Distributional Representation for Resting-state Functional Brain Connectivity Analysis

Jiating Zhu and Jiannong Cao

The Hong Kong Polytechnic University, Hong Kong, China
sophie. z@163.com, csjcao@polyu.edu.hk

Abstract. Most analyses on functional brain connectivity across a group of brains are under the assumption that the positions of the voxels are aligned into a common space. However, the alignment errors are inevitable. To address such issue, a distributional representation for resting-state functional brain connectivity is proposed here. Unlike other relevant connectivity analyses that only consider connections with higher correlation values between voxels, the distributional approach takes the whole picture. The spatial structure of connectivity is captured by the distance between voxels so that the relative position information is preserved. The distributional representation can be visualized to find outliers in a large dataset. The centroid of a group of brains is discovered. The experimental results show that resting-state brains are distributed on the 'orbit' around their categorical centroid. In contrast to the main-stream representation such as selected network properties for disease classification, the proposed representation is task-free, which provides a promising foundation for further analysis on functional brain connectivity in various ends.

Keywords: Distributional representation · Functional brain connectivity · Categorical centroid· Outliers visualization.

1 Introduction

A good low dimensional representation, which captures the important characteristics from the large information pool, is helpful in discovering new information across hundreds of brains. But due to the curse of dimensionality, it is not practicable to unfold raw functional magnetic resonance imaging (fMRI) data (4D tensor) directly into a vector to be trained via machine learning methods. The raw fMRI data is in large size (up to several gigabytes) and has inherent high levels of noise [1]. Studies using machine learning methods on low dimensional embedding of functional brain networks have been attempted. However, most of the recent researches on this topic are region-based [2–4], which consider around a hundred regions. Nevertheless, region-based analysis limits the evaluation of inter-regional connectivities and restricts to certain anatomical areas. It has been discussed that voxel-based functional networks present features more prominently than region-based ones [5], such as small-world property.

Inspired from the text embedding method [6], which based on the distributed representation of words and documents, we propose a representation of the functional brain based on the connectivity distribution. The advantage of this novel representation is that it will not be affected by the errors from the alignment across brains.

The precise position and extent of the functional areas differ across individuals [7]. Even if 'perfect' anatomical alignment was possible, it would not align functional topography [8]. Currently, researchers use two approaches to compare a group of brains by either giving stimulus to the subjects or align different brains with anatomical masks. Both of the approaches will cause errors.

By investigating the proposed distributional representation, this paper aim to answer two questions:

(1) Is it possible to have a low dimensional representation for the whole functional brain connectivity in a voxel-level?

(2) How can we trust the prepossessed functional brain image data? Is it possible to find outliers only from the given data and eliminate them before further analysis?

2 Problem Definition

Definition 1. *A correlation matrix R captures the voxels' correlation information in a brain with N regions, where the entry r_{ij} is the Pearsons correlation coefficient value between regions i and j $(i, j = 1, ..., N)$. Here, a region refers to a three dimensional area in voxel-level (or higher region resolution).*

There will exist different correlation matrices for a specific brain with N regions due to different region partitions. Simply changing the numbering of the regions or assigning different anatomical positions for the regions will generate a different correlation matrix. Such a matrix has the ambiguity for representing the brain connectivity. Hence, we consider the distribution of correlation to avoid the ambiguity.

Definition 2. *A correlation distribution is a probability distribution that provides the probabilities of occurrence of different possible correlation values between voxels within a brain.*

In comparison with the correlation matrix, the correlation distribution loses the spatial information of the brain. Thus, we introduce a distance distribution to capture the information of the functional topology.

Definition 3. *A distance matrix D captures the voxels' relative position information in a brain with N regions, where the entry d_{ij} is the Euclidean distance between the centers of regions i and j $(i, j = 1, ..., N)$.*

Definition 4. *A distance distribution is a probability distribution that provides the probabilities of occurrence of different possible distance values between voxels within a brain. It depicts the spatial information of a brain by calculating the distribution of different possible values for the distance between a randomly selected pair of voxels.*

Definition 5. *A correlation-distance joint distribution is a probability distribution that provides the probabilities of co-occurrence of different possible correlation values and distance values between voxels within a brain.*

However, different distance values will give different contribution of spatial information to the joint probability distribution, because the distance distribution is a non-uniform distribution. To reduce this side effect, we consider a conditional distribution of correlation given distance.

Definition 6. *A correlation-distance conditional distribution is a probability distribution that provides the probabilities of occurrence of different possible correlation values between voxels when each of different possible values of the distance between voxels are given within a brain.*

Definition 7. *Distributional representation refers to an expression of functional brain connectivity by correlation distribution, correlation-distance joint distribution or correlation-distance conditional distribution.*

3 Distributional Representation Analysis

3.1 Data Description

The data we analyzed in this paper is from the ADHD-200 sample which collected resting-state fMRI scans from subjects with ADHD and typically developing controls in 8 international imaging sites. In this study, we only analyzed the preprocessed data from NYU with the largest number of subjects. The preprocessed data used NIAK pipeline and has extracted time courses from around 3000 regions of interests [9]. One or two resting-state fMRI scans were acquired for each subject in the NYU data. We picked the first scans (216 subjects) for analysis in this paper. The detailed information about the data can be found at http://preprocessed-connectomes-project.org/adhd200/.

3.2 Single Brain Analysis

We use a histogram to estimate the correlation distribution for a single subject. The data sample used for calculating the histogram is the upper triangle of the correlation matrix R, which is denoted by R_{tri}, where the entry in R_{tri} is $r_{ij}(i < j)$.

The histogram $R(r_k)$ for the correlation distribution is calculated by $R(r_k) = c_r(k)/(w_r n_r)$, where w_r is the bin width, n_r is the number of bins, $c_r(k)$ is the number of entries that fall into the kth bin. Thus, $r_k = (k-1)w_r - 1, k = 1, 2, ..., n_r$. Figure 1 shows the correlation distributions for 4 subjects by histograms, each with 200 bins over the range $[-1, +1]$. The correlation distributions in Figure 1 look like Gaussian distributions.

Similarly, we use a histogram to estimate the distance distribution, taking the upper triangle D_{tri} of the distance matrix D as the data sample. The histogram

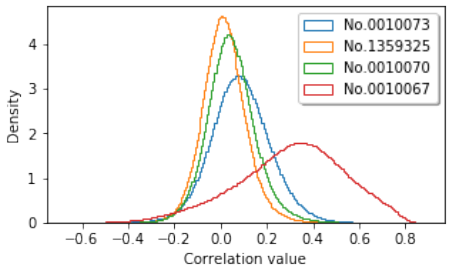

Fig. 1. Correlation distributions

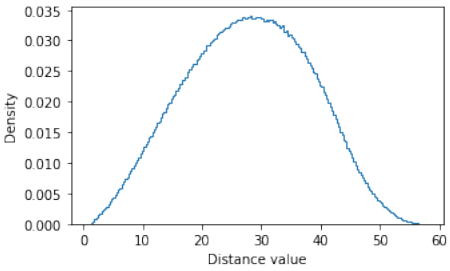

Fig. 2. The distance distribution

$D(d_k)$ for the distance distribution is calculated by $D(d_k) = c_d(k)/(w_d n_d)$, where w_d is the bin width, n_d is the number of bins, $c_d(k)$ is the number of entries that fall into the kth bin. Thus, $d_k = (k-1)w_d, k = 1, 2, ..., n_d$.

Figure 2 shows the distance distribution by a histogram with 200 bins over the range $[0, 60]$. The distance distribution also looks like a Gaussian distribution. Theoretically, the spatial information (distance distribution) for every brain should be the same.

We use a two-dimensional histogram to estimate the correlation-distance joint distribution for a single subject. The two dimensional histogram is

$$h(r_i, d_j) = c_{rd}(i, j)/(w_r w_d n_r n_d), \tag{1}$$

where $c_{rd}(i, j)$ is the number of entries that fall into both the ith bin in $R(r_i)$ and the jth bin in $D(d_j)$. The two-dimensional histograms for the four subjects in Figure 1 are shown in Figure 3. We can see that the first three subjects from left to right look like two-dimensional Gaussian distributions. However, the last one (subject No.0010067) has an obviously abnormal shape, whose alien features can not be seen prominently in correlation distributions in Figure 1. While the first three are marked 'pass', the subject No.0010067 is marked 'questionable' in the quality control that come along with the preprocessed ADHD-200 NYU data. This means that our finding is consistent with the ground truth information.

In Figure 3, the peaks (the smallest ovals) of the two-dimensional histograms all appear around the distance value 30. The reason is that the distance distri-

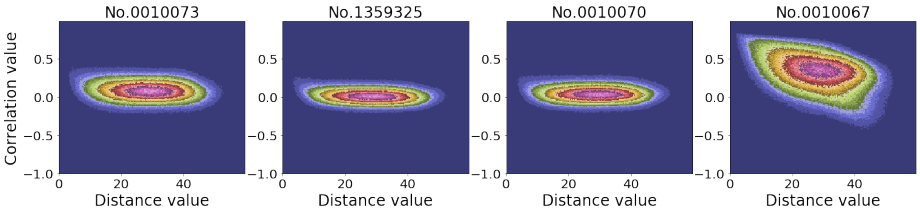

Fig. 3. Correlation-distance joint distributions

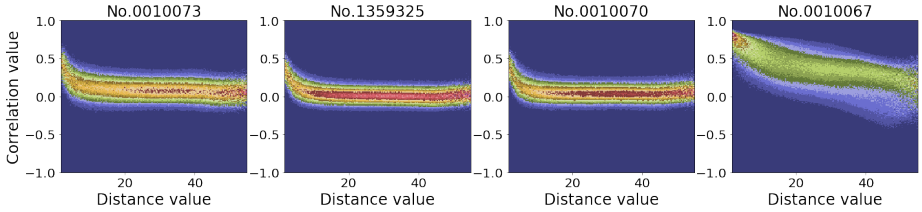

Fig. 4. Correlation-distance conditional distributions

bution, which also reaches its peak around 30 (see Figure 2), affects the joint distribution. To eliminate such side effect, we introduce a conditional distribution to have a deeper investigation of the data. The two-dimensional histogram for estimating the correlation-distance conditional distribution for a single subject is

$$h(r_i|d_j) = h(r_i, d_j)/D(d_j). \tag{2}$$

Figure 4 shows the conditional distribution by a histogram with 200 bins in distance over the range $[1.55, 55]$ (the data around the two ends are removed due to the noise for visualization). As we can see from Figure 4, correlation values tend to be higher around the shortest distance, which is consistent with our common knowledge that interactions are more active between neighboring voxels.

3.3 Group Analysis

To study the group characteristic in distributional representation, we first explore the centroid of the group and then focus on the divergence from the centroid in the group.

The following are four ways to represent the centroid.

– AJH: The accumulative two-dimensional joint histogram $H(r_i, d_j)$ as the centroid of a group with M subjects is calculated by

$$H(r_i, d_j) = \frac{1}{M} \sum_{b=1}^{M} h_b(r_i, d_j), \tag{3}$$

where $h_b(r_i, d_j)$ is the two-dimensional histogram for a subject numbered b in the group.

– EGD: The estimated two-dimensional Gaussian distribution $f_g(r, d)$ as the centroid is optimized by

$$f_g(r, d) = \arg\min_{f} D_{KL}(H(r, d)||f(r, d)), \tag{4}$$

where D_{KL} is the Kullback-Leibler (KL) divergence measure,

$$D_{KL}(H(r, d)||f(r, d)) = \sum_{i,j} H(r_i, d_j) \log \frac{H(r_i, d_j)}{f(r_i, d_j)}, \tag{5}$$

and $f(r, d)$ is a two-dimensional Gaussian distribution function

$$f(x, y) = A \exp(-a(x - x_0)^2 + 2b(x - x_0)(y - y_0) + c(y - y_0)^2). \quad (6)$$

– ACH: The accumulative two-dimensional conditional histogram $H(r_i|d_j)$ as the centroid is calculated by

$$H(r_i|d_j) = H(r_i, d_j)/D(d_j). \quad (7)$$

– ECD: The estimated two-dimensional conditional distribution $f_g(r|d)$ as the centroid is optimized by

$$f_g(r|d) = \arg\min_f D_{KL}(H(r|d)||f(r|d)), \quad (8)$$

where $f(r|d) = f(r, d)/f(d)$, $f(d)$ is a one-dimensional Gaussian distribution function.

Figure 5 shows the centroid of a group in the four ways. The group contains 84 typically developing children in ADHD-200 NYU dataset. We kicked off the data that is marked as questionable in quality control, or the data that shows abnormal shape like the subject No.0010067 in Figure 3.

We calculate the KL divergence from the group centroid to each subject in the dataset. Furthermore, we fit two probability distribution functions to the KL divergence distribution. Figure 6 shows the KL divergence from the group centroid in four different aspects. The representation of each individual is the correlation-distance joint distribution if the centroid is represented by AJH or EGD, while the correlation-distance conditional distribution is applied when the centroid is represented by ACH or ECD. We can see that most of the KL divergence values are below 0.2. This implied that most of the subjects cluster around the centroid. However, the shape of the probability distribution suggests that individuals in a group are gathered on an 'orbit' around the group centroid, not simply clustered closely around the group centroid. In Figure 6(c),

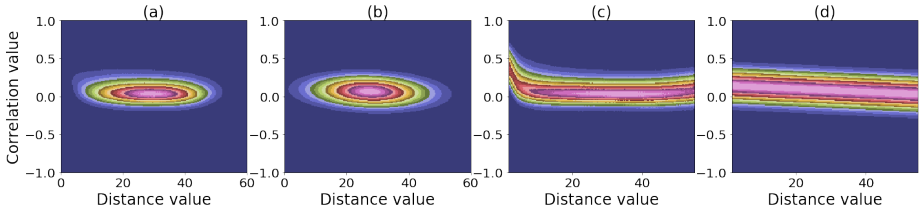

Fig. 5. Centroid for the group of typically developing children in ADHD-200 NYU dataset. (a) Accumulative two-dimensional joint histogram for the group centroid. (b) Estimated two-dimensional Gaussian distribution for the group centroid. (c) Accumulative two-dimensional conditional histogram for the group centroid. (d) Estimated two-dimensional conditional distribution for the group centroid.

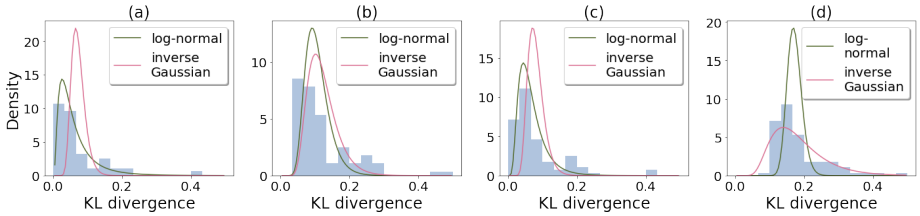

Fig. 6. Distributions of the KL divergence from the group centroid of typically developing children in ADHD-200 NYU dataset. The figure shows the histogram of the KL divergence values with 15 bins. (a) The divergence from the AJH centroid. (b) The divergence from the EGD centroid. (c) The divergence from the ACH centroid. (d) The divergence from the ECD centroid.

for instance, the probabilities for the individuals with very small KL values drop dramatically.

With a proper probability distribution function, the 'obrit' will be well depicted. It seems that the log-normal distribution is more suitable for estimating the KL divergence from the centroids of AJH and ACH, while the inverse Gaussian distribution is more suitable for the centroids of EGD and ECD.

A maximum likelihood estimation is used to estimate the parameters in the two probability distribution functions. Optimized parameters θ_g for a probability distribution function $f(kl_b|\theta)$ are calculated by

$$\theta_g = \arg\max_{\theta} \sum_{b=1}^{M} \log f(kl_b|\theta), \tag{9}$$

where $kl_b = D_{KL}(h_b||f_g)$ for the subject numbered b, $\theta = \mu$ if the probability distribution f is the single parameter inverse Gaussian distribution $f(x;\mu,\mu^2)$, or $\theta = (\mu,\sigma)$ if the probability distribution f is a log-normal distribution.

4 Results and Discussion

4.1 Outliers

We have a look at the correlation-distance joint distributions of total 216 subjects in preprocessed ADHD-200 NYU dataset. 15 obviously abnormal shapes are found (see Figure 7), where 11 (73%) of them are marked as questionable in the quality control sheet.

The other four subjects are marked pass in the dataset. However, two of them (No.3662296 and No.6568351) have the same look (only a horizontal line appeared across the center) as the two questionable subjects (No.0010015 and No.0010003) shown in the first row in Figure 7. The third one (No.0010114), on the second row at the second column, has the characteristic similar to the questionable subject No.0010005 at the first column, where there suddenly appears a horizontal line across the center. The forth one (No.1517240), next to

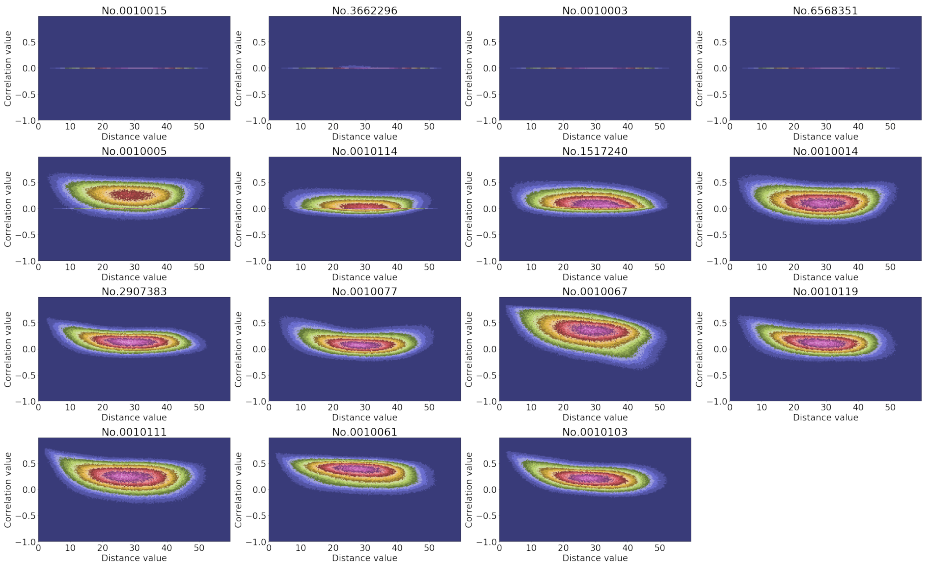

Fig. 7. Obviously abnormal shapes found manually.

the subject No.0010114, also has the similar phenomenon, which seems to be collapsed to a horizontal line across the center, unlike other normal shapes that are symmetric-like. Thus, we believe that our distributional representation can provide useful information for finding alienated data.

4.2 Group Difference

We use two approaches to distinguish which group a subject belongs to. The typically developing children in ADHD-200 NYU dataset are regarded as a group named Typical, while subjects with ADHD-Combined, ADHD-Hyperactive /Impulsive or ADHD-Inattentive are regarded as a group named ADHD.

Simply comparing KL values $KL_{Typical}$ and KL_{ADHD} of an individual, we guess the individual falls into the group that has the smaller KL divergence to it. $KL_{Typical}$ is the KL divergence from the centroid of the group Typical to the individual distribution, while KL_{ADHD} is the KL divergence from the centroid of the group ADHD to the individual distribution. The number of individuals that are correctly found from each group is shown in the row labeled 'Compare KL' in Table 1.

The second way to guess the category of each individual is to estimate the probability of its KL value in each of the two groups Typical and ADHD, respectively, by an estimated probability distribution function $f(kl|\theta_g)$. We guess the individual falls into the group with the higher probability for its corresponding KL value. The results for the two probability distribution functions, namely log-normal and inverse Gaussian, are shown in the rows labeled 'Log-normal' and 'Inverse Gaussian', respectively, in Table 1.

Table 1. Different ways to recognize a group of subjects with four different centroid representations. The number of subjects in the group Typical is 84 and the number of subjects in the group ADHD is 89.

Centroid type		AJH	EGD	ACH	ECD
Compare KL	Typical	52 (62%)	55 (65%)	**72 (86%)**	51 (61%)
	ADHD	45 (50%)	38 (42%)	**64 (72%)**	54 (61%)
Inverse Gaussian	Typical	34 (40%)	52 (62%)	65 (77%)	26 (31%)
	ADHD	34 (40%)	24 (27%)	32 (34%)	58 (65%)
Log-normal	Typical	51 (61%)	37 (44%)	37 (44%)	17 (20%)
	ADHD	41 (46%)	49 (55%)	41 (46%)	69 (78%)
Centroid divergence		0.00103	0.00109	0.00217	0.00121

As shown in Table 1, the ACH approach has the best performance, where 86% of individuals in the group Typical are identified as Typical and 72% of individuals in the group ADHD are identified as ADHD. This performance is explainable. The divergence from the ACH centroid of the group ADHD to the ACH centroid of the group Typical is the biggest (see the bottom row). Hence, the two groups can be differentiated more easily. The ACH representation also has a significant improvement than the AJH representation, which justified that conditional distributions do eliminate the side effect of the unbalanced distance contributions to the joint distributional representations. Despite the fact that the two estimated probability distribution functions we attempted, namely Inverse Gaussian and Log-normal, do not show better performance, it is still possible to find appropriate estimation to improve the performance.

Because of the rough estimation, the estimated representations EGD and ECD do not have good performance on distinguishing groups (see columns labeled 'EGD' and 'ECD'). We can see from Figure 5(c) and Figure 5(d) that the estimated two-dimensional conditional distribution (ECD) loses some information at lower distance values in comparison with the accumulative two-dimensional conditional histogram (ACH). With a better estimation approach, we believe that the ECD representation will have a better performance.

5 Conclusion and Future Work

The greatest benefit of our distributional representation is that we can avoid the alignment errors when we compare a large group of brains. With our distributional representation, we can get a preview of the data and the data can be visualized to find outliers which might not be detected with a standard quality control procedure. The distributional representation persevered the majority connectivity information below the correlation value 0.7, which is around the commonly picked threshold for functional brain network analysis [5]. Although the dominant part of the representation stores the lower correlation values, it conveys the higher correlation information inside the distribution.

Instead of voxels' positions, the distance distribution between voxels is taken into consideration to preserve the spatial information in the functional brain, so

that the position alignment can be avoided. The unbalanced contribution from the distance distribution is relived by a conditional distribution. Our quantitative results verified the effectiveness of the proposed conditional distribution (see Table 1).

An interesting thing is observed. The divergence distribution functions as an 'orbit', on which individuals are gathered around a group centroid. This phenomena, which needs further exploration, is beyond the conventional concept that similar subjects inhabit on the whole neighboring area around the centroid.

In the future, we will look for better distributional representation [10] and test our method on larger and various datasets to validate our findings. Novel analysis algorithms will be designed for various tasks, such as clustering and classification, on a basis of the distributional representation.

Moreover, it is possible to apply a distribution-based embedding method, such as the techniques in text embedding [6], to construct a task-free low dimensional representation for the whole functional brain connectivity in a voxel-level.

References

1. Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V.: Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends in cognitive sciences **10**(9) (2006) 424–430
2. Cao, B., He, L., Wei, X., Xing, M., Yu, P.S., Klumpp, H., Leow, A.D.: t-bne: Tensor-based brain network embedding. In: Proceedings of the 2017 SIAM International Conference on Data Mining, SIAM (2017) 189–197
3. Wang, S., He, L., Cao, B., Lu, C.T., Yu, P.S., Ragin, A.B.: Structural deep brain network mining. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2017) 475–484
4. Shen, H., Wang, L., Liu, Y., Hu, D.: Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fmri. Neuroimage **49**(4) (2010) 3110–3121
5. Hayasaka, S., Laurienti, P.J.: Comparison of characteristics between region-and voxel-based network analyses in resting-state fmri data. Neuroimage **50**(2) (2010) 499–508
6. Tang, J., Qu, M., Mei, Q.: Pte: Predictive text embedding through large-scale heterogeneous text networks. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2015) 1165–1174
7. Nishimoto, S., Nishida, S.: Lining up brains via a common representational space. Trends in cognitive sciences **20**(8) (2016) 565–567
8. Chen, P.H.: Multi-view Representation Learning with Applications to Functional Neuroimaging Data. PhD thesis, Princeton University (2017)
9. Bellec, P., Chu, C., Chouinard-Decorte, F., Benhajali, Y., Margulies, D.S., Craddock, R.C.: The neuro bureau adhd-200 preprocessed repository. Neuroimage **144** (2017) 275–286
10. Zhu, J., Cao, J.: Group analysis by visualized distributional representation for resting-state functional brain connectivity. In: Proceedings of the 14th International Conference on Semantics, Knowledge and Grids., IEEE (2018)