Probabilistic graph alignment applied to brain connectomes

Teresa Lázaro[1], Roger Guimerà[1],and Marta Sales-Pardo[1]
[1]URV, teresa.lazaro@urv.cat

The graph alignment problem refers to the task of finding the correspondence mapping between two or more graphs. This problem has been considered in many disciplines such as in computational chemistry (to map molecules), computational biology (to annotate proteins by aligning protein connectomes), information sciences (to identify similar roles in different knowledge graphs) [4][5]. In addition, recent technological developments have enabled the reconstruction of full brain volumes to obtain brain connectomes at synapse resolution levels for several specimens of *C. elegans* in different developmental stages [1], and for two specimens of larva and female adult in *D. melanogaster* [3]. The availability of such rich datasets call of the development of tools that enable the comparison of these connectomes.

Current approaches to the graph alignment problem focus on the alignment of pairs of graphs [2]. Here, we propose a probabilistic approach that allows to simultaneously align multiple graphs. In contrast to current approaches, our approach is not only able to find the best alignment of multiple graphs simultaneously, but also able to collect statistics about the different alignment configurations of the different graphs.

A probabilistic model for (connectome) graph alignment. Our approach is based on the empirical observation that connectomes of different specimens have many features in common. We thus assume that there exists an underlying blueprint $\mathbf{L}$ of the connectome graph of a given species, so that the brain of specimen k of that species $\mathbf{A}^k$ is a copy of $\mathbf{L}$ with a certain level of error q. In general, the labels of the nodes in $\mathbf{A}^k$ is a permutation π^k of the labels of the nodes in $\mathbf{L}$.

In our probabilistic approach, the problem of finding the best alignment of multiple graphs reduces is equivalent to finding the blueprint $\mathbf{L}$ and the set of permutations $\{\pi^k\}$ that relate node identities in each graph to the blueprint that maximize the posterior probability $p(\mathbf{L}, \{\pi^k\}|\{\mathbf{A}^k\})$.

To sample this posterior probability (and find the configuration that maximizes the posterior) we use a Markov chain Monte Carlo algorithm with Parallel Tempering that incorporates biological information about the nodes (neurons), such as the knowledge of the neuroblast group to which each neuron belongs or the *a priori* knowledge of certain node identities (which we call anchors).

Our probabilistic approach allows us to recover with good accuracy the alignment of synthetic an real networks. Figure (1, a-b) shows results for the alignment of four *C. elegans* connectomes corresponding to late larval and adult stages of development form [2] ; and of the two hemispheres of the larva of *D. melanogaster*, as well as to obtain their associated blueprints.

Additionally, we are able to infer the group label of some neurons in cases where there are lack of information about the neuroblast they originated from. We show results for the connectomes of *C. elegans* in Fig. (1, c-e). This extra possibility of "group annotation" is really interesting because it allows us to classify the nodes into groups when the information is missing, that can happen often with the experimental data.

References

[1] Joshua T. Vogelstein, John M. Conroy, et al.*Fast Approximate Quadratic Programming for Graph Matching.*, PLOS ONE, 10(4):e0121002, April 2015.

[2] Daniel Witvliet, Ben Mulcahy, James K. Mitchell, et al. *Connectomes across development reveal principles of brain maturation.*, Nature, 596(7871):257–261, August 2021

[3] Winding M, Pedigo BD, Barnes CL, Patsolic HG, Park Y, Kazimiers T, Fushiki A, Andrade IV, Khandelwal A, Valdes-Aleman J, Li F, Randel N, Barsotti E, Correia A, Fetter RD, Hartenstein V, Priebe CE, Vogelstein JT, Cardona A, Zlatic M. *The connectome of an insect brain.*, Science, 379(6636): eadd9330.

[4] Guiyang Liu, Canghong Jin , Longxiang Shi, Cheng Yang, Jiangbing Shuai and Jing Ying*Enhancing Cross-Lingual Entity Alignment in Knowledge Graphs through Structure Similarity Rearrangement* Sensors 2023, 23, 7096. https://doi.org/10.3390/s23167096

[5] Warith Eddine Djeddi, Sadok Ben Yahia, Engelbert Mephu Nguifo. *A Novel Computational Approach for Global Alignment for Multiple Biological Networks.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2017, IEEE/ACM Transactions on Computational Biology and Bioinformatics

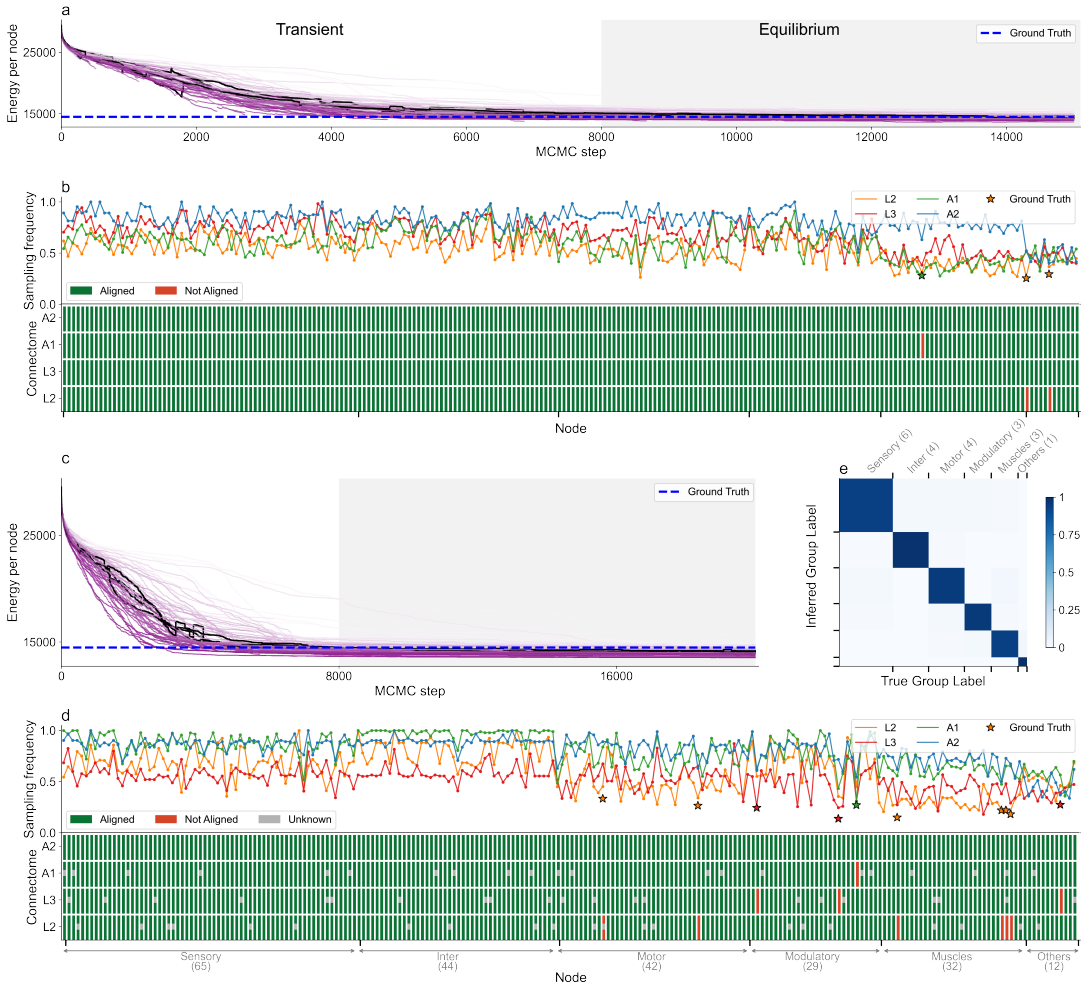

Figure 1: **Alignment of four C.elegans connectomes corresponding to four different stages of development**.
(a-b) All group labels are known: (a) Minimization of the energy (associated to the maximization of the posterior probability), using Markov chain Monte Carlo algorithm with Parallel Tempering. The blue dashed line correspond to the ground truth. In this case the ground truth follows the annotation of the label nodes provided by the experimental results [1]. **(b)** Sampling at temperature $\beta = 1$ of the Equilibrium zone in (a): The figure below correspond to the most probable alignment for each node for each connectome, dark green if they are aligned, red if not. The figure above plot their their respective sampling probabilities. The stars, labeled as "Ground Truth", correspond to the probability of the "right" alignment of the nodes in cases where the most probable align is not the ground truth one.
(c-e) Some group labels are unknown: (c) is analogous to (a) but in this case some nodes are not included in any of the group and thus they do not have the movement restrictions by groups. **(d)** is analogous to (b), but nodes whose group is unknown have a gray rectangle in the center . As it can be seen, the only layer with all group labels known is the adult one, the A_2 connectome. **(e)** Probability matrix for the Inferred group label of the nodes without group label (the number of group unlabeled nodes of each group is proportional to the group size, annotated in gray above). When all the probability is in the diagonal, the inferred group label is in agreement with the true group label, as it is the case