

Article

Application of Multi Network Alignment Algorithms for Connectomes Study

Marianna Milano ^{1,‡}, Pietro Hiram Guzzi ^{1,‡*} and Mario Cannataro¹

¹ Department of Surgical and Medical Sciences, University of Catanzaro; m.milano,hguzzi,cannataro@unicz.it

* Correspondence: hguzzi@unicz.it

‡ These authors contributed equally to this work.

Abstract: A growing area in neurosciences is focused on the modeling and analysis the complex system of connections in neural systems, i.e. the connectome. Here we focus on the representation of connectomes by using graph theory formalisms. The human brain connectomes are usually derived from neuroimages; the analyzed brains are co-registered in the image domain and brought to a common anatomical space. An atlas is then applied in order to define anatomically meaningful regions that will serve as the nodes of the network - this process is referred to as parcellation. Recently, it has been proposed to perform atlas-free random brain parcellation into nodes and align brains in the network space instead of the anatomical image space to define network nodes of individual brain networks. In the network domain, the question of comparison of the structure of networks arises. Such question is tackled by modeling the comparison of brain network as a network alignment (NA) problem. In this paper, we first defined the NA problem formally, then we applied three existing state of the art of multiple alignment algorithms (MNA) on diffusion MRI-derived brain networks and we compared the performances. The results confirm that MNA algorithms may be applied in cases of atlas-free parcellation for a fully network-driven comparison of connectomes.

Keywords: graph alignment; brain network; human connectome

1. Introduction

The human brain is a complex organ organized into a dense system of connections, also known as connectome. Recent studies have shown that this system of connections is responsible for the brain activity, and an alteration of connections (decreased or increased connectivity) can led to the insurgence of neurological diseases [1,2]. For this reasons, many researches in neuroscience have been focused on mapping and analysis of the human brain connectome [3]. Connectome may be analyzed using different zoom, e.g. by focusing on single components, i.e. neurons and axons, or grouping them into regions. Usually the analysis of single components is defined to as anatomic connectivity, while the analysis of regions is called functional connectivity because regions are in general performing different functions. Typically, the human brain connectome can be mapped using neuroimaging techniques, such as Magnetic Resonance Imaging (MRI), Electroencephalography (EEG), and Electromyography (MEG) enabling to take a picture of the brain connections of patients [4]. Among the others, the main source for deriving information about connectomes is Magnetic Resonance Imaging (MRI) [5] able to achieve both informations about anatomic connectivity and functional connectivity.

Once obtained, the connectome data need to be characterized through sophisticated analytic strategies. A useful strategy is based on graph theory [6], that ensures the modeling of such data into a network model. Different studies have used the network models to extract clinically relevant information [7,8] due to the capability to summarize the characteristics of a complex network with few measures and to understand the organization of both entire networks and individual network elements [6].

36 Graph theoretical approaches model the human brain as set of nodes linked by edges. The
37 nodes typically represent region of interest (ROI) and the edges represent functional or anatomical
38 connections. A typical MRI experiment produces a series of images, either from intra-subject or
39 inter-subject, then the MRI images are modeled as networks. A further step consists of the coregistration
40 among the network and a brain atlas in order to obtain anatomically meaningful regions [9]. Recently,
41 Tymofiyeva et al. [10] proposed an alternative method based on the application of atlas-free parcellation
42 and on the construction of individual connectomes only in the network space. In the network domain,
43 an appropriate analytic strategy consists of the comparison of studied networks by recurring to
44 network alignment (NA) approaches. The techniques for the alignment of biological networks fall into
45 sub-categories: local alignment, to find small conserved motifs across networks, and global alignment,
46 which attempts to find a best mapping between all nodes of the two networks or pairwise alignment
47 to align two networks and multiple alignment that align multiple networks. Different studies have
48 widely used the NA approaches for the analysis of biological networks. In previous works [11,12] we
49 explored the possibility to apply the NA methods for the analysis of to MRI connectomics. At first we
50 tested different global alignment algorithms to build the alignments among the diffusion MRI-derived
51 brain networks. Then, we analyzed the alignment results in term of topological quality measures
52 and according to these analyses, we identified the best alignment algorithm to align the diffusion
53 MRI-derived brain networks. However, recent studies have demonstrated in an independent way
54 that the multiple alignment algorithms are able to exact deeper information than pairwise alignment
55 algorithm when these one are applied to molecular biology analysis [13].

56 According to these studies, we choose to apply multiple alignment algorithms on MRI
57 connectomics. Here, we selected three existing state of the art multiple alignment algorithms to
58 build the alignment of diffusion MRI-derived brain networks. The algorithms tested here are
59 MultiMAGNA++ [13], GEDEVO-M [14] and IsoRankN [15]. The algorithms are applied to build
60 the multiple alignment among the diffusion MRI-derived brain networks. After the alignments were
61 built, we compared the performance of these algorithms.

62 2. Methodology for Brain Network Analysis

63 In this section, we present the workflow of analysis that can be preformed on the brain network
64 starting from the building of connectome from MRI images. The Figure 1 shows a workflow of
65 Methodology for the Brain Network Analysis from the building of representative network from
66 experimental data to the comparison of brain network applying multiple alignment algorithm.

67 2.1. Building a Brain Network

68 The building of brain network starts with a set of anatomical or physiological observations [16],
69 then the structural and functional connectivity data are processed into network model exploiting graph
70 theory.

71 However, the application of graph theory to the study of connectomes presents some challenges
72 related to the description of nodes and edges [17]. For example, an ideal definition of nodes should
73 group a set of neurons according to maximal functional homogeneity within nodes and the maximal
74 functional heterogeneity among different nodes. According to this, there is no clear evidence for the
75 optimal definition of both nodes and edges. A common approach to define the nodes of a brain network
76 consists of the subdivision of the brain into homogeneous, non-overlapping and large-scale, regions
77 respect to information provided, generally, by techniques based on magnetic resonance imaging (MRI)
78 [18], also known as “parcellation process”. Especially, MRI has allowed to obtain information about
79 anatomical connectivity, functional connectivity, or task-related activation.

80 Currently, there exist three different approaches applied to parcellation of connectome:

- 81 1. Parcellation of the brain by using **predefined anatomical templates** that consists of the
82 registration of the structural images from MRI to anatomical atlas based on the Brodmann

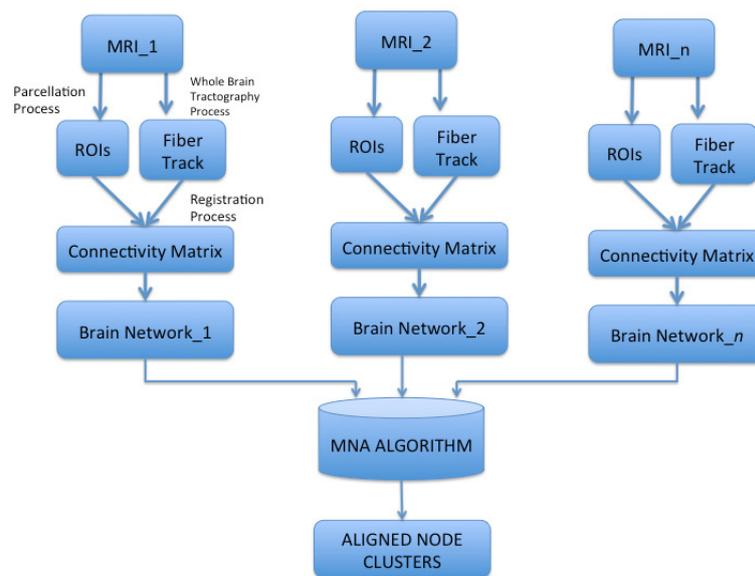


Figure 1. Building a representative network from experimental data: example of a workflow. Diffusion or functional MRI images are acquired for a subject according to the study to be conducted. The MRIs are used to perform whole-brain parcellation by selecting a suitable method. Starting from the parcellated whole brain the computation of connections is performed the connectivity matrix is constructed. Then, the resulting brain network is obtained. This process is performed for each studied subject. A MNA algorithm takes as input the brain networks and produces aligned node clusters between more than two networks.

- 83 areas [19]. This approach enables to subdivide the whole brain into labeled regions according to
 84 the different labels regions of the templates;
 85 2. Parcellation of the brain by using **randomly generated templates** [20] that ensures to divide the
 86 whole brain into parcels (brain region) of roughly equal size;
 87 3. **Connectivity-based parcellations** that aim to delineate brain regions according to the similarities
 88 in structural or functional connectivity patterns.

89 Due to the different approach, the choice of a parcellation scheme is fundamental for subsequent
 90 analysis on brain network. In fact, each parcellation method presents some pitfalls.

91 For example, the parcellation of the brain by using predefined anatomical template raises the
 92 question of the accuracy of mapping. Since atlas based on the Brodmann areas are originally defined
 93 using cytoarchitectural differences between brain regions, in the registration step a mismatch among
 94 the cortical surface analyzed and the borders of the Brodmann areas may occur [10]. Thus, this
 95 approach is limited by inter-subject variability and can be especially problematic in the context of brain
 96 maturation. In this paper, we focus on the random, atlas-free definition of nodes in individual subjects
 97 (see [12] for a deep description), which can allow for a fully network-driven way of looking at the
 98 brain and comparing brains of different subjects and, potentially, species [10].

99 The definition of the edges is also currently an open challenge related to a) the type of connectivity
 100 measured, and b) the method used to quantify it. As mentioned above, brain connectivity can refer
 101 to different aspects of brain organization including (i) *anatomical connectivity* consisting of axonal
 102 fibers connecting cortical and subcortical regions inferred from diffusion imaging, and (ii) *functional*
 103 *connectivity* defined as the observed statistical correlations of the BOLD signal between brain regions.

104 Once the nodes and the edges are defined, the pattern of connections between brain regions
 105 (nodes) can be stored into the Connectivity Matrix [21]. The Connectivity Matrix is symmetric matrix
 106 where rows and columns represent different brain regions, and the entries correspond to connection

107 (edge) between the regions. This representation lends itself to be mapped to a graphical model which
108 ensures to quantify different topological aspects of the connectome.

109 2.2. Comparison of Brain Network: Network Alignment

110 A crucial point in the connectome analysis regards the comparison of the brain networks. Thus,
111 the detection of an correct node mapping between atlas-free networks may uncover significant aspects
112 on the comparison of brains or structure of groups of subjects, such as healthy versus diseased subjects.
113 Many different network alignment methods have been proposed in biological fields [22].

114 Formally, a graph G is defined as $G = \{V, E\}$, where V is a finite set of nodes and E is a finite set of
115 edges. Let $G_1 = \{V_1, E_1\}$ and $G_2 = \{V_2, E_2\}$ be two graphs, where $V_{1,2}$ are sets of nodes and $E_{1,2}$ are sets
116 of edges, a graph alignment is the mapping between the nodes of the input networks that maximizes
117 the similarity between mapped entities. From a theoretical point of view, the graph alignment problem
118 consists of finding an alignment function (or a mapping) $f : V_1 \rightarrow V_2$ that maximizes a cost function
119 Q . The similarity between the graphs is defined by a cost function, $Q(G_1, G_2, f)$, also known as the
120 quality of the alignment. Let f be an alignment between two graphs G_1 and G_2 , given a node u from
121 G_1 , $f(u)$ is the set of nodes from G_2 that are aligned under f to u . Q expresses the similarity among
122 two input graphs with respect to a specific alignment f and the formulation of Q strongly influences
123 the mapping strategy.

124 There exist different formulations of Q that fall into following the classes:

125 **Topological Similarity:** Graphs are aligned by considering only edge topology, so that the perfect
126 alignment is reached when input graphs are isomorphic.

127 **Node Similarity:** Such function considers the similarity among mapped nodes. Nodes of the
128 aligned graphs can be more or less similar to each other. Thus the alignment should align each node of
129 one graph to the most similar node of the other one given a node similarity functions, $s(v_1, v_2) \rightarrow R$,
130 $v_1 \in V_1, v_2 \in V_2$.

131 **Hybrid approaches:** Some recent formulations of Q take into account of both of the approaches
132 by linear combination.

133 The network alignment problem can be formulated according to: i) the kind of input, *pairwise or*
134 *multiple alignment* and ii) the scope of node mapping required, *local or global alignment*. In general, the
135 network alignment can be classified as local alignment or global alignment. The *local alignment* typically
136 finds multiple and unrelated regions of isomorphism among the input networks, each region implying
137 a mapping independently of the others. Therefore, the computed correspondences may involve
138 overlapping subgraphs. The output of local network alignment is a set of pairs of possibly overlapping
139 subgraphs of the input networks. The literature contains many algorithms that address local graph
140 alignment problem. For example, AlignNemo [23] and AlignMCL [24] algorithms. The *global alignment*
141 aims to find a mapping that should cover all of the nodes of the input networks. Global alignment
142 returns a unique overall alignment between the input networks, such that a one-to-one correspondence
143 is found between of a network with one node of the other network. Most popular existing methods of
144 global alignment are MAGNA [25], NETAL [26], GHOST [27], WAVE [28]. For a complete review on
145 global and local network alignment algorithms and their advantages or disadvantages see [29].

146 Also, the network alignment methods can be *pairwise or multiple alignment*.

147 The pairwise network alignment (PNA) aligns two networks at a time and produces aligned node
148 pairs between two networks. The multiple network alignment (MNA) aligns three or more networks
149 to each other at once and produces aligned node clusters. PNA and MNA can be local or global, with
150 one-to-one or many-to-many node mappings. The difference between one-to-one and many-to-many
151 mapping in the pairwise alignment refers the previous discussion on global and local alignment. The
152 PNA can search the similar small subnetworks exploiting many-to-many mapping between nodes of
153 the compared network or can look for the best overlap of the whole compared networks exploiting
154 one-to-one node mapping. MNAs are one-to-one MNA methods when produce an aligned cluster

155 containing at most one node per network, whereas MNAs are many-to-many MNA methods when an
156 aligned cluster contains more than one node from a single network.

157 In literature, both PNA and MNA are applied to build the alignment protein interaction networks
158 (PINs) [30]. Since, MNA can capture functional knowledge that is common to multiple species,
159 it was detected that MNA leads to deeper biological information than PNA. However, MNA is
160 computationally much harder than PNA because the complexity of the network alignment problem
161 increases exponentially with the number of analyzed networks.

162 There exist different proposed multiple network alignment algorithms in literature such as
163 MultiMAGNA++ [13], GEDEVO-M [14] and IsoRankN [15].

164 In this work, three multiple alignment algorithms were chosen to build the multiple alignment of
165 brain networks. We give hereafter a short conceptual description.

166 A popular existing method of multiple alignment is MultiMAGNA++ [13]. MultiMAGNA++ is a
167 a global one-to-one MNA aligner that simulates a population of alignments that evolves over time
168 by applying a genetic algorithm and a function for the crossover of two alignments into a superior
169 alignment. Since the genetic algorithm simulates the evolutionary process guided by the survival
170 of the fittest principle, only alignments, i.e. those that conserve the most edges, survive. Thus,
171 MultiMAGNA++ proceeds to the next generation, until the alignment accuracy cannot be optimized
172 further.

173 The second multiple aligner is GEDEVO-M [14] a global one-to-one MNA aligner. GEDEVO-M is
174 an extension of GEDEVO [31] tool for efficient global graph alignment. Underlying the GEDEVO-M
175 method is the Graph Edit Distance model (GED), where a graph is transferred into another one with a
176 minimal number of edge insertions and deletions. Thus, GEDEVO-M uses the GED as optimization
177 model for finding the best alignments and then minimizes the sum of GEDs between every pair of
178 input networks.

179 The last multiple aligner is IsoRankN [15], a global many-to-many MNA alignment tool based a
180 spectral clustering method to find dense and clique modules when the global alignment of multiple
181 networks is computed.

182 3. Results

183 3.1. Dataset

184 The dataset consisted of 24 diffusion MRI-derived structural networks of human brain: 12
185 networks with a number of nodes equal to 95 and the 12 networks with a number of nodes equal to
186 1000. The brain networks are related to three different stages of development by including newborns
187 (NE), six-month-old infants (6M), and adults (AD). See *Materials and Methods Section* for a complete
188 description.

189 3.2. Building of brain network multiple alignment

190 We built the multiple alignment of all networks with 95 and 1000 nodes (for convenience we call
191 the two dataset *networks*₉₅ and *networks*₁₀₀₀) related to same growth stages (NE, 6M, AD) by applying
192 MultiMAGNA++ [13], GEDEVO-M [14] and IsoRankN [15].

193 We ran all MNA methods on the same Linux machine with Intel Core i5 and 4GB of RAM. We
194 selected the following MultiMAGNA++ parameters: *CIQ* as measure of Edge Conservation, the α
195 parameter equal to 0, in order to consider only topology, whereas the population size, number of
196 generation, fraction of elite members were set to default values. We tested different parameters and
197 obtained best results with the default parameters for GEDEVO-M: *pop* parameter that controls the
198 number of new individuals per iteration set equal to 1000 and *maxsame* that controls the stop after N
199 iterations without any score improvement were equal to 3000. To build the multiple alignment with
200 IsoRankN we set: the max number of iterations *K* equals to 30, the threshold *thresh* equals to $1e - 4$,
201 *maxveclen* equals to 1000000 and the α parameter equal to 1 in order to consider only network data.

202 The Table 1 reports the execution time of MultiMAGNA++, GEDEVO-M and IsoRankN to build the
 203 multiple alignment on the networks with 95 nodes and on the networks with 1000 nodes.

Table 1. Execution Time to build the multiple alignment with MultiMAGNA++, GEDEVO-M and IsoRankN for the networks with 95 nodes and the networks with 1000 nodes

	Execution Time for network with 95 nodes	Execution Time for network with 1000 nodes	Processor	Memory
MultiMAGNA++	5 seconds	7 seconds	Intel Core i5	4 GB
GEDEVO-M	8 seconds	11 seconds	Intel Core i5	4GB
IsoRankN	6 seconds	10 seconds	Intel Core i5	4 GB

204 3.3. Topological alignment quality evaluation

205 Here, we aim to evaluate the quality of the multiple alignments built with MultiMAGNA++,
 206 GEDEVO-M and IsoRankN algorithms. The topological quality is related to two alignment algorithm
 207 capability as the reconstruction of the true node mapping and the conservation of as much as possible
 208 edges. Typically, the Node Correctness (NC) is the measure widely used to evaluate how an alignment
 209 reconstructs the true node mapping correctly. Instead, different measures are used to evaluate how
 210 well the edges are conserved on an alignment, such as EC, ICS or S^3 (see the previous Section). In
 211 general, the Edge Correctness is defined as the number of edges conserved under an alignment f with
 212 respect to the total number of edges of input networks. Thus, once the multiple alignments were built,
 213 we performed an evaluation of alignment quality by comparing the Edge Correctness (EC) [25] related
 214 to the alignments built with MultiMAGNA++, GEDEVO-M and IsoRankN.

215 The Table 2 and Table 3 report the global Edge Correctness computed on the multiple alignment
 216 of all networks with 95 nodes and with 1000 nodes related to same growth stages NE, 6M, AD by
 217 applying MultiMAGNA++, GEDEVO-M and IsoRankN algorithms.

Table 2. Comparison the Edge Correctness of the multiple alignments built with MultiMAGNA++, GEDEVO-M and IsoRankN.

Edge Correctness	NE	6M	AD
MultiMAGNA++	0.5	0.55	0.49
GEDEVO-M	0.441	0.441	0.48
IsoRankN	0.477	0.477	0.485

Table 3. Comparison the Edge Correctness of the multiple alignments built with MultiMAGNA++, GEDEVO-M and IsoRankN.

Edge Correctness	NE	6M	AD
MultiMAGNA++	0.14	0.16	0.19
GEDEVO-M	0.089	0.091	0.099
IsoRankN	0.095	0.099	0.1

218 Figure 2 shows an overview of edge conservation comparison on $networks_{95}$ whereas Figure
 219 3 shows an overview of edge conservation comparison on $networks_{1000}$. We note that the best
 220 results in terms of edge conservation were obtained when applying MultiMAGNA++ as global
 221 aligner both on $networks_{95}$ and $networks_{1000}$. In fact, the mean edge correctness values are higher
 222 on the alignments built with MultiMAGNA++ than mean edge correctness scores on alignments
 223 obtained with GEDEVO-M and IsoRankN. The reason is related to the strategy of MultiMAGNA++ to
 224 construct the multiple alignment. In fact, MultiMAGNA++ is the unique multiple aligner that directly
 225 optimizes edge conservation in addition to node conservation by using a genetic algorithm, whereas
 226 the MNA algorithms optimize node conservation only. In this way, the quality of alignment built with

227 MultiMAGNA++ results improved. This entails an inferior behavior of GEDEVO-M and IsoRankN
 228 compared to the MultiMAGNA++.

229 We also note that values of EC for *networks*₉₅ are higher than EC for *networks*₁₀₀₀.

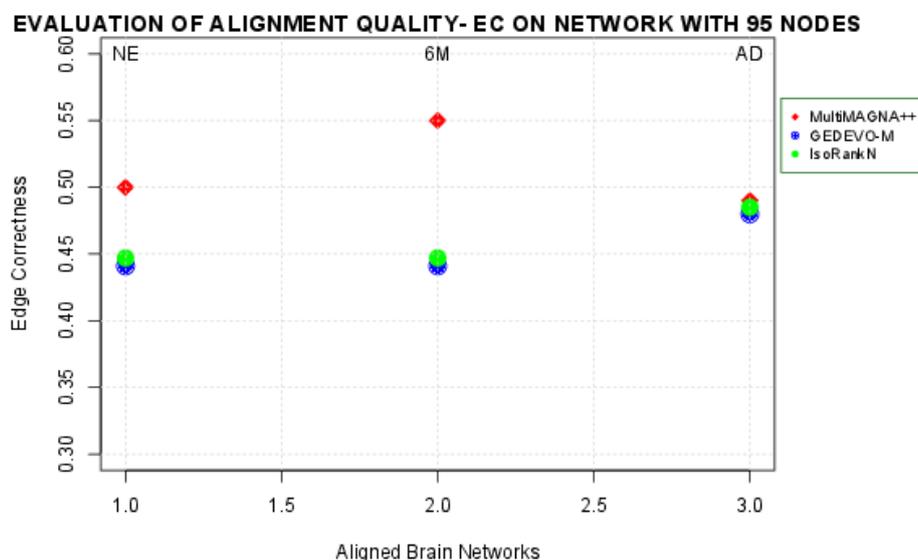


Figure 2. The topological evaluation of alignments built with MultiMAGNA++ (red marker), GEDEVO-M (blue marker), IsoRankN (green marker). The Figure shows the mean Edge Correctness scores of alignments built among the networks with 95 nodes by applying the selected three multiple aligners.

230 4. Discussion

231 The brain connectivity refers to different aspects of brain organization including i) anatomical
 232 connectivity consisting of axonal fibers across cortical regions and ii) functional connectivity defined
 233 as the observed statistical correlations of the BOLD signal between regions of interest. Understanding
 234 brain connectivity can shed light on the brain cognitive functioning that occurs via the connections
 235 and interaction between neurons. Brain connectivity can be modeled and quantified with a large
 236 number of techniques. A useful formalism to represent the brain connectivity derives from graph
 237 theory. The graph theoretical modeling of the human connectome has enabled important discoveries
 238 by comparing the brain networks of studied subjects. In this study we proposed to apply three multiple
 239 alignment algorithms MultiMAGNA++, GEDEVO-M and IsoRankN to align atlas-free human brain
 240 networks at three developmental stages. We decided to apply MNA algorithms to the study of brain
 241 networks because, in previous studies conducted on PINs, MNA were able to lead to deeper biological
 242 compared to PNA, by capturing conserved network regions between multiple networks. We analyzed
 243 the multiple alignment results in term of topological quality measures, by comparing the EC related to
 244 each alignment. According to these analyses, MultiMAGNA++ resulted in the best alignment. The
 245 reason is related to the strategy underlying MultiMAGNA++ to construct the multiple alignment
 246 by optimizing simultaneously the edge conservation and node conservation. Our ongoing study is
 247 focused on the implementation of an ad hoc algorithm for connectome alignment. Since there are
 248 many conditions in which the classical parcellation is not useful, we retain that this seminal work may
 249 open the way for the use of multiple network alignment in atlas-free parcellation.

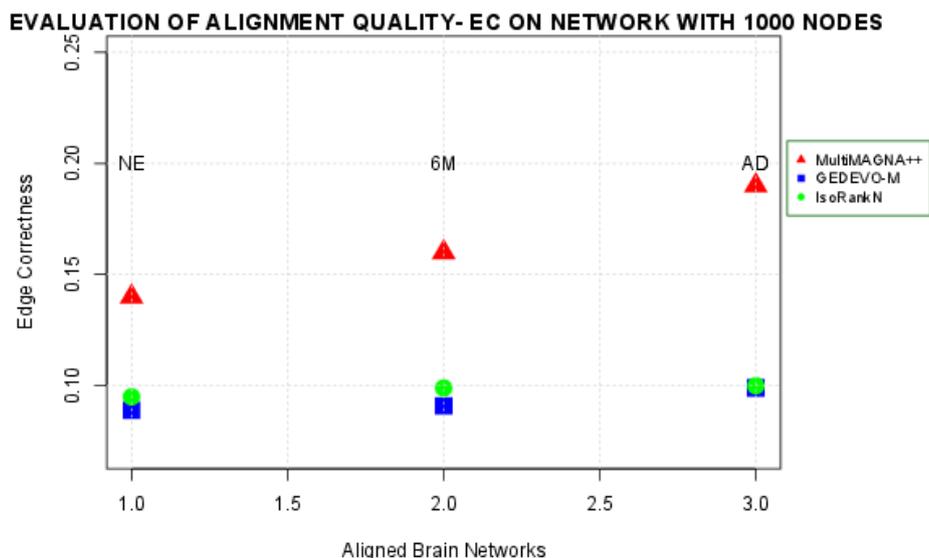


Figure 3. The topological evaluation of alignments built with MultiMAGNA++ (red marker), GEDEVO-M (blue marker), IsoRankN (green marker). The Figure shows the mean Edge Correctness scores of alignments built among the networks with 1000 nodes by applying the selected three multiple aligners.

250 5. Materials and Methods

251 5.1. Dataset

252 The dataset consisted of diffusion MRI-derived structural networks of human brain at different
 253 stages of development, starting with neonates [10]. Acquisition of the MRI data was compliant with
 254 the Health Insurance Portability and Accountability Act (HIPAA) and the study was approved by
 255 the Committee on Human Research (CHR) of the University of California, San Francisco. Three age
 256 groups were included: 4 neonates imaged in the first 4-5 days of life (NE), 4 six-month-old infants
 257 (6M), and 4 adults (age 24-31 years) (AD). The two pediatric groups had transient encephalopathy
 258 at birth, but none of the patients had clinical or imaging evidence of brain injury. The subjects were
 259 scanned on a 3T GE MR scanner using a spin echo (SE) echo planar imaging (EPI) diffusion tensor
 260 imaging DTI sequence with parameter described in [10]. Tensor calculation, tractography, cortical
 261 parcellation into N equal-area nodes (Figure ??), and construction of the connectivity matrices was
 262 performed as described previously [10]. All networks were binarized with a threshold of 1 streamline.
 263 Starting from the images we obtained two different datasets. The first dataset consist of 12 networks
 264 with number of nodes equal to 95 depending on parcellation step. For convenience we call this dataset
 265 *networks₉₅*. Table 4 shows the networks parameters. About the second dataset, the 12 networks were
 266 constructed by setting the number of equal-area nodes for the cortical parcellation equal to 1000. Since
 267 all cortical areas of the brain are connected, a fine parcellation should ensure the interconnectedness of
 268 the whole brain, leaving no nodes isolated. In [10] the authors demonstrated that the highest number
 269 of nodes at which this condition is fulfilled in equal to 95. For this reason, the networks of the second
 270 dataset showed the isolated nodes that were not computed in the construction of the connectivity
 271 matrices. For convenience we call this dataset *networks₁₀₀₀* even though the nodes number is different
 272 from 1000. Table 5 shows the network parameters.

Table 4. Details of brain networks with 95 nodes used for experiments

Network	Nodes	Edge
NE01	95	341
NE02	95	341
NE03	95	334
NE04	95	320
6M01	95	353
6M02	95	333
6M03	95	333
6M04	95	338
AD1	95	449
AD2	95	406
AD3	95	438
AD4	95	416

Table 5. Details of brain networks with 1000 nodes used for experiments

Network	Nodes	Edge
NE01	889	2555
NE02	904	2618
NE03	900	2585
NE04	899	2298
6M01	902	2458
6M02	849	2182
6M03	805	1928
6M04	851	2087
AD1	902	3146
AD2	869	2691
AD3	878	3262
AD4	853	2907

273 5.2. Alignment Algorithms

274 In this section we describe in detail the multiple alignment algorithms selected to align the
275 diffusion brain networks.

276 MultiMAGNA++ [13] is a global one-to-one MNA algorithm based on a genetic algorithm to
277 build an improved alignment. By simulating the evolutionary process, guided by the survival of
278 the fittest principle, the genetic algorithm directly optimizes both edge and node conservation while
279 the alignment is constructed. In details, MultiMAGNA++ uses the genetic algorithm to simulates a
280 population of alignments that evolves over time and then applies new function for the crossover of
281 parent alignments into a superior child alignment that allows for aligning multiple networks.

282 The genetic algorithm requires an initial population of a given number of members. In
283 MultiMAGNA++, the members of population are multiple alignments. A multiple network alignment
284 (MNA) of k networks, ordered in terms of the number of nodes from the smallest to the largest one,
285 is represented by using $k - 1$ permutations which are bijective mappings between pairs of networks
286 adjacent. The permutations are set of disjoint node clusters that cover nodes in the k networks. So
287 MNA can be defined as multi permutation. The members of a population crossover with each other to
288 produce new members. Only the fitted members are more likely to crossover. Thus, the child alignment
289 resulting from a crossover function reflects each parent. In MultiMAGNA++, the crossover function is
290 defined as the midpoint of the shortest path between two permutations. In this way, the child MNA
291 shares the characteristics of each of the two parent MNAs. To avoid the size of the population to grow
292 without bound, the size is kept constant across all generations, with only the fittest members surviving
293 from one generation to the next. The fitness function is a combined measure of edge conservation S_E
294 and node conservation S_N maximized as follow:

$$\alpha S_E + (1 - \alpha) S_N \quad (1)$$

295 where α controls the contribution of each node and edge conservation measures and takes the
296 values between 0 and 1.

The edge conservation measure used in MNA is Conserved Interaction Quality (CIQ). CIQ is a weighted sum of edge conservation between all pairs of aligned a and b clusters and is defined as:

$$S_E = CIQ = \frac{\sum_{a,b} |E_{a,b}| cs(a,b)}{\sum_{a,b} |E_{a,b}|} \quad (2)$$

297 where, $|E_{a,b}|$ is the number of edges that connect the clusters, and $cs(a,b)$ is edge conservation
298 between two clusters. Let $r(a,b)$ be the number of networks that the edges which connect the clusters
299 belong to and $s(a,b)$ be the number of networks that contain at least one node in both clusters, $cs(a,b)$
300 is equal to 0 if $r(a,b) \leq 1$, also $cs(a,b)$ is equal to $\frac{r(a,b)}{s(a,b)}$.

The node conservation measure for MNA refers to internal cluster quality, i.e, the nodes in each cluster should be highly similar to each other with respect to some node cost function.

$$S_N = \frac{1}{n} \sum_{i=1}^n \frac{1}{\binom{|a_i|}{2}} \sum_{(u,v) \in P(a_i)} s(u,v) \quad (3)$$

301 where $s(u,v)$ is the similarity between nodes u and v with respect to some node cost function, a_i
302 is a aligned clusters with $i = 1, \dots, n$, $|a_i|$ is the size of a_i and $P(a_i)$ is the set of all pairs of nodes in a_i .

303 The genetic algorithm produces newer generations until the alignment quality cannot be
304 optimized further.

305 GEDEVO-M [14] is a global one-to-one MNA algorithm based on an evolutionary algorithm that
306 uses the Graph Edit Distance (GED) as optimization model for finding the best alignments. The GED
307 is defined as the minimum insertions and deletions of edges required to transfer a graph into another
308 graph. GEDEVO-M applies the Graph Edit Distance to multiple graph models and considers the
309 alignment building as Topological Multiple one-to-one Network Alignment (TMNA). TMNA problem
310 aims to find a multiple mapping F on a set of graphs G , such that the multiple Graph Edit Distance
311 $mGED_F$ is minimal over all possible multiple mapping on G . By minimizing the $mGED_F$, the number
312 of edges that are aligned in multiple networks simultaneously is maximized. The GEDEVO-M builds
313 the alignment by generating an initial multiple mapping with random permutations. A one-to-one
314 MNA of G graphs consists of a set of disjoint clusters. Each cluster can be represented as a tuple.
315 Initially, GEDEVO-M fixes a threshold, defined as the average over all tuple scores and then it randomly
316 swaps the tuples that have scores higher than the threshold. The tuples with lower than the threshold
317 are also given a certain chance to be swapped. After that, GEDEVO-M uses a crossover operator to
318 construct a new multiple mapping from two or more parent individuals of the previous generation. At
319 first, GEDEVO-M computes the tuple scores for every possible subset of G . Then, GEDEVO-M iterates
320 over the corresponding tuple scores by starting with larger subsets of G and assigns some of these
321 tuples to a new multiple mapping until every subset is considered. Finally, GEDEVO-M evaluates the
322 quality of the multiple mapping by using the score S . The score S depend on the multiple Graph Edit
323 Distance ($mGED$) and Graphlet-degree signature distance (GSD) of multiple mapping that computes
324 the difference in neighboring topologies of potentially matched nodes.

325 IsoRankN [15] is a global many-to-many MNA alignment tool based a spectral partitioning
326 method to find dense and clique clusters on multiple-network alignment.

327 IsoRankN builds a multiple network alignment by local partitioning the graph of pairwise
328 functional similarity scores. Initially, IsoRankN computes the functional similarity scores of every pair
329 of nodes of k networks. In this way, a functional similarity graph, where each edge is weighted by its
330 functional similarity score, is obtained. Then, IsoRankN applies a star spread method on functional
331 similarity graph to obtain a multiple alignment as highly similar cliques. In detail, IsoRankN computes,

332 for each node, every neighbor connected with an edge whose weight is greater than a threshold; this
 333 represent the star of a node S .

334 Then, IsoRankN orders the nodes according to the total weight of the star S . For each the star S , a
 335 subset with highly weighted neighborhood is found. This subset represents a functionally conserved
 336 interaction cluster. Finally, IsoRankN performed a merging stars process, by looking at the neighbors
 337 of the neighbors of a node and by merging the stars of two nodes if every member of star related to a
 338 node 1 has the node 2 as a neighbor and vice versa. The process is repeated until all nodes are assigned
 339 to a cluster.

340 **Acknowledgments:** PHG,MM and MC have been partially supported by the following research project
 341 funded by the Italian Ministry of Education and Research (MIUR): BA2Know-Business Analytics to Know
 342 (PON03PE_00001_1). The authors wish to thank Olga Tymofiyeva for her suggestions to this research activity.

343 **Author Contributions:** PHG and MM conceived the main idea of the algorithm and designed the tests. MC
 344 supervised the design of the algorithm. PHG and MM designed the functional requirements of the software tool.
 345 All authors read and approved the final manuscript.

346 **Conflicts of Interest:** The authors declare no conflict of interest.

347 Abbreviations

348 The following abbreviations are used in this manuscript:

349	NA	Network Alignment
	PNA	Pairwise Network Alignment
	MNA	Multiple Network Alignment
	MRI	Magnetic Resonance Imaging
	ROIs	Region of Interest
	DTI	Diffusion Tensor Imaging
	PINs	Protein Interaction Networks
	BOLD	Blood Oxygenation Level Dependent
	NE	Newborns
	6M	Six-Month-Old
350	AD	Adults
	HIPAA	Health Insurance Portability and Accountability Act
	CHR	Committee on Human Research
	S^3	Symmetric Substructure Score
	EC	Edge Correctness
	S_E	Edge Conservation
	S_N	Node Conservation
	CIQ	Conserved Interaction Quality
	GED	Graph Edit Distance
	TMNA	Multiple one-to-one Network Alignment
	$mGED_F$	Multiple Graph Edit Distance

351 References

- 352 1. Kiani, R.; Cueva, C.J.; Reppas, J.B.; Peixoto, D.; Ryu, S.I.; Newsome, W.T. Natural grouping of neural
 353 responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* **2015**, *85*, 1359–1373.
- 354 2. Bargmann, C.I.; Marder, E. From the connectome to brain function. *Nature methods* **2013**, *10*, 483–490.
- 355 3. Sporns, O.; Tononi, G.; Kötter, R. The human connectome: a structural description of the human brain.
 356 *PLoS Comput Biol* **2005**, *1*, e42.
- 357 4. Xia, M.; He, Y. Magnetic resonance imaging and graph theoretical analysis of complex brain networks in
 358 neuropsychiatric disorders. *Brain connectivity* **2011**, *1*, 349–365.
- 359 5. Toga, A.W.; Clark, K.A.; Thompson, P.M.; Shattuck, D.W.; Van Horn, J.D. Mapping the human connectome.
 360 *Neurosurgery* **2012**, *71*, 1.

- 361 6. Bullmore, E.; Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional
362 systems. *Nature Reviews Neuroscience* **2009**, *10*, 186–198.
- 363 7. Cannataro, M.; Guzzi, P.H.; Veltri, P. Protein-to-protein interactions: Technologies, databases, and
364 algorithms. *ACM Computing Surveys (CSUR)* **2010**, *43*, 1.
- 365 8. Lesne, A. Complex Networks: from Graph Theory to Biology. *Letters in Mathematical Physics* **2006**,
366 *78*, 235–262.
- 367 9. Yap, P.T.; Wu, G.; Shen, D. Human Brain Connectomics: Networks, Techniques, and Applications [Life
368 Sciences]. *IEEE Signal Processing Magazine* **2010**, *27*, 131–134.
- 369 10. Tymofiyeva, O.; Ziv, E.; Barkovich, A.J.; Hess, C.P.; Xu, D. Brain without anatomy: construction and
370 comparison of fully network-driven structural MRI connectomes. *PloS one* **2014**, *9*, e96196.
- 371 11. Milano, M.; Tymofiyeva, O.; Xu, D.; Hess, C.; Cannataro, M.; Guzzi, P.H. Using Network Alignment for
372 Analysis of Connectomes: Experiences from a Clinical Dataset. Proceedings of the 7th ACM International
373 Conference on Bioinformatics, Computational Biology, and Health Informatics. ACM, 2016, pp. 649–656.
- 374 12. Milano, M.; Guzzi, P.H.; Tymofiyeva, O.; Xu, D.; Hess, C.; Veltri, P.; Cannataro, M. An extensive assessment
375 of network alignment algorithms for comparison of brain connectomes. *BMC bioinformatics* **2017**, *18*, 235.
- 376 13. Vijayan, V.; Milenkovic, T. Multiple network alignment via multiMAGNA++. *arXiv preprint*
377 *arXiv:1604.01740* **2016**.
- 378 14. Ibragimov, R.; Malek, M.; Baumbach, J.; Guo, J. Multiple graph edit distance: simultaneous topological
379 alignment of multiple protein-protein interaction networks with an evolutionary algorithm. Proceedings
380 of the 2014 Annual Conference on Genetic and Evolutionary Computation. ACM, 2014, pp. 277–284.
- 381 15. Liao, C.S.; Lu, K.; Baym, M.; Singh, R.; Berger, B. IsoRankN: spectral methods for global alignment of
382 multiple protein networks. *Bioinformatics* **2009**, *25*, i253–i258.
- 383 16. Sporns, O. *Networks of the Brain*; MIT press, 2010.
- 384 17. Fornito, A.; Zalesky, A.; Breakspear, M. Graph analysis of the human connectome: promise, progress, and
385 pitfalls. *Neuroimage* **2013**, *80*, 426–444.
- 386 18. Thirion, B.; Varoquaux, G.; Dohmatob, E.; Poline, J.B. Which fMRI clustering gives good brain parcellations?
387 *Frontiers in neuroscience* **2014**, *8*, 167.
- 388 19. Geyer, S.; Weiss, M.; Reimann, K.; Lohmann, G.; Turner, R. Microstructural parcellation of the human
389 cerebral cortex—from Brodmann’s post-mortem map to in vivo mapping with high-field magnetic resonance
390 imaging. *Frontiers in human neuroscience* **2011**, *5*, 19.
- 391 20. Fornito, A.; Zalesky, A.; Bullmore, E.T. Network scaling effects in graph analytic studies of human
392 resting-state FMRI data. *Resting state brain activity: Implications for systems neuroscience* **2010**, p. 40.
- 393 21. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: uses and interpretations.
394 *Neuroimage* **2010**, *52*, 1059–1069.
- 395 22. Meng, L.; Striegel, A.; Milenkovic, T. Local versus Global Biological Network Alignment. *arXiv preprint*
396 *arXiv:1509.08524* **2015**.
- 397 23. Ciriello, G.; Mina, M.; Guzzi, P.H.; Cannataro, M.; Guerra, C. AlignNemo: A Local Network Alignment
398 Method to Integrate Homology and Topology. *PloS one* **2012**, *7*, e38107.
- 399 24. Mina, M.; Guzzi, P.H. AlignMCL: Comparative analysis of protein interaction networks through Markov
400 clustering. Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on.
401 IEEE, 2012, pp. 174–181.
- 402 25. Saraph, V.; Milenković, T. MAGNA: maximizing accuracy in global network alignment. *Bioinformatics*
403 **2014**, *30*, 2931–2940.
- 404 26. Neyshabur, B.; Khadem, A.; Hashemifar, S.; Arab, S.S. NETAL: a new graph-based method for global
405 alignment of protein–protein interaction networks. *Bioinformatics* **2013**, *29*, 1654–1662.
- 406 27. Patro, R.; Kingsford, C. Global network alignment using multiscale spectral signatures. *Bioinformatics* **2012**,
407 *28*, 3105–3114.
- 408 28. Sun, Y.; Crawford, J.; Tang, J.; Milenković, T. Simultaneous optimization of both node and edge conservation
409 in network alignment via WAVE. International Workshop on Algorithms in Bioinformatics. Springer, 2015,
410 pp. 16–39.
- 411 29. Guzzi, P.H.; Milenković, T. Survey of local and global biological network alignment: the need to reconcile
412 the two sides of the same coin. *Briefings in Bioinformatics* **2017**, p. bbw132.

- 413 30. Stelzl, U.; Worm, U.; Lalowski, M.; Haenig, C.; Brembeck, F.H.; Goehler, H.; Stroedicke, M.; Zenkner,
414 M.; Schoenherr, A.; Koeppen, S.; others. A human protein-protein interaction network: a resource for
415 annotating the proteome. *Cell* **2005**, *122*, 957–968.
- 416 31. Ibragimov, R.; Malek, M.; Guo, J.; Baumbach, J. GEDEVO: an evolutionary graph edit distance
417 algorithm for biological network alignment. OASISs-OpenAccess Series in Informatics. Schloss
418 Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013, Vol. 34.