



Connectivity-Driven Brain Parcellation via Consensus Clustering

Anvar Kurmukov^{1,2}, Ayagoz Musabaeva¹, Yulia Denisova¹, Daniel Moyer⁴,
and Boris Gutman^{1,3}(✉)

¹ The Institute for Information Transmission Problems, Moscow, Russia

² National Research University Higher School of Economics, Moscow, Russia

³ Illinois Institute of Technology, Chicago, USA

`bgutman1@iit.edu`

⁴ University of Southern California, Los Angeles, USA

Abstract. We present two related methods for deriving connectivity-based brain atlases from individual connectomes. The proposed methods exploit a previously proposed dense connectivity representation, termed continuous connectivity, by first performing graph-based hierarchical clustering of individual brains, and subsequently aggregating the individual parcellations into a consensus parcellation. The search for consensus minimizes the sum of cluster membership distances, effectively estimating a pseudo-Karcher mean of individual parcellations. We assess the quality of our parcellations using (1) Kullback-Liebler and Jensen-Shannon divergence with respect to the dense connectome representation, (2) inter-hemispheric symmetry, and (3) performance of the simplified connectome in a biological sex classification task. We find that the parcellation based-atlas computed using a greedy search at a hierarchical depth 3 outperforms all other parcellation-based atlases as well as the standard Dessikan-Killiany anatomical atlas in all three assessments.

1 Introduction

The ability to quantify how the human brain is interconnected in vivo has opened the door to a number of possible analyses. In nearly all of these, brain parcellation plays a crucial role. Variations in parcellation significantly impact connectome reproducibility, derived graph-theoretical measures, and the relevance of connectome measures with respect to biological questions of interest [16]. A natural approach is then to use individual densely sampled connectomes to drive the parcellation directly, leading to a more compact, connectivity-aware set of brain regions and resulting graph, as done in e.g. [10]. A comprehensive review of parcellation methods and their effects on the derived connectome quality is given in [17]. Because individual connectivity data is at once very informative and highly redundant, there is a great flexibility in how parcels can be derived from dense, highly resolute graphs. It is possible for example to derive (1) a unified population-based atlas, (2) individual-level parcellations with cross-subject

label mapping, or (3) individual parcellations with no inter-subject label correspondence. While the first approach is appealing for its simplicity and ease of interpretation, the second and third may enable the researcher to reveal some individual aspect of the connectome that is lost in the aggregate atlas.

In this work, we attempt to bridge these three approaches by first constructing maximally flexible hierarchical parcellations, and then finding a unifying set of labels and parcels to maximize individual agreement. We use the a continuous representation of a brain connectivity [8] as our initial dense connectome representation. Continuous connectivity is a parcellation-free representation of tractography-based, or “structural” connectomes that is based on the Poisson point process. Once individual parcellations are computed, we obtain a group-wise parcellation using partition ensemble algorithm. We assess quality of the resulting parcellations in three ways. (1) We use the continuous connectome framework to compare parcellation-approximate and exact edge distribution functions. (2) We compare performance of the resulting graphs on a gender classification task. (3) We also show that without any explicit knowledge of brain geometry and based solely on graph connectivity we obtain comparatively symmetric parcellations.

2 Methods

2.1 Continuous Connectome

The continuous connectome model (ConCon) treats each tract as an observation of an inhomogeneous symmetric Poisson point process with the intensity function given by

$$\lambda : \Omega \times \Omega \rightarrow \mathbb{R}^+, \quad (1)$$

where Ω denote union of two disjoint topologically spherical brain hemispheres, representing cortical white matter boundaries. In practice, ConCon uses cortical mesh vertices as nodes of connectivity graph. From such a representation, a “discrete” connectivity graph could be computed from any particular cortical parcellation P . We follow definitions from [8] and call $P = \{E_i\}_{i=1}^N$ a parcellation of Ω if $E_1 \dots E_k \subseteq \Omega$ such that $\cup_i E_i = \Omega$, and N is the number of parcels (ROIs). Edges between regions E_i and E_j can then be computed by integration of the intensity function:

$$\mathcal{C}(E_i, E_j) = \iint_{E_i, E_j} \lambda(x, y) dx dy, \quad (2)$$

Due to properties of the Poisson Process, $\mathcal{C}(E_i, E_j)$ is the expectation of the number of observed tracts between E_i and E_j . In the context of connectomics, this is the expected edge strength.

2.2 Graph Clustering

Once we obtain all individual continuous connectomes, we partition each independently into a set of disjoint communities. For graph clustering we use the

Louvain modularity algorithm [1], as it has shown good results in multiple neuroimaging studies [5, 7, 9, 12]. This algorithm consist of two steps. The first step combines locally connected nodes into communities, while the second step builds new meta graph. The nodes of the meta-graph are communities from the previous step, and the edges are defined as the sum of all inter-community connections of the new nodes. The algorithm in [1] cycles over these steps iteratively, converging when further node clustering leads to no increase in modularity. We follow the hierarchical brain concept [7], repeating the clustering procedure iteratively. After the initial parcellation, we further cluster each individual parcel as an independent graph. In this work, we repeat the process three times. For each (i 'th) continuous connectome this procedure yields a three-level hierarchically embedded partition: $P_i^I, P_i^{II}, P_i^{III}$, (see Fig. 1).

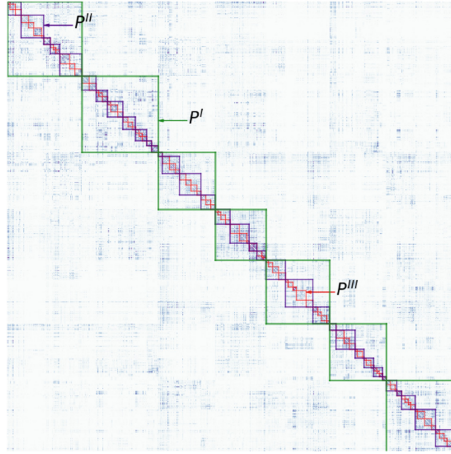


Fig. 1. Adjacency matrix of a sample continuous connectome. Rows and columns are reordered according to partition of the third hierarchical level. Boxes of different color represents clusters of different hierarchical levels. P^I clusters are obtained first, next we reapply clustering on each detected P^I cluster and obtain P^{II} . This is repeated once more to obtain P^{III} (Color figure online)

2.3 Consensus Clustering

In order to obtain a unified parcellation for all subjects, we use consensus clustering. The concept was developed for aggregating multiple partitions of the same data into a single partition. We define the *average* partition over all individual partitions $\{P_i\}$ as:

$$\bar{P} == \operatorname{argmin}_P \sum_i d(P, P_i), \quad (3)$$

where \bar{P} is used to denote desirable *average* partition, K is a number of averaged partitions, $d(P_i, P_j)$ is a distance measure between two partitions and we want to minimize average distance from \bar{P} to all given partitions P_i . All partitions are represented by a vector of length M , where M is a number of clustered objects (vertices of a graph in our case). It contains values from 1 up to N , where N is a number of clusters (parcels). This task is generally NP complete [14], but there are many approximate algorithms. We use two approaches: Cluster-based Similarity Partitioning Algorithm (cspa) [11] and greedy algorithm from [2].

CSPA defines a similarity between data points based on co-occurrence in a same cluster across different partitions, and then partitions a graph induced by this similarity. Specifically, given multiple partitions P_1, \dots, P_K of a data points x_1, \dots, x_M . One can define similarity between points x_i, x_j as follow:

$$S(x_i, x_j) = \sum_{k=1}^K \delta(P_k(x_i), P_k(x_j)), \quad (4)$$

Here δ is Kroneker delta. Thus $S(x_i, x_j)$ is just number of partitions in which points x_i and x_j were in the same cluster. Next we build a graph, with nodes correspond to data points and edge between node x_i and x_j is equal to $S(x_i, x_j)$. We then partition this graph into communities using some clustering algorithm and the resulting partition is our clustering consensus partition.

Another way to find such *average* clustering is to optimize loss function given by Eq. 3.

The authors of [2] propose a greedy approach (Hard Ensemble - HE). Given multiple partitions $P_1 \dots P_K$ it combines them iteratively, first it finds average of $\bar{P}_{1,2} = \min_{\bar{P}}(d(\bar{P}, P_1) + d(\bar{P}, P_2))$, next average of $\bar{P}_{1,2}$ and P_3 and so on. As a measure of distance the authors take the average square distance between membership functions:

$$d(P_i, P_j) = \frac{1}{N} \sum_{k=1 \dots N} \|p_i^k - p_j^k\|^2, \quad (5)$$

Exclusively for this definition we use another way to encode object's memberships: P_i is a matrix of size $M \times N$ (number of objects times number of clusters)

$$P_i^{m,n} = \begin{cases} 1 & \text{if } m\text{'th object belongs to } n\text{'th cluster} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

In Eq. 5 p_i^k and p_j^k are k^{th} rows of memberships matrices P_i and P_j respectively. They correspond to membership vector of the k^{th} object. Since we are looking for disjoint clusters, only a single element of such row vector is equal to 1. This representation is defined up to any column permutation π of matrix P , thus the optimization procedure is done subject to all possible column permutations.

2.4 Comparison Metrics

Once we find individual partitions and combine them into an average partition, we want to access their quality. We use two different approaches.

First, we compare representation strength of different parcellations by measuring distance between original $\lambda(x, y)$ and its piece-wise approximation given by:

$$\gamma(x, y) = \frac{1}{|E_i||E_j|} \mathcal{C}(E_i, E_j), \quad (7)$$

where $x \in E_i$ and $y \in E_j$. Natural way to compare two statistical distributions is to measure distance between their probability density functions, we will use Kullback-Leibler divergence [4]. For two probability distributions with densities $\lambda(x)$ and $\gamma(x)$ the KL divergence is:

$$KL(\lambda, \gamma) = \int_{-\infty}^{\infty} \lambda(x) \log \frac{\lambda(x)}{\gamma(x)} dx, \quad (8)$$

It takes values close to 0 if two distributions are equal almost everywhere. Similar but symmetrized version of KL divergence is Jensen-Shannon divergence [6]. Again for two probability distributions with densities $\lambda(x)$ and $\gamma(x)$ it is given by:

$$JS(\lambda, \gamma) = \frac{1}{2}(KL(\lambda, r) + KL(\gamma, r)), \quad (9)$$

where $r(x) = \frac{1}{2}(\lambda(x) + \gamma(x))$.

Second, we compare performance of different parcellations on a gender classification task. We use Logistic Regression model with (small) l_1 regularization on a vectors of edge weights (the upper triangle of adjacency matrix excluding diagonal). Classification performance is measured in terms of ROC AUC score, which is typical for binary classification tasks.

Finally, in order to quantify goodness of consensus clustering and access hemisphere symmetry we use Adjusted Mutual Information [15]. It measures similarity between two partitions, with value 1 corresponds to identical partitions and values close to zero for partitions that are very different. Given set X of n elements, $X = \{x_1, x_2, \dots, x_n\}$ let us consider two partitions of X : $U = \{U_1, U_2, \dots, U_l\}$ and $V = \{V_1, V_2, \dots, V_k\}$. These partitions are *strict* (or *hard*):

$$\bigcap_{j=1}^k V_j = \bigcap_{i=1}^l U_i = \emptyset$$

and *complete*:

$$\bigcup_{j=1}^k V_j = \bigcup_{i=1}^l U_i = X$$

We can construct the following $l \times k$ contingency table:

U, V	V_1	V_2	\dots	V_k	$\sum_{j=1}^k s_{ij}$
U_1	s_{11}	s_{12}	\dots	s_{1k}	s_1
U_2	s_{21}	s_{22}	\dots	s_{2k}	s_2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
U_l	s_{l1}	s_{l2}	\dots	s_{lk}	s_l
$\sum_{i=1}^l s_{ij}$	s^1	s^2	\dots	s^k	

Here s_{ij} denotes a number of common objects between U_i and V_j :

$$s_{ij} = |U_i \cap V_j|$$

then Mutual Information is given by:

$$MI = \sum_{i=1}^l \sum_{j=1}^k P(i, j) \log \frac{P(i, j)}{P(i)P'(j)}, \tag{10}$$

where $P(i)$ is the probability of a random sample occurring in cluster U_i , $P'(j)$ is the probability of a random sample occurring in cluster V_j :

$$P(i) = \frac{s^i}{n}, \quad P'(j) = \frac{s_j}{n}$$

and $P(i, j)$ - probability of an object occurs in U_i and V_j simultaneously:

$$P(i, j) = \frac{s_{ij}}{n}$$

Adjustment scheme as proposed by Hubert and Arabie [3] has the following general form:

$$\text{Adjusted Index} = \frac{\text{Index} - \text{Expected Index}}{\text{Max Index} - \text{Expected Index}} \tag{11}$$

Using AMI we assess ensemble goodness (how good clustering ensemble algorithm combines multiple partitions) using modified 3:

$$\text{Ensemble goodness} = \sum_i^K \text{AMI}(\bar{P}, P_i), \tag{12}$$

We compute parcellation symmetry by comparing hemisphere parcels (labels):

$$\text{Symmetry} = \text{AMI}(\bar{P}_{\text{LH}}, \bar{P}_{\text{RH}}). \tag{13}$$

3 Experiments

3.1 Data Description

We use construct continuous connectomes of 400 subjects from the Human Connectome Project S900 release [13] following [8]. We use an icosahedral spherical sampling, at a resolution of 10242 mesh vertices per hemisphere. We used Dipy’s implementation of constrained spherical deconvolution (CSD) to perform probabilistic tractography. Prior to clustering, we exclude all mesh vertices that were labeled by FreeSurfer as corpus callosum or cerebellum.

3.2 Experimental Pipeline

Our experiments are summarized as follows:

1. For each subject we reconstruct its Continuous Connectome.
2. For each Continuous Connectome we iteratively run Louvain clustering algorithm, as described above. Subgraphs of having less than 1% of original graph vertices were not divided.
3. Next we aggregate individual subject partitions and obtain consensus clustering. Aggregation was done over 400 HCP subjects. Further, after finding the optimal parcellation, we obtain two parcellations based on two disjoint sets of 200 HCP subjects in order to compute reproducibility.
4. We aggregate partitions of the same level (I-II-III) using CSPA and HE.
5. We compare obtained partitions between themselves and with FreeSurfer’s Desikan-Killiani parcellation using Kullback-Leibler and Jensen-Shannon divergence. We compute goodness of an ensemble and parcellation symmetry using AMI.
6. We compare performance of simplified connectomes on a binary classification task using Logistic Regression with l_1 penalty. Classification results are measured in terms of ROC AUC score, with averaging over 10 cross-validation folds.

3.3 Results

Table 1 represent all comparison results. First we can see that CSPA algorithm failed to find good clustering ensemble which result in poor classification performance and high KL and JS divergences. Greedy algorithm performed on P^{III} on the other hand outperforms standard Desikan atlas across all comparison metrics (except number of parcels, 68 versus 83). Surprisingly, greedy ensemble of second level partition (P^{II}) performs comparatively with Desikan, despite having twice as lower number of parcels (30 versus 68).

Another interesting property that we get automatically is parcellation symmetry. Our clustering algorithm known nothing about brain topology (all information was contained in graph connectivity), still reconstruct parcellations which

Table 1. All results are rounded to 2 significant digits. Where it possible results are reported with standard deviation. Best result in each row is colored. KL, JS divergences, **lower is better**; binary Gender Classification was measured in terms of ROC AUC score, higher - better; Ensemble goodness and Hemisphere symmetry were measured using AMI, Ensemble goodness is an average AMI between consensus partition and all individual partitions, higher - better.

	cspa P^I	cspa P^{II}	cspa P^{III}	HE P^I	HE P^{II}	HE P^{III}	DKT
KL	1.22 ± .07	1.18 ± .07	1.15 ± .07	1.16 ± .07	.86 ± .05	.66 ± .04	.83 ± .05
JS	.20 ± .00	.19 ± .00	.19 ± .00	.20 ± .00	.17 ± .00	.14 ± .00	.16 ± .00
Gender Classification	.63 ± .04	.64 ± .04	.69 ± .03	.64 ± .03	.75 ± .03	.86 ± .02	.81 ± .03
Hemisphere symmetry	.15	.24	.32	.26	.55	.66	.64
Ensemble goodness	.47 ± .06	.40 ± .02	.35 ± .00	.53 ± .05	.64 ± .02	.70 ± .01	–
Number of ROIs	5	7	8	7	30	83	68

are highly symmetrical. For standard Desikan atlas hemisphere symmetry is 0.64, and for our best parcellation this value even higher (0.66), and still remains quite high for second level partition (0.55).

Finally we check if our best ensemble parcellation, which combines 400 individual partitions is stable. We split 400 subjects into 2 groups of 200 subjects and independently combine their partitions. We compare resulting parcellations: $\bar{P}_{1,200}$ and $\bar{P}_{201,400}$ between themselves and with original \bar{P} (which is an ensemble of all 400 subjects) again using Adjusted Mutual Information. Both $\bar{P}_{1,200}$ and

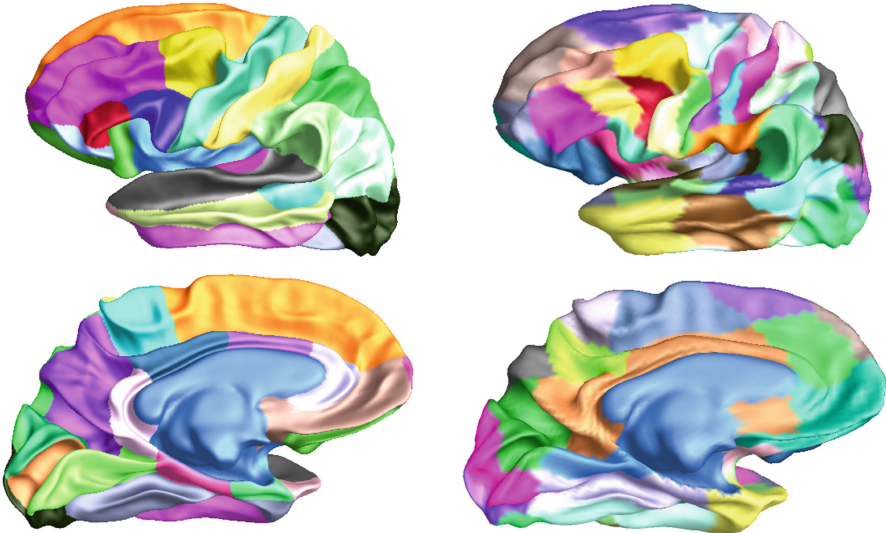


Fig. 2. Left column: Desikan-Killiany parcellation. Right column: HE P^{III} parcellation. Lateral and Medial views, left hemisphere.

$\bar{P}_{201,400}$ shows AMI value greater than 0.80 (0.83 and 0.82 respectively) when compare with \bar{P} , they also highly similar between themselves (Fig. 2).

4 Conclusion

We have presented an approach for generating unified connectivity-based human brain atlases bases on consensus clustering. The method is based on finding a pseudo average over the set of individual partitions. Our approach outperforms standard a anatomical parcellation on several important metrics, including agreement with dense connectomes, improved relevance to biological data, and even improved symmetry. Because our approach is entirely data driven an requires no agreement between individual parcellation labels, it combines both the flexibility of individual parcellations and the interpretability of simple unified atlases.

Acknowledgements. This work was funded in part by the Russian Science Foundation grant 17-11-01390 at IITP RAS.

References

1. Blondel, V.D., et al.: Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**(10), P10008 (2008)
2. Dimitriadou, E., Weingessel, A., Hornik, K.: A combination scheme for fuzzy clustering. *Int. J. Pattern Recogn. Artif. Intell.* **16**(07), 901–912 (2002)
3. Hubert, L., Arabie, P.: Comparing partitions. *J. Classif.* **2**(1), 193–218 (1985)
4. Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951)
5. Kurmukov, A., et al.: Classifying phenotypes based on the community structure of human brain networks. In: Cardoso, M.J. (ed.) *GRAIL/MFCA/MICGen 2017*. LNCS, vol. 10551, pp. 3–11. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67675-3_1
6. Lin, J.: Divergence measures based on Shannon entropy. *IEEE Trans. Inf. Theory* **37**(14), 145–151 (1991)
7. Meunier, D., Lambiotte, R., Bullmore, E.T.: Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* **4**, 200 (2010)
8. Moyer, D., et al.: Continuous representations of brain connectivity using spatial point processes. *Med. Image Anal.* **41**, 32–39 (2017)
9. Nicolini, C., Bordier, C., Bifone, A.: Community detection in weighted brain connectivity networks beyond the resolution limit. *Neuroimage* **146**, 28–39 (2017)
10. Parisot, S., Glocker, B., Schirmer, M.D., Rueckert, D.: GraMPa: graph-based multimodal parcellation of the cortex using fusion moves. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9900, pp. 148–156. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46720-7_18
11. Strehl, A., Ghosh, J.: Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* **3**, 583–617 (2002)
12. Taylor, P.N., Wang, Y., Kaiser, M.: Within brain area tractography suggests local modularity using high resolution connectomics. *Sci. Rep.* **7**, 39859 (2017)
13. Van Essen, D.C., et al.: The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013)

14. Vega-Pons, S., Ruiz-Shulcloper, J.: A survey of clustering ensemble algorithms. *Int. J. Pattern Recogn. Artif. Intell.* **25**(03), 337–372 (2011)
15. Vinh, N.X., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837–2854 (2010)
16. Petrov, D., et al.: Evaluating 35 methods to generate structural connectomes using pairwise classification. *arXiv e-prints*, eprint = 1706.06031, June 2017
17. Arslan, S., Ktena, S.I., Makropoulos, A., Robinson, E.C., Rueckert, D., Parisot, S.: Human brain mapping: a systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage* **170**, 5–30 (2018)