analytical variability with, for instance, the selection of representative pipelines.

Acknowledgment

This work was partially funded by Region Bretagne (ARED MAPIS) and Agence Nationale pour la Recherche for the program of doctoral contracts in artificial intelligence (project ANR-20-THIA-0018).
Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.
We thank Jeremy Lefort-Besnard for its advice and its code on Louvain communities.

References

[1] W. Yin, L. Li, and F.-X. Wu, "Deep learning for brain disorder diagnosis based on fMRI images," *Neurocomputing*, vol. 469, pp. 332–345, 2022.
[2] O. Firat, L. Oztekin, and F. T. Y. Vural, "Deep learning for brain decoding," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2784–2788.
[3] J. Carp, "The secret lives of experiments: Methods reporting in the fMRI literature," *NeuroImage*, vol. 63, no. 1, pp. 289–300, 2012.
[4] R. Botvinik-Nezer, F. Holzmeister, C. F. Camerer, A. Dreber, J. Huber, M. Johannesson, M. Kirchler, R. Iwanir, J. A. Mumford, R. A. Adcock *et al.*, "Variability in the analysis of a single neuroimaging dataset by many teams," *Nature*, vol. 582, no. 7810, pp. 84–88, 2020.
[5] A. Bowring, C. Maumet, and T. E. Nichols, "Exploring the impact of analysis software on task fmri results," *Human Brain Mapping*, vol. 40, no. 11, pp. 3362–3384, 2019.
[6] J. Dafflon, P. F. Da Costa, F. Váša, R. P. Monti, D. Bzdok, P. J. Hellyer, F. Turkheimer, J. Smallwood, E. Jones, and R. Leech, "A guided multiverse study of neuroimaging analyses," *Nature Communications*, vol. 13, no. 1, p. 3758, 2022.
[7] D. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. Curtiss, S. Della Penna, D. Feinberg, M. Glasser, N. Harel, A. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. Petersen, F. Prior, B. Schlaggar, S. Smith, A. Snyder, J. Xu, and E. Yacoub, "The human connectome project: A data acquisition perspective," *NeuroImage*, vol. 62, pp. 2222–2231, 2012.
[8] D. A. Feinberg, S. Moeller, S. M. Smith, E. Auerbach, S. Ramanna, M. F. Glasser, K. L. Miller, K. Ugurbil, and E. Yacoub, "Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging," *PLOS ONE*, vol. 5, no. 12, pp. 1–11, 2010.
[9] K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, and L. L. Wald, "Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty," *Magnetic Resonance in Medicine*, vol. 67, no. 5, pp. 1210–1224, 2012.
[10] J. Xu, S. Moeller, J. Strupp, E. Auerbach, L. Chen, D. A. Feinberg, K. Ugurbil, and E. Yacoub, "Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband epi," in *Proceedings of the 20th Annual Meeting of ISMRM*, vol. 2306, 2012, pp. 1907–1913.
[11] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, oct 2008.
[12] K. Gorgolewski, "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python," *Frontiers in Neuroinformatics*, p. 15, 2017.
[13] "Neurodocker." [Online]. Available: https://github.com/ReproNim/neurodocker
[14] E. Germani, E. Fromont, and C. Maumet, "Software heritage archive for the gitlab repository "pipeline_distance"." https://archive.softwareheritage.org/browse/origin/directory/?origin_url=https://gitlab.inria.fr/egermani/pipeline_distance, 2023.
[15] W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. E. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2011.
[16] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL." *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.
[17] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Frontiers in Neuroinformatics*, 2014.
[18] "Comparing GLM modelling approaches of SPM and FSL." [Online]. Available: https://gist.github.com/fladd/9d3707f7a457cf8272042e742952c950
[19] W. Olszowy, J. Aston, C. Rua, and G. B. Williams, "Accurate autocorrelation modeling substantially improves fMRI reliability," *Nature Communications*, vol. 10, p. 1220, 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6428826/