Graph alignment across connectomes using Bayesian Inference

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

he graph alignment problem refers to the task of finding the correspondence mapping between two or more equivalent graphs, i.e. finding the right permutation between their node labels. This is an active research area used in several interdisciplinary fields, including, for instance, the comparison of protein structures or the identification communities in a social network. In our study, we have developed a Bayensian inference-algorithm to align N graphs simultaneously.

In particular, our objective is to align four experimental connectomes obtained from the brains of C.elegans at different developmental stages [1]. Since individuals that belong to the same species share common brain features, it is meaningful to search for the mappings between their connectomes: We can consider their brains as variations of a common "brain template", also called the blueprint. Therefore, our aim is to infere the right permutations between the node labels of the four brain graphs, as well as identifying the most likely blueprint that generates these connectomes.

In the bayesian framework, we first must to formalize a generative process (with q the copy error probability of an edge/neuronal connection) for the different networks we observe, ($\mathbf{A}^k$ are the observed adjacency matrices). The graph alignment problem in this case is then to match each network to the latent matrix, or, in other words, to find the permutation of network indices π^k for each network $\mathbf{A}^k$ that match the underlying blueprint. In what follows, we will call this blueprint as the latent (directed) adjacency matrix $\mathbf{L}$.

Finally, our goal will be then to find the set of permutations $\{\pi^k\}$ and the latent matrix $\mathbf{L}$ that maximize the posterior probability $p(\mathbf{L}, \{\pi^k\}|\{\mathbf{A}^k\})$, Eq. (1).

$$p(\mathbf{L}, q, \{\pi^k\}|\{\mathbf{A}^k\}) = \frac{p(\{\mathbf{A}^k\}|\mathbf{L}, q, \{\pi^k\})p(\mathbf{L})p(q)p(\{\pi^k\})}{p(A)} \quad (1)$$

To maximize this Bayesian probability (the best solution), we have developed a Parallel Monte Carlo algorithm that incorporates biological assumptions about the nodes, such as the knowledge of the neuroblast group to which each neuron belongs or the *a priori* knowledge of certain node labels (called the anchors).

This probabilistic approach allows us to recover with good accuracy the alignment of the four C.elegans connectomes. Until now these kinds of alignments were done only for pairs [2], in easier problems and without our possibilities of fitting the generative model to other problems. Additionally, we are able to infer the group label of some neurons in cases where their neuroblast groups is unknown.

[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):257261, August 2021